I

STATISTICS OF CRIME AND DELINQUENCY: PROGRESS IN MEASUREMENT

Chairman, D. M. Barbour, U. S. Bureau of the Budget

FBI STATISTICS ON OFFENSES AND OFFENDERS

Jerome J. Daunt, Federal Bureau of Investigation

Criminal statistics are difficult to compile, understand and interpret. Police statistics such as published in the Uniform Crime Reports are no less so. Yet progress in the police control of crime as well as other phases in the administration of criminal justice requires more accurate and pertinent statistics on offenses and offenders. Stated very simply, what makes up these numbers and what do these numbers mean?

Uniform Crime Reporting is a voluntary cooperative program of local law enforcement agencies and the FBI to provide a nationwide view of crime, its extent, fluctuation and distribution. The system was originally developed by the International Association of Chiefs of Police in conjunction with social scientists and experts on the law and public administration. Since its inception in 1930, it has been administered by the FBI and continues to be endorsed and actively supported by the Committee on Uniform Crime Records, IACP.

Basically, Uniform Crime Reports is a collection of data on crimes known to the police either through citizen complaints or police arrest of offenders. The most widely known and used portion and perhaps the most important because of its intended purpose is the Crime Index. This Index consists of 7 offenses: murder and nonnegligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny $50 and over in value and auto theft. It was designed to serve as a convenient measure of criminality in a community. We call these 7 offenses Index or serious crimes because they are serious by their very nature or because they occur in such volume they are serious community and police problems. These are perhaps best described as predatory crimes. This Index obviously does not indicate the extent of organized crime or what is commonly referred to as "white collar" crime nor does it attempt to. It is, however, a practical measure of police activity as it relates to the most common local crime problems.

The uniform definitions of these crimes are primarily based on legal or the more common statutory terms. The rule is that all contributing agencies count and classify these offenses based on the uniform definitions without regard to local statutes or prosecutive policy. Necessarily, the definitions are quite broad and leave an area of discretion although the vast majority of criminal acts can be readily classified within the framework of these definitions. Nonetheless, because of the wide variation in state penal codes, it is necessary to frequently remind contributors that the uniform crime definitions must be applied and local law ignored for the purposes of crime reporting.

For the past 3 years we of the FBI staff and the police executives who comprise the Committee on Uniform Crime Records have been giving special attention to improvement of crime reporting through a more detailed description of the criminal acts involved. Because of the role played by the Crime Index our efforts have been primarily focused on the Index offenses. We have conducted special monthly surveys on a nationwide basis looking into the nature of certain crimes such as burglary, aggravated assault, homicide, auto theft, etc. The purpose of the surveys is two-fold; first, to examine the consistency of reporting and secondly, to develop practical subdivisions of these broad crime categories and thereby make available a better statistical description of these criminal acts.

We feel that through the additional reporting of certain specific information concerning general crimes, more consistent crime reporting will result. This additional information is normally included as essential data in police offense and investigative reports and so provides each contributing police agency with a better description of the types of criminal acts to be reported in each classification.

Of equal importance is the fact that through a more detailed itemization of criminal acts law enforcement is in a far better position to identify their nature and the extent to which the incidence of certain crimes can or cannot be controlled by law enforcement. This should be done by analyzing the occurrence of these crimes in relation to law enforcement's limited functions in crime control; namely, prevention by exposing the nature of crime, principally the effect of contributing community factors; suppression through properly oriented and alert patrol activity; and deterrence through successful investigation and apprehension of the offender. Police success then depends heavily on local citizen understanding and support, the action of the courts and the effectiveness of the correctional and rehabilitative process.

There is here available for illustration a revised monthly Return A which will be introduced in the Uniform Crime Reporting Program beginning in January, 1964. This is the vehicle for collecting the number of reported Index and Part 1 offenses from contributing police agencies nationally. Since basic police records and crime reporting had a common beginning in this country tally sheets and an instructional manual for establishing a basic record system to provide such a crime count is distributed free of charge to all contributors.

The approach by subdividing the broad crime classifications is apparent in forcible rape, robbery, assault and burglary which were previously collected as totals without differentiation. With respect to homicide it should be noted that in the summer of 1961 a Supplementary Homicide Report was initiated in the Uniform Crime Reporting Program. Rather than simply a request for a numerical count of willful killings, this new form requires the age, sex and race of the victim, the type of weapon used and the motive or circumstances surrounding the killing when it is known to the police. The Supplementary Homicide Report has become an excellent supervisory tool in verifying the count of willful killings by eliminating duplication and killings not properly classified within the uniform definition. It has also made available considerable information some of which was set forth in Uniform Crime Reports - 1962 concerning the character of these willful killings and the extent to which they were beyond control of the police.

These revisions in the Uniform Crime Reporting Program in no way change the standard definitions. Perhaps an explanation of the breakdown of assault and burglary best describe the reasoning involved. Because of the wide variation in state laws as well as prosecutive practice, these two crime classifications present a serious problem in uniform crime reporting. While all assaults, both simple and aggravated, now become an offense known to the police (Part 1) the Index will still be limited to those which are of an aggravated nature, namely 4A through 4D. As a result of a survey on aggravated assault in August, 1960, a ratio of weapons used was established. That survey indicated that over 1/2 of the aggravated assaults were committed with the use of 4A gun and 4B knife or other cutting instrument. This becomes a useful tool in supervising reporting. Keep in mind that all assaults with a dangerous weapon and with an intent to commit serious injury are classified as aggravated even though no injury results or the injury is minor. While we would agree that assaults are not consistently reported to the police, the extension of Part 1 offenses to include all assaults provides each contributor with a better description of the types of criminal acts that should be included in this crime classification. At the same time, because of where and among whom they happen (2/3 occur within the family or among neighbors and acquaintances) assaults perhaps best demonstrate law enforcement's limited ability to control crime and the extent to which it is a social ill.

Burglary has been defined in the Uniform Crime Reporting Program from the beginning as any breaking or unlawful entry of a structure to commit a theft. Yet, in many jurisdictions a charge of burglary is only possible after establishing forcible entry. We know from a survey in October, 1961, that 70 percent of the burglary results from forcible entry, 20 percent from unlawful entry and 10 percent were attempts or means of entry was unknown. It is felt this new subdivision of burglary will provide a more uniform count as well as more meaningful statistics. It is possible that in the foreseeable future the Crime Index burglary will be limited to that resulting from forcible entry.

While no change is recommended in the larceny classification for 1964, it is still receiving serious consideration. Although a survey of state statutes in 1961 revealed that the median breaking point between grand and petty larceny is still $50 among the 50 states, there is strong opinion that larceny by dollar evaluation should be dropped from the Crime Index. Larceny by type under the following categories will be collected on a supplementary return in 1964: (A) Pocket-picking (B) Purse-snatching (C) Shoplifting (D) From Autos (Except E) (E) Auto accessories (F) Bicycles (G) From Building (Except C) (H) From any coin-operated device or machine (Except G) (I) All other. We feel that 4 or 5 of the above types of larceny as a group should represent the Crime Index larceny offense without respect to the value of property stolen. This would provide greater uniformity in reporting and more consistency in trend information in view of the difficulty in obtaining accurate and comparable property loss values and the ever present cost of living factor.

RETURN A
MONTHLY RETURN OF OFFENSES KNOWN TO THE POLICE

TO BE FORWARDED TO THE FEDERAL BUREAU OF INVESTIGATION, U. S. DEPARTMENT OF JUSTICE, WASHINGTON, D. C., BY THE SEVENTH DAY AFTER CLOSE OF MONTH. See other side for instructions - also tally sheet.

1 CLASSIFICATION OF OFFENSES (PART I CLASSES)	2 OFFENSES REPORTED OR KNOWN TO POLICE (INCLUDE "UNFOUNDED" AND ATTEMPTS)	3 UNFOUNDED, I. E., FALSE OR BASELESS COMPLAINTS	4 NUMBER OF ACTUAL OFFENSES (COLUMN 2 MINUS COLUMN 3) (INCLUDE ATTEMPTS)	5 NUMBER OF OFFENSES CLEARED BY ARREST THIS MONTH a TOTAL OFFENSES CLEARED	5 b BY ARREST OF PERSONS UNDER 18 (INCLUDED IN 5a)
1. CRIMINAL HOMICIDE a. MURDER AND NONNEGLIGENT MANSLAUGHTER					
b. MANSLAUGHTER BY NEGLIGENCE					
2. FORCIBLE RAPE TOTAL					
a. RAPE BY FORCE					
b. ASSAULT TO RAPE - ATTEMPTS					
3. ROBBERY TOTAL					
a. ARMED - ANY WEAPON					
b. STRONG-ARM - NO WEAPON					
4. ASSAULT TOTAL					
a. GUN					
b. KNIFE OR CUTTING INSTRUMENT					
c. OTHER DANGEROUS WEAPON					
d. HANDS, FISTS, FEET, ETC. - AGGRAVATED					
e. OTHER ASSAULTS - NOT AGGRAVATED					
5. BURGLARY TOTAL					
a. FORCIBLE ENTRY					
b. UNLAWFUL ENTRY - NO FORCE					
c. ATTEMPTED FORCIBLE ENTRY					
6. LARCENY - THEFT (EXCEPT AUTO THEFT) a. $50 AND OVER IN VALUE					
b. UNDER $50 IN VALUE					
7. AUTO THEFT					
GRAND TOTAL					

TOTAL ARRESTS THIS MONTH For ALL OFFENSES Except Traffic	
ADULTS	JUVENILES (age as defined by your state)

DATE

PREPARED BY TITLE

CHIEF, COMMISSIONER, OR SUPERINTENDENT

MONTH AND YEAR

DO NOT USE THIS SPACE	
	INITIALS
RECORDED	
REVIEWED	
PUNCHED	
VERIFIED	
ADJUSTED	

CITY AND STATE

Auto theft was the subject of a special survey in October, 1962. Subdivision of this crime classification is also under continuing study by the FBI staff and the Committee on Uniform Crime Records, IACP. As indicated in the November, 1962, survey, about 80 percent of the autos stolen were for the purpose of transportation only or the purpose of theft was unknown. Eight percent were taken for the purpose of stripping for parts, 5 percent were used in another crime or escape and the remainder for the purpose of resale. While police agencies cannot always distinguish between the so-called "joy ride" and theft for profit or other purpose solely on the recovery of a stolen car, some such breakdown on a continuing basis would be useful not only in police administration but in a better statistical description of this offense.

In addition, changes in Uniform Crime Reporting as noted under 5B of the Return A will bring about a more complete Index as to the volume of crime being committed by young persons in each community. Heretofore, the measure available in Uniform Crime Reports as to juvenile criminality was limited to police arrests for specific criminal acts. The new measure is based on the number of offenses which are cleared-up either by the arrest of a person under 18 years of age or through exceptional clearances. These clearances will also be related directly to the subdivisions of the Part 1 crime classifications which will give us a better understanding of the specific nature of the criminal acts in which young people are becoming involved. Other changes in the Uniform Crime Reporting Program and which time does not permit lengthy discussion include a breakdown of narcotic and gambling arrests by type as well as separate additions to the Uniform Classifications of Offenses such as arson, vandalism, etc.

While Uniform Crime Reports furnishes extremely useful information on offenses, there is little information on the offender. In Uniform Crime Reports, this is limited to age, sex and race. In January, 1963, the FBI launched a new statistical program to develop meaningful data on known offenders. This might be described as a follow-up on known offenders. Briefly, the criminal history of offenders as known through FBI identification records is being processed and stored on tape for computer analysis. A flagging mechanism within the Identification Division of the FBI which handles the fingerprint inquiries will make available all subsequent violations on each offender as recorded in fingerprint identification. The key to this new statistical program is the FBI one-number system for each offender. The information normally available on these criminal histories is date, place and identity of arresting agency, charge at the time of the arrest, court disposition, correctional action and the age, sex and race of the offender. The data is being coded in detail in order that a wide variety of analyses may be made. The first experimental tabulations utilizing this information will be available in a few months. We are particularly interested in recidivism studies measured by arrest, conviction and commitment, analysis of criminal behavior patterns, mobility of offenders and the extent to which the same offenders contribute to our crime counts year after year. Records being processed at the present time are basicly those offenders who are being handled currently in some phase of the federal law enforcement system. We anticipate that at the end of the first calendar year, criminal histories on 50,000 individuals will have been processed.

Law enforcement is making valuable use of crime statistics particularly in the area of man power distribution but more needs to be done in using statistics as the tool to express a better identification of the nature of crime and the many types of offenders. Armed with the knowledge gained from additional analysis, law enforcement is better able to direct its efforts and at the same time supply valuable information to the courts and those in the correctional process as well as social scientists and the general public since all of us have a "stake" in the problem.

MEASUREMENT TECHNIQUES USEFUL IN EVALUATING THE EFFECTIVENESS OF PROGRAMS FOR PREVENTING AND CONTROLLING DELINQUENCY

Edward B. Olds, Social Research Consultant

Juvenile delinquency like many other behavior based problems is extremely difficult to measure unless delinquency is defined narrowly in terms of officially ajudicated violations of laws or ordinances. Many studies have shown that only a small proportion of offensive, law violating behavior results in arrests and only a portion of the arrests results in recorded cases of juvenile delinquency.*

The measurement of juvenile delinquency can be regarded as an end in itself or as one of several aspects of preventing and controlling this social problem. In this paper the focus is on measurement techniques useful in evaluating programs of prevention and control.

There are at least three levels of sophistication in the use of delinquency measurements in relation to prevention and control. At the first level, counts of referrals to the juvenile court, officially ajudicated delinquency cases or police arrests of juveniles are commonly used to sound the alarm or document the success of various programs. Even though such data are usually converted to ratios based on the estimated population at risk, there is little attention given to problems such as random variation in rates and the effects of changes in the age, sex, race, and socio-economic composition of the population. Editorial writers, civic leaders, and administrators of agencies are inclined to make uncritical uses of delinquency rates. Actions taken on the basis of questionable data may entail the passage of new laws, enlargement of budgets and staff or tightening up on law enforcement. The net effect of such actions may be a further increase in the reported delinquency cases as a result of the new laws, more policemen to make arrests, and a larger ratio of apprehensions.

At a second level, the reports on individual juvenile delinquency cases are coded according to the census tract of the delinquent's residence. Tabulations of such data by sex, race, and census tract converted into rates provide a measure of the comparative incidence of reported delinquency cases in different types of neighborhoods. Through correlation analysis based on indices of the social and economic characteristics of census tracts, it is possible to identify variables associated with the apprehended and reported delinquency. Regression analysis can be employed to make predictions of delinquency rates for comparison with the actual rates. The selection of sites for demonstration programs to prevent and control delinquency may be improved through access to such data. Later in this paper an illustrative use of regression analysis techniques will be described in relation to data for Washington, D. C.

At the third level, a variety of carefully tailored measurement techniques is sometimes used in evaluating the effectiveness of neighborhood based delinquency control programs. Changes in the measurements taken before and after conducting action projects facilitate drawing conclusions concerning the extent to which objectives are achieved. Measurements are also needed in control areas to determine the extent to which changes can be credited to the planned actions.

Although sound measurement techniques are commonly used in fields such as biological and agricultural experimentation they are not widely used in research to solve social problems such as juvenile delinquency. However, there is a small trend developing towards a greater use of such methods as indicated by the demonstration projects to prevent and control delinquency under development in a number of large cities through financing from the President's Committee on Juvenile Delinquency and Youth Crime. Criteria used by the Review Panel in making grants include adequate plans for systematic evaluation as well as explication of a well conceived theoretical framework to provide a basis for the planned actions.

Community programs directed towards preventing and controlling delinquency are tending towards a comprehensive approach involving many of the major institutions such as school systems, welfare departments, recreation agencies, employment offices and places of employment as well as the correctional agencies. Programs conducted in school systems include vocational education, special classes, counselling, sheltered employment experience, and cultural

* James F. Short and F. Ivan Nye, "Extent of Unrecorded Juvenile Delinquency: Tentative Conclusions," the Journal of Criminal Law, Criminology and Police Science Vol. 49, Nov.-Dec., 1958 296-302.

enrichment. Public welfare departments may use reduced case loads with specialized workers to provide services for delinquent prone children in families receiving public assistance, group counselling for parents of delinquent youth, and work relief programs. Recreation departments may provide specially trained workers who are assigned to work with delinquent gangs, after-school recreation programs, and leisure activities to meet the need for adventure and excitement as well as new experience and relaxation. Correctional institutions test various levels of group as well as individual counselling.

Theorectical assumptions underlying the design of programs may be broadly classified into those which stress the personality and psychological attributes of individual youth as causative factors and those which stress social systems such as friendship groups, neighborhoods and subcultures. Variants of these theories stress the importance of effective social controls. Some programs reflect a major emphasis upon the family as an influential and continuing social system determining much of individual behavior. Since many factors separately and jointly produce delinquency, each brand of theory tends to find some justification. However, the galaxy of theories and programs complicates evaluation and is confusing to citizen leaders.

Evaluation techniques need to draw upon experience from many fields in the use of experimental designs and related statistical theory. The target population towards which the action is directed needs to be carefully defined as well as the environmental setting in which the action takes place. Furthermore, the exact nature of the action must be specified. The kind and amount of change attributable to the actions should be measured as well as the unexpected negative effects. It is desirable to have an assessment of the likely long term effects of the program. From a broader policy standpoint it is essential to know the costs of the program as well as the investment of volunteer effort. A more subjective but obviously significant criterion of success relates to the effect of the project in convincing leaders of the necessity for allocating more resources to carry on continued research and experimentation.

Measurement techniques may be broadly classified as to whether they are directed towards assessing the "net impact" or total effect of a whole set of programs or whether they are directed towards determining the effect of specific programs. They may also be grouped according to whether the data are used primarily as aggregations for small areas such as census tracts or primarily in relation to specific individuals.

Measurement data may also be grouped according to source and method of collection. On the one hand official or semi-official data are collected by police, court and treatment agencies. For the most part these data have not been standardized although the Children's Bureau and the Federal Bureau of Investigation have made progress in this direction. Since the official reports do not reflect the bulk of delinquent behavior, other measurement techniques are being explored. The "self-report" method entails the use of group administered questionnaires listing major classifications of delinquent acts in relation to which youth are asked to report the degree of their own participation. These questionnaires are usually administered without signatures so as to encourage accurate reporting. Experience indicates that where great care and skill is exercised in administering such questionnaires, quite plausible data can be obtained. However, most of the studies using this method have been conducted in smaller cities.

Judgements concerning the volume of offensive behavior occurring in specific neighborhoods have been sought from residents, business men, and community service workers. Interviewees may also be asked to nominate specific youth presenting serious behavior problems. The effectiveness of this method obviously depends upon the extent to which informants are intimately acquainted with the neighborhood situation and the youth population.

A most promising measurement technique entails the use of a semi-projective questionnaire which can be administered to youth assembled in groups. One such instrument, the Jessness Inventory, has been used to differentiate the delinquents from the non-delinquents in school populations. It is claimed that this instrument has been able to identify correctly about 85 percent of the delinquents and about the same proportion of the non-delinquents in a particular school where fairly accurate information was available on delinquency from other sources.

The measurement of the specific effects of individual programs within a large set of programs entails the use of somewhat different measurement techniques. One approach is to use specially designed interview or questionnaire methods directed towards participants and non-participants in the specific program. Adults intimately acquainted with the individual youth may also be interviewed to obtain judgements concerning the effects. Programs may be carefully observed by trained personnel to identify features with positive and negative impacts. Case records on individuals may be analyzed in depth to obtain reflections of the changes attributable to particular programs.

An interesting use of statistical techniques in evaluating the results of programs of prevention and control is illustrated by the work done in the District of Columbia to appraise the selection of census tracts in which demonstration programs are to be conducted. Through the use of many statistical and operational criteria, 18 contiguous census tracts were selected by Washington Action for Youth* to constitute a "target area" in which to locate a set of action programs to prevent and control delinquency. One of the major criteria for evaluating the net effect of the action programs is assumed to be the change in the percentage of youth in each census tract referred to the Juvenile Court. If analysis should indicate that at the beginning of the project the 18 census tracts had delinquency rates considerably above prediction, the probability would be high that many of the census tracts would exhibit a reduced delinquency rate after a year or two even though no action programs were conducted. Such a trend is to be expected through the operation of the "regression toward the mean" principal. Many studies have found that extreme deviates from a mean position tend to change by regressing towards the mean without any assignable cause for such change. Accepting this principle it is important to determine how census tracts in a demonstration or target area are distributed with respect to the ratio between the actual and predicted delinquency rate at the starting and termination dates. If the tracts are approximately randomly distributed with respect to the ratio between the actual and predicted rate at the starting date and change in the direction of a lower ratio at the termination date, it could be concluded that the action program was effective. Statistical analysis could be used to determine the degree of significance to be accorded the observed change.

However, before conclusions can be drawn it would be necessary to obtain answers to questions such as the following:

1. Was there any change in the procedures or criteria used in reporting offensive behavior in the target area, which could explain the observed changes in rates?
2. Was there an unusual change in the make-up of the population at risk sufficient to account for the observed changes?
3. Did unplanned events or actions take place within the target area between the starting and termination dates which could influence delinquency rates to a greater extent in the target area than in the rest of the city?

Personnel responsible for the final evaluation would need to keep such questions in mind in planning their research.

The study to be described was concerned only with the development of procedures for predicting delinquency rates within the target areas on the basis of the relationships prevailing among census tracts throughout the city between the delinquency referral rate and an efficient set of predictors. It was hoped that the identification of patterns in the District of Columbia might be useful in selecting predictors in other cities.

The first step was to obtain a census tract tabulation of delinquency referrals to the D.C. Juvenile Court during a three year period (July 1, 1959 to June 30, 1962). Data were sought for three years rather than for one year so as to reduce the random error in rates which results when rates are based on too small a population. The analysis was restricted to the male delinquents so as to have a more homogeneous population than if female delinquents were included, with rates amounting to less than one-fifth of the rates for the males. The male delinquency referrals were divided into two groups: (1) a white group comprising 14.4 percent of the male delinquency referrals; (30.7 per cent of the population 10 to 17 years of age in the D. of C. was white in 1960) (2) a nonwhite group comprising 85.6 percent of the male delinquency referrals; (69.3 percent of the population 10 to 17 was nonwhite). This resulted in a rate of 8.5 white male referrals to the Juvenile Court during the three year period per 100 white boys 10 to 17 years of age in 1960, as contrasted with a corresponding rate of 22.3 for the nonwhites.

Because of the sizeable differences

---

* This is the Washington, D.C. organization financed largely by the President's Committee on Juvenile Delinquency and Youth Crime.

in delinquency referral rates between the white and nonwhite males it was decided to conduct the regression analysis in two parts:

Part A: using data on referrals of white male delinquents and independent variables based on the white population.

Part B: using data on referrals of nonwhite male delinquents and independent variables based on the nonwhite population.

Accordingly, ratios were computed for the white and nonwhite segments of each census tract for each of the 15 independent variables shown in Table 1. The 15 variables are arrayed according to the average correlation coefficient between each variable and the male delinquency referral rate. Ranks are shown for the variables in each of the two groups on the basis of the correlation between the independent variables and the delinquency referral rate. The last column shows the difference in ranks between the white and nonwhite segments. The coefficient of rank correlation between the ranking of the coefficients based on the white and nonwhite segments was only .294. This indicates that there is only slight similarity between the pattern of relationship with the independent variables in the white and nonwhite segments. It will be noticed, however, that for six of the 15 variables there was a correlation of .55 or higher with juvenile delinquency for both the white and nonwhite segments. Three of these six variables ranked higher among the white segments and two among the nonwhite. It so happened that these six variables were eventually selected as the predictors in the final multiple regression equations.

Through the use of facilities at the National Bureau of Standards it was possible to explore the effect of special combinations of items as well as the effect of editing the data. With an Omnitab program developed for use on the IBM 7090 computer, the following output for a particular set of variables was obtained in about three minutes of machine time:

1. Computation of the regression coefficients, standard error of estimate and many additional statistical measures.
2. Calculation of squared residuals from which the multiple correlation coefficient was easily derived.
3. Plotting of scatter diagrams showing the distribution of the census tracts according to delinquency rate and each of the independent variables.
4. Computation of the predicted delinquency rate based upon the regression equation as well as the difference between the actual and predicted rates for each census tract segment.
5. The above operations carried out separately for the white and nonwhite segments of the census tracts.

By slight changes in the parameter cards the order and combination of the independent variables were easily changed. Furthermore, editing was accomplished by simply removing the appropriate census tract cards.

The first run using all 69 white census tract segments and all 82 nonwhite segments with all 15 independent variables produced a multiple correlation coefficient of .821 for the white segments and .809 for the nonwhite segments. However, the standard error of estimate for the nonwhite segments was much too high (11.53) as contrasted with the comparable figure for the white segments (5.50). Examining the scatter diagrams a peculiar pattern was observed for the nonwhite segments with ratios above 40. This pattern is illustrated in Figure 1-B with the bulk of the nonwhite tracts distributed similarly to the white segments in a relatively smooth regression pattern. On the other hand, the nonwhite segments which had rates above 40 seemed to be responding to a special set of forces with little relationship to the independent variables. Examination of the scatter diagrams for the other 14 variables showed that a similar condition prevailed in all but three of the diagrams. Accordingly, it was decided to separate the 16 nonwhite segments exhibiting the unusual tendency and analyze them separately as will be described later.

Other editing rules were developed for the isolation of 20 white segments and four additional nonwhite segments. Most of these segments were isolated because they had too small a population of boys 10 to 17 (under 60). It was believed that random variation was introducing "noise" attributable largely to the small population in each of these segments. Another criterion for excluding several census tract segments was residence in group quarters by more than 20 percent of the total population in the census tract. It was reasoned that the inclusion of areas with a large proportion of the population living in

Table 1

ZERO ORDER CORRELATION COEFFICIENTS BETWEEN THE MALE DELINQUENCY REFERRAL RATE* AND EACH OF 15 INDEPENDENT VARIABLES FOR WHITE AND NONWHITE SEGMENTS OF CENSUS TRACTS

Independent variables derived from census tract data published by the Census Bureau	Variable no.	Correlation with male delinquency referral rate		Rank of correlation coefficient		
		White	Nonwhite	White	Nonwhite	Diff
% of employed males in white collar or skilled jobs	5	-.733#	-.725	2	2	0
% of families with income over $5000 in 1959	4	-.747#	-.621#	1	6	-5
Separated, widowed or divorced males as % of ever married males	8	.714#	.629#	3	4	-1
% of population 25 & over with 12 or more years of school	3	-.588#	-.695#	8	3	5
Index of overcrowding**	7	.679#	.558#	4	9	-5
% of employed females in white collar or skilled jobs	6	-.473	-.726#	11	1	10
% of children under 18 living with both parents	2	-.552#	-.628#	9	5	4
% of male labor force unemployed	10	.667	.473	5	11	-6
% of housing units owner occupied	13	-.494	-.605#	10	7	3
Separated, widowed or divorced females as % of ever married females	9	.380#	.602#	12	8	4
% of males 10 to 17 nonwhite	15	.662	.209	6	14	-8
One person households as a % of all households	14	.334#	.521#	13	10	3
% of total population nonwhite	16	.592#	.247	7	13	-6
% of females 14 and over in the labor force	11	.062	-.313	15	12	3
Population in households per household	12	-.212	-.035	14	15	-1

* Male referrals to the D.C. Juvenile Court between July 1, 1959 and June 30, 1962 as a percent of males 10 to 17 years of age according to the 1960 U.S. Census.

** Computed by Metropolitan Population Project by adding percent of occupied housing units with 1.01 to 1.50 persons per room to three times the percent with 1.51 or more persons per room.

\# Variables included in the second set of computer runs to develop and make the predictions.

Figure 1-A

WHITE SEGMENTS OF CENSUS TRACTS PLOTTED ACCORDING TO WHITE MALE DELINQUENCY RATE AND ACCORDING TO PERCENT OF WHITE MALES IN WHITE COLLAR OR SKILLED JOBS

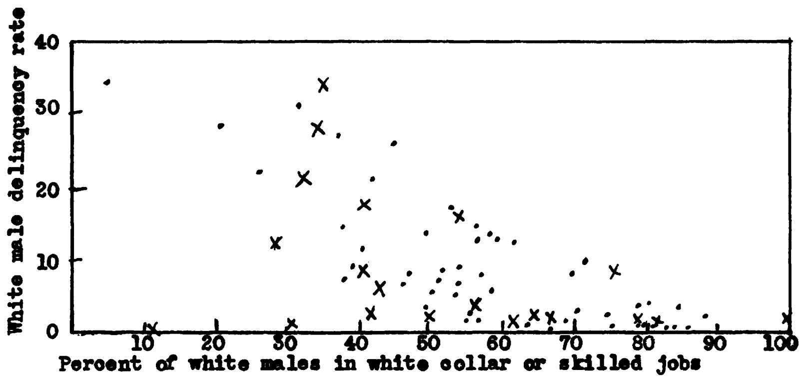


Figure 1-B

NONWHITE SEGMENTS OF CENSUS TRACTS PLOTTED ACCORDING TO NONWHITE MALE DELINQUENCY RATE AND ACCORDING TO PERCENT OF NONWHITE MALES IN WHITE COLLAR OR SKILLED EMPLOYMENT

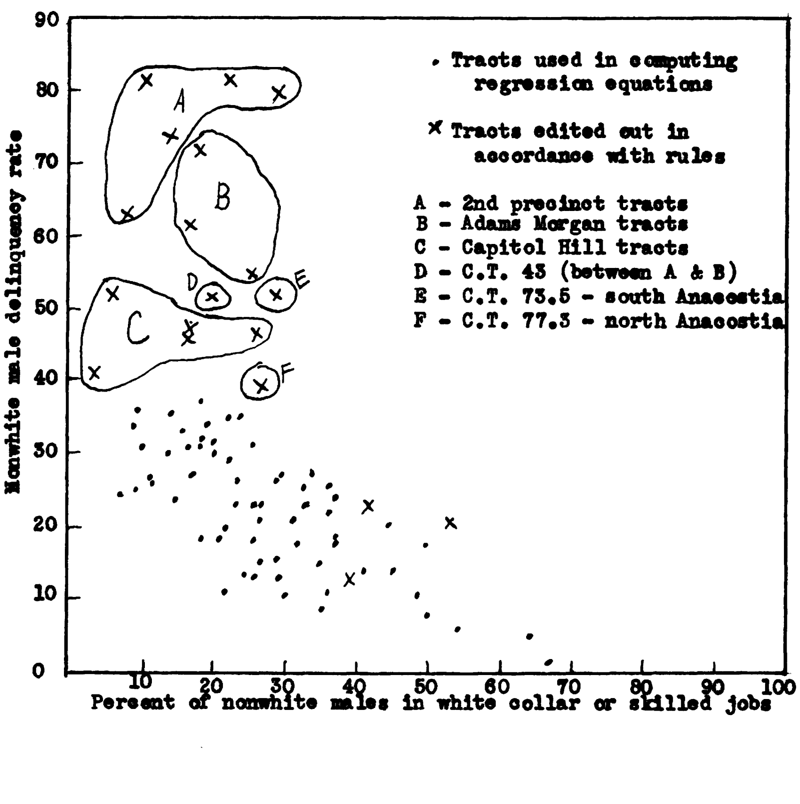


Table 2-A

INTERCORRELATIONS BETWEEN NINE INDEPENDENT VARIABLES USED IN COMPUTING THE MULTIPLE REGRESSION EQUATION FOR PREDICTING DELINQUENCY RATES IN WHITE SEGMENTS OF CENSUS TRACTS IN D.C.

.707

Independent variable*	Var. no.	Variable number 3	4	5	7	16	14	8	9	2
Over 12 yrs. of schooling	3	1.000	.768	.913	-.689	-.588	.253	-.323	.086	.372
Over $5000 income	4	.768	1.000	.862	-.803	-.562	.135	-.557	-.207	.572
Males in white collar jobs	5	.913	.862	1.000	-.762	-.634	.032	-.454	-.098	.492
Overcrowding index	7	-.689	-.803	-.762	1.000	.484	.003	.430	.020	-.366
% of population nonwhite	16	-.588	-.562	-.634	.484	1.000	.096	.562	.330	-.477
One person households	14	.253	-.135	.032	.003	.096	1.000	.694	.896	-.537
Separated, wid. & div. males	8	-.323	-.557	-.454	.430	-.562	.694	1.000	.783	-.721
Separated, wid. & div. females	9	.086	-.207	-.098	.020	.330	.896	.783	1.000	-.675
Children with both parents	2	.372	.572	.492	-.366	-.477	-.537	-.721	-.675	1.000

.293 .718

Table 2-B

INTERCORRELATIONS BETWEEN TEN INDEPENDENT VARIABLES USED IN COMPUTING THE MULTIPLE REGRESSION EQUATION FOR PREDICTING DELINQUENCY RATES IN NONWHITE SEGMENTS OF CENSUS TRACTS IN D.C.

.844

Independent variable*	Var. no.	Variable number 3	4	5	6	7	13	14	8	9	2
Over 12 yrs. of schooling	3	1.000	.861	.944	.932	-.849	.829	-.411	-.632	-.577	.636
Over $5000 income	4	.861	1.000	.816	.852	-.789	.911	-.441	-.492	-.505	.585
Males in white collar jobs	5	.944	.816	1.000	.932	-.820	.794	-.439	-.631	-.571	.654
Females in white collar jobs	6	.932	.852	.932	1.000	-.797	.812	-.478	-.673	-.621	.638
Overcrowding index	7	-.849	-.789	-.820	-.797	1.000	-.723	.261	.416	.358	-.493
Percent of homes owned	13	.829	.911	.794	.812	-.723	1.000	-.585	-.558	-.587	.649
One person households	14	-.411	-.441	-.439	-.478	.261	-.585	1.000	.791	.739	-.631
Separated, wid. & div. males	8	-.632	-.492	-.631	-.673	.416	-.558	.791	1.000	.914	-.772
Separated, wid. & div. females	9	-.577	-.505	-.571	-.621	.358	-.587	.739	.914	1.000	-.795
Children with both parents	2	.636	.585	.654	.638	-.493	.649	-.631	-.772	-.795	1.000

.537 .774

jails, army barracks, college dormitories, etc. would produce irrelevant variation. The position of these excluded areas with respect to delinquency and percent in white collar or skilled jobs is shown by crosses instead of dots in Figures 1-A and 1-B.

Simultaneously with the separation of the deviant areas as described above, the number of independent variables was reduced from 15 to 9 for the white segments and from 15 to 10 for the nonwhite segments. The variables to be included were selected largely on the basis of the data shown in Table 1. In this table the identity of the specific variables included in this run is shown by the number symbol (#) following the correlation coefficient. The multiple correlation coefficients obtained from this run were as follows:

White segments:	.874
Nonwhite segments:	.784

The most important gain noted in this run was the reduction of the standard error of estimate from 11.53 for the nonwhite segments down to 5.95. There was also a reduction in the standard error of estimate for the white segments from 5.50 down to 4.35.

To identify constellations of highly intercorrelated items with comparatively low correlations within constellations, a correlation matrix was prepared as shown in Tables 2-A and 2-B. Only those variables are included in this table with a reasonably high correlation with the delinquency referral rate. The columns and rows in this matrix were rearranged by trial and error until the illustrated pattern was achieved showing two constellations of items. The major constellation in the upper left quadrant of each table revealed a high intercorrelation between the socio-economic variables (.707 for the white segments and .844 for the nonwhite). A second constellation shown in the lower right quadrant presents high intercorrelations between family structure items (.718 for white segments and .774 for nonwhite). However, it was found that the average correlation between the variables in these two constellations was considerably less (.293 for the white segments and .537 for the nonwhite). The decision suggested by this analysis was the selection of four variables from the socio-economic items and two from the family structure items to make up a final predictive equation. Accordingly, a run based on variables number 2, 3, 4, 5, 7, and 8 for both the white and nonwhite segments produced the following multiple regression coefficients and standard errors of estimate:

	Multiple regress. coeff.	Stand. error of est.
White segments	.865	4.29
Nonwhite segments	.766	5.92

It will be noted that the coefficients were nearly as high as those obtained through using a much larger number of variables (nine for the white segments and ten for the nonwhite).

A supplementary regression analysis was conducted using the 16 nonwhite segments which were separated because they showed extremely deviant patterns in the scatter diagram as illustrated in Figure 1-B. A composite socio-economic index and a composite family structural index was developed manually for each of the 16 segments. This index was computed by summing the ranks of the 16 census tract segments in the six variables predictive of socio-economic status. The index of family structure was similarly computed using the four variables predictive of family structure. A regression analysis using these two indexes as independent variables produced a multiple correlation coefficient of .5805 and a standard error of estimate of 12.8. This was judged to be close to the threshold of significance at the .05 level.

It will be recalled that a major purpose for the multiple regression analysis was to appraise the selection of census tracts included in the target area for the demonstration project. When the 17 nonwhite and three white census tract segments included in the target area were classified according to the actual delinquency rate as a percentage of the predicted rate, a reasonably normal distribution was found as shown below:

Observed delinquency rate as percent of actual rate	Number of census tract segments	
140 and over	1	
130 to 139	2	
120 to 129	4	
110 to 119	3	
110 and over		10
100 to 109	3	
90 to 99	4	
80 to 89	1	
70 to 79	2	
70 to 109		10

Even though the distribution seems to be centered at 110 rather than 100 it is quite likely that the scatter is sufficiently random to conclude that the selected census tracts meet the test. However, it should be noted that only three of the 20 segments were white so that the demonstration project located in this area would have to be focussed largely on the nonwhite delinquents.

One of the issues which frequently arises in the discussion of delinquency in cities with racially mixed populations concerns the explanation of the sizeable differences found in the rates for white and nonwhite neighborhoods. A commonly given explanation is that differences in socio-economic conditions and family structures in the white and nonwhite neighborhoods account for the observed differences. A test of this explanation is provided by a special regression run involving the 19 census tracts which had both white and nonwhite segments. Two regression equations were derived: one for the white segments of the 19 census tracts and another for the nonwhite segments of these tracts.* Among the 19 census tracts was one (C.T. 95.2) which had almost the same delinquency rate for the white and nonwhite segments (8.0 and 7.6) as well as similar socio-economic characteristics. However, when the equation based on the <u>white segments</u> was used to predict the delinquency rate for this tract using data for the <u>white segment</u> of census tract 95.2 as the input, a prediction was obtained of only 4.5 or about 56 percent of the actual rate. When the same data were used as input for the nonwhite equation, the predicted rate increased to 10.3. Similarly, when the <u>nonwhite</u> equation was used to predict the delinquency rate using the data for the <u>nonwhite segment</u> of census tract 95.2, the predicted rate was 12.5 or 64 percent higher than the actual rate. The white equation using the data for the nonwhite segment produced a rate of 4.1 or 54 percent of the actual rate. These findings suggest that there are undefined factors operating in nonwhite and white areas which produce higher delinquency rates in the nonwhite than in the white population even though they may have nearly identical socio-economic and family structural characteristics.

It is possible that special studies will be conducted in conjunction with the Washington Action for Youth demonstration project to identify the factors explaining the extremely high nonwhite male delinquency referral rates in 16 census tracts in the District of Columbia as well as the factors beyond those measured by available census indicators which result in higher delinquency rates among nonwhite than white youth. Without definitive research we can only speculate concerning the explanation of these phenomena.

In summary, a number of measurement techniques have been outlined for evaluating delinquency prevention and control projects. An exploratory project has been described involving the use of multiple regression techniques to make predictions concerning delinquency rates in specific census tracts. From these forecasts, ratios of actual to predicted delinquency rates were computed and used to assess the selection of census tracts included in the target area for a demonstration project within the District of Columbia. This assessment indicated that there was an acceptable distribution of census tracts in the selected area with respect to the ratio between the actual and predicted rates.

There is a large need for the extionsion of sound measurement techniques to the evlauation of a wide range of action programs purporting to demonstrate effectiveness in solving a variety of social problems. There is an important application of statistical theory and method in such endeavors.

---

* The two regression equations were as follows:

White delinquency rate:

$$X_1 = 19.76 + .01439X_2 - .01134X_3 - .02978X_4 - .00781X_5 + .01927X_7 + .05363X_8$$

Nonwhite delinquency rate:

$$X_1 = 62.71 - .01544X_2 - .02538X_3 + .00240X_4 - .04089X_5 - .00746X_7 - .03848X_8$$

See Table 1 for description of the six independent variables.

RECENT DEVELOPMENTS IN CANADIAN STATISTICS ON CRIME

William A. Magill, Dominion Bureau of Statistics

I wish to thank you, Mr. Chairman, for inviting me to talk today about recent developments in Canadian Statistics on Crime. We at the Dominion Bureau of Statistics have a strong bond with our opposite numbers in the United States, at least I believe that we do.

I am not going to attempt to contrast the organization of Federal Statistics in Canada with Federal Statistics in the United States. I think it is enough to say that the Dominion Bureau of Statistics is a centralized statistical agency. Thus in the Judicial Section of the Bureau are located police statistics, adult and juvenile court statistics, institutional statistics, both for the federal penitentiaries and for provincial institutions including the training schools, and parole statistics. Other sections of the Bureau deal with other social statistics and, of course, economic statistics. Sections are organized into Divisions, generally speaking on a common subject matter basis and serviced by centrally operated data processing and other divisions.

Historically, the Bureau has published police statistics since 1920. Although the Bureau had provided standard forms for the collection of police statistics and supported these with instructions, the fact was that not all respondents adhered to the requirements of the system. Instead, some respondents used their own systems based on their own definitions. Under such conditions, it was not possible to produce meaningful aggregates for Canada as a whole. Without standard units of count adhered to by all respondents, uniform statistics did not exist. Further, some of the units of count were not of a type to facilitate the integration of police statistics with other statistical series collected by the Judicial section, notably court statistics.

An examination of program requirements was needed, a design to meet the needs of the users, and a system which would not place an undue burden on reporting units. Another major consideration was that the respondents should be able to comply with the program once it was laid down. It was clear that some respondents would have to modify their records systems, and others change to new ones in order to meet the requirements of a uniform program.

The Canadian Association of Chiefs of Police played an important role in the development of the new system. In 1958 this Association formed a Committee on the Uniform Recording of Police Activities. Representatives of federal, provincial and municipal police were named to the committee, as well as members of the DBS staff, and a work group established in the Ottawa area. As developmental work proceeded, the work group kept in touch with the committee members through correspondence. This was supplemented at a later stage by field visits of Bureau officers to individual police departments and members of the Committee. This working arrangement provided close liaison with the respondents in the police reporting system. It also provided a close link with a large segment of the users of police statistics, the departments themselves, and introduced a strong note of what is practical, an essential in the work of the statistician - reconcilling what is desirable in a statistical system with what is possible.

Fortunately, while we were undertaking the developmental work, a good deal of material of much value was available to us. Canadian police departments supplied the Committee with copies of their record systems, as did many departments in the United States and some in Europe. The Uniform Crime Reporting Section of the Federal Bureau of Investigation also supplied us with material. One document that was particularly useful to our committee, was the report of the Consultant Committee on Uniform Crime Reporting.

As developmental work progressed, the Committee reported its findings to the members of the Association at their Annual Conference. As a result, a bridge was built between the respondent police departments and the central statistical agency. The communication flow over this bridge has improved steadily and was given new impetus in 1960 when several experienced police officers were recruited to the Judicial Section. In 1961 developmental work was completed and the product, the Uniform Crime Reporting Program, was adopted at the Annual Conference of the Canadian Association of Chiefs of Police.

Respondents in the Uniform Crime Reporting Program include Municipal police, Royal Canadian Mounted Police, both as Federal police and as Provincial police in eight provinces, Ontario Provincial Police, Quebec Provincial Police, the Railway police and the Provincial Fire Marshals. There are 937 separate respondent sources, counting municipal police departments, districts, subdivisions and detachments of federal and provincial police and the fire marshals as separate reporting units. The present cut-off point for Municipal police departments is set at communities with 750 population or over, maintaining either their own municipal department, or contracting for one with the RCMP, or the Ontario Provincial Police. To be a respondent the department must also have the responsibility for enforcing, in their jurisdiction, the Criminal Code of Canada, which unlike the United States, is uniform throughout Canada, other Federal Statutes, Provincial Statutes and Municipal By-laws. Consideration is being given to raising the lower population level of communities to 2,000 or 2,500 population.

The Uniform Crime Reporting System provides data which gives the user a statistical picture of police administration, of crimes and of traffic enforcement. Separate schedules are used for collecting this information.

In the first schedule, which is for police administration data, information is collected on full-time personnel of the respondent departments, classifying police, civilians, trainees and other full-time employees separately. Information is obtained on actual strength, authorized strength, engagements, retirements and other separations from service. Data are also obtained on police transport - automobiles, motorcycles and other motor vehicles, boats, aircraft and horses.

A separate form is used for collecting statistics on crimes. Information is obtained on the number of crimes reported or known to the police, the number unfounded following police investigation and on the actual number of offences. Everyone agrees that offences known or reported to the police are only a sample of crimes that are committed, that the universe of crimes committed is unknown, that the proportion that are unknown is not known. We are aware that the unknown proportion varies, depending on the type of crime, and it has been indicated that the intensity of police activity has an effect, in part, on the relationship between known crime and total crime. From a practical point of view there must be a beginning at some point; this is the point we have chosen. It provides users with a measure of crimes and traffic offences known or reported to the police.

The system requires that offences cleared be reported either as cleared by charge or cleared otherwise. To illustrate this point, I would like to refer to the Uniform Crime Reporting Manual, prepared for the use of police respondents, in which precise definitions and clear instructions for reporting data are given. Reporting procedures for a variety of situations have been illustrated. Time does not permit me to go through this Manual in detail with you but I would like to draw your attention to one or two points which will serve to illustrate some of the units of count and the system generally.

I mentioned that offences cleared may be reported as cleared by charge or cleared otherwise. An offence is cleared by charge when an Information is laid against at least one person. This includes arrests, summonses to appear and warrants to apprehend. Offences and not arrests are counted. This instruction is not affected by any subsequent acquittal, dismissal or withdrawal. An offence may be cleared by charge if any other charge is laid in connection with the same offence. If several persons commit a crime and only one is arrested and charged, then the crime is to be reported as cleared by charge. When the other offenders are charged, respondents are required not to list a clearance by charge a second time for the same offence.

Respondents may report offences as cleared otherwise if three questions can be answered in the affirmative: (1) has the offender been identified; (2) is there enough evidence to support the laying of an information; (3) is there a reason outside of respondent's control that prevents them from laying an information and prosecuting the offender. The limitations of cleared otherwise are indicated by ten examples. The clearance of offences provides the user with a measurement of the extent to which crimes and traffic offences known to the police have been dealt with by these agencies.

Data are collected on the number of persons charged, such persons being classified as either adult or juveniles, male or female. If police statistics are to be integrated with court statistics and ultimately with those of correctional facilities, it is essential that a change over occur from an accounting of offences to an accounting of persons. This is a particularly useful measure for our integration program, which I shall refer to briefly at the end of this paper.

Data on offences known, unfounded, actual crimes, clearances and on persons charged are obtained for certain specific offences under the Criminal Code of Canada, which is a federal document administered by the provinces. The fact that in Canada we have one Criminal Code means that many of the problems of uniformity have been resolved. Data are collected on murder, attempted murder, manslaughter, rape, other sexual offences, wounding, assaults (other than indecent assaults which are included under other sexual offences), robbery, breaking and entering, theft of motor vehicle, theft over $50, theft $50 and under, possession of stolen goods, frauds, prostitution, gaming and betting and offensive weapons. A residual category for other offences under the Criminal Code is provided. This residual category excludes traffic offences under the Criminal Code, which information is collected separately. Information is also collected on offences against Federal Statutes, Provincial Statutes and Municipal By-laws, excluding motor vehicle traffic and narcotic offences.

Statistics on Narcotic Offences are collected separately by the Royal Canadian Mounted Police on forms supplied by the Bureau. Fire crime statistics are also collected separately on forms supplied by the Bureau by the provincial fire commissioners in co-operation with the police.

Referring again to the manual, there are twelve general rules of scoring offences. Offences against a person are scored differently from offences against property. In the case of offences against the person one offence is counted for each victim, while in offences against property the number of distinct or separate operations are counted. A distinct operation means the same time, location and circumstance of the offence. Specific instructions for each offence are also given in the manual.

The third statistical return is used for the collection of traffic enforcement statistics. Data are collected on traffic offences under the Criminal Code which are criminal negligence causing death, criminal negligence causing bodily harm, criminal negligence in the operation of a motor vehicle, failure to stop at the scene of an accident, dangerous driving, driving while intoxicated, driving while impaired and driving while disqualified. Traffic offence data are also obtained in separate categories for Federal Statutes, Provincial Statutes and Municipal By-laws from which have been excluded parking violations. The latter are reported separately.

The number of fatal traffic accidents are reported and, as well, the number of non-fatal (injury) accidents, property damage accidents over $100 and those with damages $100 or less. Information is obtained on the number of persons killed in these accidents, whether drivers or passengers, pedestrians, cyclists, or others. The number of persons injured is also reported.

There are a number of other items we collect such as motor vehicles stolen and recovered, missing persons, and policemen killed on duty, whether accidentally or by criminal action.

The Uniform Crime Reporting System was implemented on January 1, 1962. Much of the field work has been done during the first year of reporting under the new system. Field work has been facilitated through the organization of Uniform Crime Reporting Committees in Provincial and Regional Associations of Chiefs of Police and, in some areas, these organizations have appointed zone liaison officers for the program. These are police officers drawn from association membership who work quite closely with our staff at the Bureau on reporting problems. Generally, the Provincial and Regional Associations have followed the lead of the National Association in working closely with the Bureau in the interests of the program.

The Uniform Crime Reporting Manual has been supplied to all police respondents and is now in its third reprinting. It has been used extensively in seminars conducted by DBS staff members in different parts of the country. Some of these seminars have been organized by provincial police associations as a part of their staff training programs held at universities. Seminars and field visits are conducted by the Bureau staff members who have had police experience, which enables them to interpret the program to the respondents in a common language. This communications bridge is not a one-way street however, and it is to the Bureau's advantage that the reporting problems of police respondents can be interpreted to the Bureau.

We are now at the point of beginning to see the tangible results of our developmental work. Three publications will result from the Uniform Crime Reporting Program - Police Administration Statistics, Crime Statistics and Traffic Enforcement Statistics, each of which will be an annual statistical report.

In these reports data will be available for Canada as a whole, for each province, for twelve selected municipal police metropolitan areas, and for communities having municipal police departments. Community data will be classified according to population size. In addition, data will be summarized for each responding department and will appear in a detailed list at the end of each report.

I have given you a brief overview of the history, development, implementation, field organization and co-ordination of the Uniform Crime Reporting System in Canada, one of several programs for which the Judicial Section has responsibility. I would like now to turn very briefly to another area.

In our work, we are attempting to integrate statistical series in the law enforcement, judicial and correction fields. The goal of integration has been described in a number of ways: one is that the individual statistical series is of limited interest in itself. Almost invariably users of statistical data want to employ different statistical series jointly. Reliable individual series may be produced in a number of different ways with different definitions and concepts chosen with equal validity as the primary foundation. Recognition of the possible joint use of statistics means, therefore, that all relevant series must be designed from the beginning to facilitate joint use. In this view the most useful statistics require that the whole statistical output should be conceived as an integrated statistical system. The application of integration to criminal statistics has been stated by a number of individuals, one of whom is Mr. Ronald H. Beattie. He indicates that ideally information should be available which would show for every 1,000 persons arrested, individual characteristics, the time consumed and the disposition made of each person in each step of the legal process involving police, courts and the various correctional facilities. All of such detailed information has to be gathered from separate agencies and put together in order to obtain a reliable and integrated description of the administration of criminal justice.

An advantage to the integrated approach, of course, is that segments of the integrated system can be made the subject of more detailed study. The integrated frame work provides the bench marks against which the results of these studies can be assessed. Such is the case in a survey which we are now writing up on auto thefts in Canada, the results of which will be ready for release in a few days time.

The pre-requisite to integrated data are conceptual integration, operational integration and controlled reporting. All three are equally important. In order to link individual statistical returns, positive identification of each individual return is required. The only positive means of identification is the fingerprint method. In order to relate individual statistical series, standard classifications are required. We are concentrating on a standard offence classification, a standard geographical classification and a standard age-sex classification.

We are conducting a study in which data from the various sources I have mentioned are being integrated. This study, which is concerned with murder, will be useful in itself. Also important, the study will lay the ground work and serve as a model for broader application in our work.

II

MATCHING OF MEDICAL, SOCIAL AND ECONOMIC RECORDS FOR RESEARCH PURPOSES

Chairman, O. K. Sagen,
National Center for Health Statistics

RESEARCH NEEDS FOR RECORD MATCHING

Sam Shapiro, Health Insurance Plan of Greater New York
and
Paul M. Densen, New York City Department of Health

The value for research of bringing together information on two or more unrelated records for an individual or family has long been recognized. Indeed, the present article could just as readily have been prepared 10 years ago as today. In fact a proposal for establishing a permanent record matching system on a national scale was advanced by Dunn, 15-20 years ago.(1) But this proved to be too comprehensive and complex and fell by the wayside.

With the passage of time, record matching has been resorted to with increasing frequency and it is likely that the use of this procedure will continue to expand. It is the purpose of this paper to examine some of the substantive and methodological circumstances that have made "record matching" a vital issue for many research interests. Illustrations are drawn primarily from research that has been completed or is in progress.

Definition of record matching and record systems under consideration

For the current discussion, record matching refers to the process of collating the records on specific individuals from two or more sets of records collected through independent means or for different reasons. The objective of this procedure is to combine the information on these records for research purposes. Usually at least one of the files of records being searched is large, relates to the total population or a defined segment of it, and is the result of a regular on-going collection process. Such a file is ordinarily designed to last for a long time.

There are today several networks of records of this type. Prominent among them are the Decennial Census with its broad range of sociological, economic and housing information regarding a sample of the population; the vital registration system with its basic facts of birth, death, marriage and divorce; the Old Age and Survivors Insurance records which now cover well over 90 per cent of the gainfully employed and will eventually provide a history of the working life of individuals employed in particular industries and geographic areas; the Internal Revenue Service's income tax returns; and the health insurance companies' claims and medical records which contain data on hospital episodes involving at least 75 per cent of the population and information about medical care outside the hospital for an expanding sector of the population. To these could be added the networks of life insurance, welfare, hospital, and educational records and the myriad of specialized sets of records that may exist in a particular community.

Substantive needs for record matching

In a sense, the growth, organization and strengthening of these vast repositories of records is a reflection of the increasing complexity of our society and of the economic, social and health problems it faces. The record systems are of course repeatedly being treated as closed, self-contained sources of information and mined for data relevant to outstanding problems of the day. But for some of the most pressing questions, this approach is inadequate. The reasons are quite apparent. The range of observations collected through a single record system is restricted and frequently the issue under study concerns events for which the system was not designed to collect information in the first place. Many of these problems or limitations are, however, overcome when data in two or more existing record systems can be made to supplement each other or when a routinely collected set of records is used as an adjunct to records generated for specific research purposes. These two types of situations are discussed more fully below in relation to the issues that led to the use of a record-matching procedure.

Studies dependent on multiple sets of established record systems

In the health field extensive use for research purposes has been made of two or more sets of established records. Many health departments have for years been matching live birth and infant death records to enlarge knowledge about risk factors affecting mortality in infancy. This activity is now so well established in many states that it is not always appreciated that it was undertaken in response to a major change in the nature of the infant mortality problem. As long as infectious diseases were the primary causes of death among the newborn and progress was being made in reducing the infant mortality rate, requirements for data could in large measure be satisfied from the items on the death record. These included race, sex, geographic area, cause and age at death.

As postnatal environmental conditions receded in importance, pressures developed for learning more about the influence of prenatal and natal circumstances on infant mortality. Despite the fact that the vital records are far from perfect instruments for comprehensive inquiries into these matters, it was recognized that through the mechanism of matched records a broad extension of relevant information could be quickly and economically accomplished. To the usual set of variables on the death records, it became possible through matching to add items on operative procedures at delivery, characteristics of hospital where the birth occurred, birth weight, ages of parents, birth order, etc., from the birth record for new types of investigations.(2)

One of the most intriguing exploitations of information in completely unrelated record systems has been carried out by Mancuso and Coulter. Their general interest was in determining "how various cohort industrial population groups, differing primarily in the exposure of the working environment in the same and different geographic areas, vary among each other in morbidity and mortality patterns".(3) This was pursued to good effect in a study of mortality risks among employees in an asbestos plant. The investigators used the BOASI records to establish a cohort of persons working in a particular asbestos manufacturing company during 1938-1939. This cohort was then followed to determine subsequent mortality. Death claims for benefits related to the cohort were used as a basis for locating death certificates on file in vital statistics offices. Death record information was supplemented where possible by reports of microscopic findings and then combined with the BOASI record data for analytical purposes. An association was found between employment in the asbestos industry and mortality due to asbestoses and cancer of the lung and of the peritoneum.

Mancuso and Coulter concluded from this and other experiences that through the merging of data on existing health, insurance, employment and earnings records, cohort studies of great importance for industrial health programs would become feasible. The link that they saw was the Social Security number. It should be noted that many of the same records are pertinent to other areas of interest. Occupational mobility and changes in economic status at various stages of an individual's working life are of considerable significance to behavioral scientists concerned with, for example, parameters of family formation and stability, changes in social class structure, and emotional disorders.

Another study of particular interest is that being conducted by Hauser and Kitagawa using the enumeration schedules filled out in the 1960 Decennial Census and the copies of certificates of deaths that occurred during the 4 month period May-August 1960.(4) The desired end result is the location of the appropriate set of Census data for each of the persons who died and the merging of all of the pertinent facts of death with the sociological and economic data collected through the census. This is unquestionably the most formidable record matching operation ever undertaken in this country, involving as it does records for the 179 million persons enumerated and for a half million deaths.

It is not the purpose here to dwell on the matching problems--that is the prerogative of those attempting to solve them--but rather to indicate that this difficult activity is aimed at relieving some of the basic shortcomings of the death record. It is no longer possible to satisfy the needs for mortality information through the limited range of parameters available on this record. However, the linking of the death record with the Census record should open the door to the exploration of the relationship between income, education, occupation, housing and other environmental factors, and death due to specific chronic and acute conditions.

Additional uses made of established record systems to satisfy research needs come to mind. Lew and associates investigated mortality risks associated with deviations from average weight and with high blood pressure.(5) Medical observations made at the time individuals applied for life insurance were linked to mortality records over a long period of time to arrive at relative risk factors which are of great utility not only for underwriting purposes but also for the current efforts to understand for example the relevance of weight and blood pressure to mortality due to coronary heart disease.

Also, Newcombe and co-workers developed a project to link routinely collected population records from which information could be obtained on differential fertility and mortality in families carrying hereditary defects.(6) The basic step in the program consisted of linking birth registrations (about 400,000 in the years 1946-1958) with the marriage registration of parents (114,000 for the same period) in the Canadian province of British Columbia. The unique feature of this study is that Newcombe applied computer technology to a large scale record matching problem.

Finally, through the linking of marriage, birth and divorce records, Christenson has probed into factors associated with child spacing, premarital pregnancy and marital dissolution.(7) The starting point was the file of records of marriages that occurred during a particular period in three local areas. These were then matched against birth record files covering varying lengths of time after the marriage. In two of the three communities the search also included divorce records. The data derived from these three sets of records proved to be useful for a number of speculations regarding, for example, the relationship of sexual permissiveness in a culture to premarital pregnancy and the effects of the pregnancy on hasty marriage and on subsequent divorce.

<u>Studies dependent on special research records linked to established record systems</u>

Even if it were possible to combine all existing record systems, there would be a large deficit in information required for most studies. This is inevitable. Routine records can not be expected to answer all questions for all time even within the field where they are located. Other methods for obtaining data are of course more appropriate for certain types of research and are being used extensively, e.g., personal interview, mail survey, special medical examinations.

Without flexibility in the approach to problems research would quickly dry up. However, an existing record system frequently serves as the most reliable and convenient source of an essential part of the desired information. For example, many long term population studies which start with biological, physical, behavioral or social observations obtained through special inquiries are concerned with the relationship between mortality and these initial characterizations. Persons who die must be identified and facts about their death retrieved from the appropriate death certificates. Matching with death records may involve only those persons known through other sources to have died or with the entire cohort in the absence of such knowledge.

Dorn, Hammond and others integrated smoking histories obtained through a mail survey or personal interview with information on death certificates for persons in their cohorts who died in succeeding years. Through follow-up procedures in the field Hammond determined periodically who was living and who had died.(8) The search for matching death records was then initiated. Despite the fact that he was dealing with a large cohort (188,000 men), Hammond was able to keep track of the survival status of the overwhelming majority in his cohort and to locate the death record for every man reported as having died (11,870 deaths in a 44 month period). As is well known, the information was then used to examine the relationship between smoking and lung cancer and other causes of death. In this study the investigators had many of the specific details needed to search for death records.

Not all studies concerned with mortality are so favorably disposed and the effort required to identify deaths through manual or punched card matching procedures is so great sometimes that it deters much needed research. However, present experiments in matching records through computer techniques will, hopefully, change the situation and improve the prospects for conducting long-term follow-up studies related to health. Reference has already been made to Newcombe's efforts and the problem that led to his use of computers for record matching purposes. Two other studies are cited below as illustrations of the opportunities for research that will occur when computer matching techniques are refined.

In both, the matching operations are being carried out by the Health Insurance Plan of Greater New York in cooperation with the New York City Department of Health. One of the studies is designed to measure the incidence and prognosis of coronary heart disease in a population of 120,000-130,000 persons aged 25-64 years.(9) Incidence is to be determined through a combination of medical and hospital data and the findings of special medical examinations. New cases are being located during a 5 year period; and the course of the disease is being studied also over a 5 year period. The follow-up of patients admitted to the prognosis cohort presents no special problem. There will be about 2,000 such cases and there will be frequent communication with each of these patients or next of kin in case of death. Mortality can therefore readily be determined and death records obtained.

However, death from coronary disease can occur suddenly among patients not in the prognosis cohort. This circumstance makes it necessary to place the entire population under continuous surveillance for mortality. It is not known to what degree deaths among H.I.P. members are reported as such to the Plan. For this reason, a test is being made which involves matching computer tapes that contain identifying information from the H.I.P. records against similar tapes prepared from records of all deaths among adults that occur in New York City during the study years.

In the other study, two matched samples of 30,000 women each (aged 40-64 years) are being followed for about 10 years to determine differences in mortality from breast cancer.(10) One of the samples is being asked to come in for periodic clinical and x-ray examinations through which it is expected that breast cancer will be diagnosed at a comparatively early stage. The other sample of women will follow their usual practices in seeking medical care. The need for a more efficient method in locating deaths in both cohorts than direct communication with them or their families is critical and the plan is to rely heavily on the same type of computer matching procedures being tested for the Coronary Heart Disease study.

The projects discussed involve large study populations under observation for long periods and require a repetition of the matching process at intervals. There are, however, many other cases in which established record systems are entered only once to amplify the information collected through special studies. Furthermore, the study population is quite small. Anderson, Feldman and Sheatsley supplemented data on hospital and medical utilization and costs obtained through household interviews by obtaining expanded and more specific information on these items from the matched records in the files of health insurance plans.(11) Bright, Lilienfeld and others are following up on persons who were covered in the Hunterdon and Baltimore chronic disease studies about 10 years ago to determine the relationship between morbidity findings on both interview and special medical examinations conducted at the time and subsequent mortality. (12) Death records have an important function in this investigation and must be located.

Methodological reasons for record matching

Record matching is often carried out in response not only to a need for otherwise unavailable data but also because of the opportunity it offers for obtaining improved information. This is true for many of the projects already discussed. In the case of the Anderson, Feldman and Sheatsley studies much of the information obtained through record matching was already covered in the household surveys. But, through recourse to records they were in a position to correct response errors as well as expand their range of information.

The Hauser-Kitagawa study based on matched census and death records has as one of its major objectives the improvement of mortality rates resulting from the availability of comparable numerators and denominators. Ordinarily rates by age, race, marital status, residence and occupation are computed by using for the numerators information on the death certificate and for the denominators corresponding items on the census schedule. It is well known that for several of these variables major differences exist between the two record sources. Use of the same document (the Census record) does not eliminate all sources of error by any means but it deals effectively with the problem of incomparability.

One additional reason for record matching to be taken up here is related entirely to methodological considerations. Studies have been carried out by means of matched records whose sole purpose has been to measure bias or errors in information obtained through a particular data collection procedure and to find ways of improving the accuracy and completeness of the information.

The National Health Survey has made such methodological studies an integral part of its on-going statistical program. During the past few years, the household survey approach of the National Health Survey has been examined intensively for strengths and weaknesses.(13) Chronic diseases reported in household interviews have been checked against medical record information to determine the completeness of such reports. Similarly indices of accuracy of reporting hospital episodes on interview or in a self enumerated survey were obtained through comparisons with hospital records. In all of these studies, the objective is to learn a great deal more than is now known not only about validity of the information being collected but also what the correlates of poor reporting are. The technique has wide applicability for the testing of other data collecting mechanisms, e.g., mail or telephone surveys, diaries.

Ultimately these methodological studies will clarify the types of data that can be collected by a particular procedure with sufficient accuracy to make them usable. This is a weighty issue for all investigators. From the standpoint of the present paper the important point is that the validation procedure is dependent on the correlation of information in two different sets of records.

The 1940 and 1950 Birth Registration Completeness Tests represent record matching on a "grand" scale to determine the accuracy of a record collection system. In both tests, special infant cards were filled out by Census enumerators for children born in the early part of the census year; in 1940 about 700,000 cards were prepared and in 1950 over 800,000.(14) Birth records were matched against these infant cards to measure the underregistration of live births, and in 1950 the underenumeration of infants. Matching procedures differed in these two tests. For the 1940 test, the approach taken was to match records manually. But in 1950, it was decided to place main reliance on a mechanical procedure based on punched cards. Through the application of a series of alternative matching criteria followed by visual inspection of paired record data, the 1950 test found the new approach efficient, accurate and highly suited to the rapid production of detailed test results.

Validity studies not involving governmental statistics have also been carried out repeatedly. One of the pioneering studies involving matched records was that of Parry and Crossley.(15) Their concern was with reports of voting; ownership of a home, driver's license, or library card; and donations in a community drive for funds. All related records available in the area were systematically scanned to determine the accuracy of the responses to the survey questions.

Conclusion

The need for record matching has been approached through a review of research that has utilized this procedure. Undoubtedly other cases can be cited in which record matching because of its complexity and cost had to be replaced by other, less satisfactory methods of obtaining information. Also, existing records are not always the most complete or precise source for data required and record matching becomes a poor substitute for other methodologies. However, there is often no alternative to record matching and, in fact, many of the studies previously referred to could not have been undertaken without the use of interrelated records. Furthermore, the ever increasing complexity of our society is resulting in the development of voluminous record keeping systems for large segments of the population and these systems represent a reservoir of information regarding many significant health, economic and social problems.

Record matching will however continue to be a laborious, difficult and sometimes impractical process unless it attracts greater attention both from persons responsible for developing record keeping systems and from statisticians concerned with methodology. By its nature, record matching requires a broad perspective of the function of record systems. Often, persons charged with the responsibility of establishing record systems think only in terms of the most immediate administrative needs that have to be satisfied. Even when the outlook is not so narrow, there is almost no attention given to the content or organization of the records that might facilitate their being matched with other sources of information. This applies to both routinely collected records and special research records.

Today the common thread in almost all records consists of age, sex and address information. This consistency has not occurred because of any interest in record matching and indeed the thread may in some instances be too thin for matching purposes. A prerequisite for a change in the present situation is to lift record matching from its present haphazard state and to consider seriously the basic elements that would facilitate it. This might lead to a convention regarding a minimum set of common items for routine and research records. For example, the addition of social security number which has been suggested by many would provide a potent item of identification for studies involving the adult population.

Not only content needs to be considered but also the organization and storage of record information and efficient techniques for record matching. Some steps have already been taken in this direction. Record systems are increasingly being placed on computer tapes and the application of computer techniques to record matching is being experimented with in several areas, including Canada, California and New York City. These new approaches will hopefully ease the burden of dealing with massive sets of records.

Finally, it must be recognized that we are faced with a set of long term needs for record matching and that the problems will not be resolved by any easy formula. Personnel is needed to concern itself with the elements discussed, i.e., content, organization and technology. The statistician has one of the greatest role in this effort and his training should reflect this responsibility. It is time that the design of record keeping systems for statistical and administrative functions was upgraded to a high priority activity among statisticians and that record matching was accorded the prominent place in the array of methodologies available for research that it merits.

Bibliography

(1) Dunn, H.L., Record Linkage. Amer. Journal of Public Health, 36, 12, Dec. 1946.

(2) National Office of Vital Statistics, Recommendations for Developing Comparable Statistics on Prematurely Born Infants and Neonatal Mortality. Dept. of Health, Education, and Welfare, Dec. 1950.

(3) Mancuso, T.F. and Coulter, E.J., Methodology in Industrial Health Studies. Archives of Environmental Health, 6, Feb. 1963.

(4) Hauser, P.M. and Kitagawa, E., Social and Economic Mortality Differentials in the U.S., 1960: Outline of a Research Project. Social Statistics Section, Amer. Stat. Assn., 1960.

(5) Society of Actuaries, Build and Blood Pressure Study, 1959. Vol. I, Chicago.

(6) Newcombe, H.B. and Rhynas, P.O., Family Linkage of Population Records, Symposium on Medical Electronic Data Processing. Proceedings of the U.N.-W.H.O. Seminar on Use of Vital and Health Statistics for Genetic and Radiation Studies, U.N., 1961.

(7) Christensen, H.T., Cultural Relativism and Premarital Sex Norms. Amer. Sociological Review, 25, Feb. 1960.

(8) Hammond, E.C. and Horn, D., Smoking and Death Rates. Journal of Amer. Med. Assn., 166, Mar. 1958.

(9) Shapiro, S., Balamuth, E., Frank, C.W., Sager, R.V. and Densen, P.M., The H.I.P. Study of Incidence and Prognosis of Coronary Heart Disease: Methodology. In press - Journal of Chronic Diseases.

(10) H.I.P. Annual Statistical Report, 1962. In Press.

(11) Anderson, O.W. and Feldman, J.J., Family Medical Costs & Voluntary Health Insurance: A Nationwide Survey. McGraw-Hill Book Co., 1956.

Anderson, O.W. and Sheatsley, P.B., Comprehensive Medical Insurance. Health Information Foundation Research Series No. 9.

(12) Personal communication. Lilienfeld, A.H.

(13) Health Statistics, U.S. National Health Survey, Health Interview Responses Compared with Medical Records. U.S. Dept. of Health, Education, and Welfare, Series D, No. 5.

Health Statistics, U.S. National Health Survey, Comparison of Hospitalization Reporting. U.S. Dept. of Health, Education, and Welfare, Series D, No. 8.

(14) National Office of Vital Statistics, Vital Statistics of the U.S., 1950, Vol. 1, pp. 108-112, U.S. Dept. of H.E.W.

(15) Parry, H.J. and Crossley H., Validity of Responses to Survey Questions. The Public Opinion Quarterly, 14, 1, Spring 1950.

RECORD MATCHING--THEORY AND PRACTICE
(Abstract)

Philip M. Hauser, University of Chicago and
Patience Lauriat, U. S. Bureau of the Census

After discussion of some of the principles and concepts behind matching of independent records, the paper presents some discussion of the experience of the U. S. Census Bureau in a study sponsored by the University of Chicago. The purpose of the study was to obtain data on the socio-economic characteristics of recent decedents by matching a sample of death certificates, representing deaths occurring during the period May through August 1960, to the Stage I and Stage II 1960 Census records. An overall match rate of about 80% was obtained. The matching was done clerically, using the address of usual residence, primarily, as well as name, color, sex, and age of the decedent. If a person could not be matched at the usual address given on the death certificate, or the usual address could not be allocated to a census enumeration district, the enumeration book for the institution where the death actually took place, if the death occurred in an institution, was searched for a possible name match. In order to obtain data on socio-economic data of decedents under one year of age, a possible match was searched using the mother's address and name as reported on the death certificate. The paper also includes some information about basic demographic characteristics of matched and unmatched decedents.

AN EVALUATION OF SOME MECHANIZED MATCHING OPERATIONS
(Abstract)

Joseph F. Daly, Herman Fasteau and Peter Ohs,
U. S. Bureau of the Census

The Bureau is now preparing to conduct Economic Censuses covering the calendar year 1963. This paper relates some experience bearing on our ability to mechanize three of the matching operations associated with these censuses. All three operations seem simple enough at first sight, but turn out to require a surprising amount of planning and control. One involves the matching of addresses (nearly 4 million in number) on a mailing list against a reference file to assign geographic codes needed for the tabulations. The other two match returned reports against master lists to determine which firms have reported and which firms need further follow-up. Semiautomated techniques for checking in the reports have been tested on the 1962 Annual Survey of Manufactures and on the precanvass of multi-unit firms with encouraging results. And the computer operations as associated with the advance location coding of the mailing list are now far enough along to be evaluated. (Related documents are available upon request to the authors.)

EXPERIENCE WITH COMPUTER MATCHING OF NAMES

William Phillips, Jr. and Anita K. Bahn*
National Institute of Mental Health

For many years, epidemiologists, administrators and biostatisticians have felt the need for a large comprehensive case register for mental illness similar to those in use for cancer and tuberculosis. Such a register can provide unduplicated patient counts for rates of diagnosed incidence and prevalence of mental disorders, data on patient movement between facilities, changes in diagnosis and other longitudinal information not systematically available through any other means. In order to initiate such studies, on July 1, 1961 a psychiatric case register was established for the State of Maryland, in cooperation with the National Institute of Mental Health (1).

When planning for the project, the main problem foreseen was the linkage of records for the individual who receives services in more than one psychiatric facility. Psychiatric services to Maryland residents are provided by over 100 facilities and each facility has its own patient numbering system.

In exploring the possible methods of record linkage, name and address was considered essential. (State legislation now protects the confidentiality of this information and assures its use for research purposes only.) Although social security number has discriminating power and is now requested, it could not be used as the primary matching factor because of the large number of child patients without a social security number, the resistance of patients to furnishing this information, and the lack of social security numbers on reports already on file as part of the ongoing State reporting system.

A second major consideration was the eventual large size of the register due to the relatively few deaths in this patient population. For example, within five years we expect a cumulative file of 120,000 psychiatric experiences representing 60,000 to 80,000 different persons. Primary clerical matching of names, although feasible for the first year or two, would soon become unwieldy and unreliable. The anticipated size of the register warranted the use of computer methods for person matching.

Probability factors for computer decision based on the frequency of names and other characteristics of our unique population were not available pending detailed population analysis and large scale matching experience (2). We planned, therefore, to establish computer methods which would be largely "trial and error" and would attempt to duplicate clerical processes and judgment in checking on a number of identifying factors simultaneously in deciding on the match of a pair of records (3).

Initial rules for computer decision as to whether a match is positive, possible, or rejected were established on the basis of reasonableness, some clerical matching experience, and our general knowledge of the reliability of the reported data. It was decided that clerical review would be made initially of all "positive" and "possible" computer matches, not only to insure against false linkages and for resolution of possible matches, but also to review the results so that our methods could be improved.

Only those computer programs which are essential to the patient matching and record linkage will be described. Our first matching operation involved a comparison of reports of patients on the rolls of psychiatric facilities on July 1, 1961, in order to establish the psychiatric case register. We will later describe the operation of checking new reports with the established register to determine whether the record is for a previous registrant or represents an accretion to the master file.

Establishment of the register file

A total of 22,869 cases were enrolled in the various facilities on July 1, 1961. Our first step was the automatic assignment of a temporary or pseudo register number to each record in order to facilitate record retrieval, correction, and linkage. Beginning with the pseudo-number 5, numbers were assigned by an arithmetic progression of tens. During matching, if two records were considered to be a "positive" match, the pseudo-number of the second listed record was replaced by the pseudo-number, minus one, of the first record. If a second match was detected, the pseudo-number for this record was also replaced, but with the first pseudo-number less two. This process provided for the identification of ten matches for the same individual. (The maximum number turned out to be three.) At a later date, all records were sorted using this pseudo register number which resulted in all data records for the same individual falling in sequence. We were then able to combine all the data for each "individual" and assign a single permanent register number to each record. (All records for each individual are maintained in the master identity file as part of the permanent record and used in subsequent matching and updating processes.)

In making a check for duplication, the ideal method would be to check each record against all other records on all common factors of identity.

* Mr. William Phillips, Jr. is Digital Computer Programmer and Dr. Anita K. Bahn is Chief, Outpatient Studies Section, Biometrics Branch, National Institute of Mental Health.

This, of course, is not feasible with current equipment. It would mean a maximum of 261 million comparisons or $\left(\frac{22869^2 - 22869}{2}\right)$ of one record against another and possibly 6,000 hours of computer time (8.5 milliseconds × number of comparisons). Therefore, we had to group records into blocks of a size which can be handled by our computer.

We chose to group the records by a phonetic code, commonly referred to as the Russell Soundex code. This is a system whereby the consonant sounds of the surname are assigned numbers which are accumulated into a code designed to compensate for the common misspellings. For example, Brown, Browne and Braun all have the same code (1650). In usual practice, the first letter of the surname comprises part of the code and is not assigned a number. We varied this system by coding the first letter as well as the remaining letters so that names such as Cohn and Kohn would be compared (see Figure 1). The codes are all four digits in length.

The Soundex code was assigned by a computer program. For married women with maiden names, the record was reported differing only in the Soundex code assigned.

By use of the four digital Soundex code the file was divided into 1,007 different Soundex groups, of which 31 had more than 150 records. The largest group was "2520" with 519 records. Although such different names as Jones, James, and King, for example, were included in this group, our concordance rules about correspondence of letters in the name eliminated such obvious mismatches (see Figure 2).

We read all of the records of each group into core memory of the computer and cross-checked every record within the group. The first check of this program compares surname, address, first name, and birth year (see Figure 2). The tolerance rules established for concordance of these factors in a positive match were considered conservative. If any field was missing, the comparison of these fields was considered not in concordance. Whenever a match was accepted as "positive," further checking of the record was discontinued. If agreement within the specified tolerances was not achieved, but the match was considered as possible, a second check was automatically made on the basis of social security number and maiden name, factors which can aid in positive identification. The final group of factors or third automatic check, if agreement was still in doubt, consisted of sex, race, and complete birth date.

We made a total of approximately 1,700,000 comparisons of one record against the other. Total computer running time was four hours. At the completion of this program, 627 "positive" and 1,011 "possible" matches were presented in list form for clerical scrutiny. The print record contained the complete name, address, sex, race, birth date, facility code, patient case number, and, in addition, a pseudo register number.

Two clerks spent two days each examining these matches. These clerks were primarily key punch operators who were familiar with the codes but not with the matching program. They determined that 553 "positive" matches were truly positive, 169 "possible" matches were true linkages, and 736 "possible" matches were not true linkages. After this preliminary check, another clerk, who had a thorough knowledge of the logic of the program, spent two weeks checking the doubtful items in detail, referring back to case records or querying facilities where necessary. The final count of positive and false linkages by decision rule is summarized in Figure 3A. A total of four percent of the records or 604 of the positive and 201 of the possible matches were classified as "duplicates."

The clerks then prepared a final list of records requiring change. This included, in addition to discrepant information, corrections to the pseudo register number for "positive" matches determined to be non-matches and for "possible" matches determined to be "positive." An additional clerk day was required to punch, verify, and review the 298 cards used to make adjustments. The program to correct the file of 26,051 records (including 3,182 maiden name records) ran approximately ten minutes.

In reviewing the efficiency of this first unduplication program, it should be noted that few social security numbers were available for assistance in checking at the second stage of the program, and, actually, all social security number comparisons came out as unequal. Also, the month and day of birth were missing in a large percentage of our cases, which caused many linkages to be listed as possible instead of positive. Clerical determination that these cases were positive was aided by the fact that most of these patients were on the books of both the State mental hospital (i.e. on convalescent leave) and the clinic attached to that hospital (see Figure 3B).

Since the establishment of the register, we have been successful in obtaining many of the missing birth dates. This will make it possible to conduct a second primary grouping of the master file on the basis of birth date in order to associate "duplicate" records whose Soundex codes are dissimilar. This month and day of birth check has already been used in the updating programs which will be described next.

Updating the master file (first year's experience)

In updating the register master files with fiscal 1962 data, we began with 22,323 admissions to psychiatric care. First we extracted from the admission (detail) records only the information necessary for our linkage checks. Pseudo register numbers were again assigned to each record for later association, and in addition we added another field to the linkage search record for inserting a located permanent register number (see Figure 4).

A preliminary matching program by facility code and case number searched for readmitted registrants who could be identified by their unit case number.

A person-matching program was used to link the remaining admissions with the master file (see Figure 5). We learned from our previous check that the addition of sex as a grouping factor would not impair the efficiency of our program and would shorten the computer running time. In addition, some changes in final decision rules were made (from "reject" to "possible") to permit clerical verification of our logic.

We read into computer memory core all of the records from the master file for the same Soundex-sex group. We then read the detail records for the same group into core memory, one by one, and compared each record to every record in the master group. When a match which we considered to be a "positive" linkage was detected, we extracted the register number from the master record and rewrote the detail record on a located register number file. We also prepared a print record showing all information for both the detail and master records for later clerical verification. Further computer checking of these records was discontinued.

When the computer program detected a possible linkage, a print record was prepared showing all details and computer checking was continued. As with our program to establish the register, there could be several "possible" linkages for each detail record. If no "positive" linkage was detected, the detail record was rewritten on an unlocated register number tape file to be processed through the next program.

In the first person-matching program used for the updating process, the computer made 1,501,690 record comparisons and classified 2,685 records as positive linkages and 3,219 records as possible linkages (see Figures 6 and 7). Each set of print records (master and detail) were examined clerically. We changed a total of 317 computer decisions (29 "positive" to "reject" and 288 "possible" to "accept") (Figures 6 and 7). Twelve of the 29 positive changes appear to be twins (A0630 category). There were 2,944 net linkages from this program. This figure is inflated slightly because of a number of duplicate linkages based on both the maiden name and the married name for the same person. This duplication will not interfere with our processing as we had planned for this eventuality. The actual number of person-linkages was 2,850.

Our analysis has pointed out several desirable refinements to this program. Two of the major modifications are the discontinuance of print records for the A0000 category whenever we have agreement on month and day of birth or social security number and the changing of A0232 to a positive linkage. These changes will reduce the amount of clerical work.

Clerical scrutiny of the listings required 30.5 hours of clerical time. In addition, four clerical days were spent in checking further into the 112 linkages where a decision could not be made from the listing, and six hours were spent in punching, verifying, and reviewing the 317 cards to correct the "located" and "unlocated" tape files.

The final rejects from this first updating program were then processed through a second program in which month of birth, day of birth, and sex were used as primary grouping factors (see Figure 8). These factors were chosen in order to check cases where the names were entirely different or so misspelled that the Soundex code did not permit them to be compared. The same general processing principles as in the previous program were used (i.e. each detail record was checked to every master record of the same month of birth, day of birth, and sex).

The adjusted unlocated register number file from the Soundex check was used as input to this program along with the master identity file. The computer listed 300 matches of which 219 were classified as "positive" and 81 as "possible" (see Figure 8). Running time for the program was one hour and 20 minutes. Again, we checked all linkages clerically and made adjustments to both the located and unlocated files. This required 6.5 clerk hours.

The yield of this program was very meager. We had a net of only 85 linkages. A few of these linkages were unique because of the fact that we had no name on the admission record (about one percent of the records have no name) and others had completely different surnames. There are several factors which could account for this small yield: the large number (2,000) of cases with month and day of birth still missing and the previous detection of most linkages by the Soundex program. This month and day of birth check may become more valuable in time when there are more possibilities of name change due to adoption of children or remarriage where maiden name is not reported.

After checking the input against the master file for a previously assigned register number, we had remaining a file of approximately 19,000 admission records which were not linked. These records represent new admissions to psychiatric service since inception of the register. As there are undoubtedly duplications within this file, we are planning to use person-matching programs similar to the Soundex check for persons on the rolls July 1, 1961 plus the month and day of birth check. These programs will include the further refinements pointed out by the 1962 updating programs.

Discussion

We envision a continually expanding case register which may ultimately contain over a quarter of a million names. At present, register maintenance and updating require the matching each year of over 20,000 additional records and 30,000 resident death certificates. It is expected that the number of records to be matched annually will increase due to the opening of new facilities. The development of computer programs for primary person matching seems warranted, therefore.

A series of rules for this operation has been established on judgmental grounds. A record is kept of the outcome of each decision rule so that its yield of positive and false matches can be determined and the rule modified by experience. Because of the large amount of missing data in several fields, we will have to distinguish missing information and other unequal comparisons. The large number of linkages with disagreement in address reflects in part differences in the punching of addresses. We have standardized this punching as much as possible, but abbreviations used in the past are still causing difficulties. It is also our intention to revise our programs in the future to decrease the number of print records and thereby reduce the amount of clerical work.

Our first computer matching operation of 23,000 records with each other to establish the register required 4.2 computer hours and 120 clerical hours for residual matching. Based upon the experience of another register one-fourth the size, it is estimated that approximately six man months of clerical time would have been required if the matching operation was entirely clerical, that is, for a clerk to review each record against the name of other records in an alphabetical listing.

In the linkage of new admissions to our file, we benefited from our first duplication check. The computer program for the linkage of records by Soundex grouping ran for three hours and 23 minutes and 68.5 hours of clerical time were required. For the check by month and day of birth grouping, one hour and 20 minutes of computer time was required, and clerical personnel spent 6.5 hours in scrutinizing the linkage and preparing correction punch cards.

Although the programing and other costs of making these checks are high in proportion to the number of cases, our costs will decrease over the years with refinements to our programs and improvement in the accuracy of our data. From the results of sample studies based upon clerical review of the alphabetical master file, we estimate that we are missing between .3 and .5 percent of the linkages. We are planning further sample studies to improve our computer methods for linking these misses. A primary sort on the basis of address is also planned.

In addition to routine statistical checks of each master record for internal consistency of data, there will be a continuous review of sample files to detect false linkages.

As soon as we have completed the updating of our master files with fiscal 1962 data, we are planning a series of "death clearance" programs. We have obtained a duplicate set of punch cards for all Maryland deaths during fiscal 1962. It is our intention to run these records through the same programs used to update our register and obtain the cause and date of death for any of our patients who have died.

There is still much work and experimentation to be undertaken to improve the efficiency of the maintenance of such a file by an electronic computer. Of course, additional identifying information on patients such as birth place and mother's maiden name would increase the discriminatory power of the computer. Such information cannot now be requested. However, we believe that the greater accuracy, consistency, and efficiency of our present computer program as compared with clerical operation for a register this size justify continued experimentation. Furthermore, these methods permit the matching of large rosters of individuals obtained from other sources, such as welfare and criminal records, to our master file.

We believe our methods to be generally applicable to other types of person-matching operations.

References

(1) Phillips, William, Jr., Gorwitz, Kurt, and Bahn, Anita K.: Electronic Maintenance of Case Registers. Pub. Health Rep. 77: 503-510, June 1962.

(2) Newcombe, H. B., Kennedy, J. M., Axford, S. J., James, A. P.: Automatic Linkage of Vital Records. Science 130: October 16, 1959.

(3) Phillips, William, Jr., Bahn, Anita K., Miyasaki, Mabel: Person-Matching by Electronic Methods. Communications of the ACM 5: 404-407, July 1962.

FIGURE 1

Phonetic Code Used in Maintaining Register

The consonants of the surname are assigned numbers according to the following schedule and rules.

Code	Letters
1	B, F, P, V
2	C, G, J, K, Q, S, X, Z
3	D, T
4	L
5	M, N
6	R
Not coded	A,E,I,O,U,W,H,Y

Rule 1

The code for any name consists of 4 digits. If a name does not have sufficient coded consonants, zeros are added to complete the code. (e.g. Lee: 4000). If there are more than 4 coded consonants, the code is truncated. (e.g. Malinauskas: 5452).

Rule 2

If 2 or more consonants which have the same coded number come together, they are coded as only one letter.

Phillips is coded: 1412

P is coded 1	I is not coded
H & I are not coded	P is coded 1
LL is coded 4	S is coded 2

Dickson is coded: 3250

D is coded 3, I is not coded, C K and S all have the same code value 2, and occuring together they are coded as one letter, O is not coded, N is coded 5, and a zero is added to complete the code.

Rule 3

Consonants having the same code number but separated by one or more vowels (a,e,i,o,u,y) are coded individually.

Diciccio is coded: 3220
Wyman is coded: 5500

Rule 4

W and H do not separate consonants. If two consonants having the same code are separated by a W or H, they are coded as one consonant.

Sachs - 2200

FIGURE 2

Soundex Check for Linkage

July 1, 1961 Population

		FIRST CHECK					SECOND CHECK			THIRD CHECK			
Reference Code	Soundex code	Surname	First name	Addr.	Birth year range	Decision	Soc. Sec.#	Maid. name	Decision	Sex & race	Birth mo. & day	Birth year	Decision
A0000	0	0	0	0	0	Accept							
A0100	0	0	0	0	1	Possible	0	0	Accept				
A0110							0	1	Accept				
A0120							1	0	Accept				
A0130							1	1	Possible	0	0	1	Accept
A0131										0	1	1	Reject
A0132										1	0	1	Possible
A0133										1	1	1	Reject
A0200	0	0	0	1	0	Possible	0	0	Accept				
A0210							0	1	Accept				
A0220							1	0	Accept				
A0230							1	1	Possible	0	0	0	Accept
A0231										0	0	1	Accept
A0232										0	1	0	Possible
A0233										0	1	1	Possible
A0234										1	0	0	Accept
A0235										1	0	1	Possible
A0236										1	1	0	Reject
A0237										1	1	1	Reject
A0300	0	0	0	1	1	Reject							
A0400	0	0	1	0	0	Possible	0	0	Accept				
A0410							0	1	Accept				
A0420							1	0	Accept				
A0430							1	1	Possible	0	0	[illegible]	Accept
A0431										0	0	1	Accept
A0432										0	1	0	Possible
A0433										0	1	1	Reject
A0434										1	0	0	Possible
A0435										1	0	1	Possible
A0436										1	1	0	Reject
A0437										1	1	1	Reject
A0500	0	0	1	0	1	Possible	0	0	Accept				
A0510							0	1	Accept				
A0520							1	0	Accept				
A0530							1	1	Possible	0	0	1	Possible
A0531										0	1	1	Reject
A0532										1	0	1	Possible
A0533										1	1	1	Reject
A0600	0	0	1	1	0	Possible	0	0	Accept				
A0610							0	1	Accept				
A0620							1	0	Accept				
A0630							1	1	Possible	0	0	0	Accept
A0631										0	0	1	Possible
A0632										0	1	0	Possible
A0633										0	1	1	Reject
A0634										1	0	0	Possible
A0635										1	0	1	Possible
A0636										1	1	0	Reject
A0637										1	1	1	Reject

(Cont'd.)

FIGURE 2
Soundex Check for Linkage
(Cont'd.)
July 1, 1961 Population

		FIRST CHECK					SECOND CHECK			THIRD CHECK			
Reference code	Soundex code	Surname	First name	Addr.	Birth year range	Decision	Soc. Sec. #	Maid. name	Decision	Sex & race	Birth mo. & day	Birth year	Decision
A0700	0	0	1	1	1	Reject							
A0800	0	1	0	0	0	Accept							
A0900	0	1	0	0	1	Possible	0	0	Accept				
A0910							0	1	Accept				
A0920							1	0	Accept				
A0930							1	1	Possible	0	0	1	Accept
A0931										0	1	1	Reject
A0932										1	0	1	Possible
A0933										1	1	1	Reject
A1000	0	1	0	1	0	Reject							
A1100	0	1	0	1	1	Reject							
A1200	0	1	1	0	0	Possible	0	0	Accept				
A1210							0	1	Accept				
A1220							1	0	Accept				
A1230							1	1	Possible	0	0	0	Possible
A1231										0	0	1	Possible
A1232										0	1	0	Possible
A1233										0	1	1	Reject
A1234										1	0	0	Possible
A1235										1	0	1	Possible
A1236										1	1	0	Reject
A1237										1	1	1	Reject
A1300	0	1	1	0	1	Reject							
A1400	0	1	1	1	0	Reject							
A1500	0	1	1	1	1	Reject							

NB: 0 indicates agreement; 1 indicates discrepancy between the records

Tolerance Rules for Concordance:

Surname - In a one-to-one correspondence of the first 8 letters, only one disagreement allowed

First name - In a one-to-one correspondence of the first 8 letters, only one disagreement allowed

Address - Agreement on street number and first 8 letters of street name

Birth year range - If current age is: / Range must be within:

If current age is:	Range must be within:
0 - 17	2 years
18 - 29	5 years
30 - 49	10 years
50 and over	15 years

Complete agreement required for social security number, maiden name, sex, race, birth month and day, and birth year

FIGURE 3

Psychiatric Case Register (Maryland)
Book Population July 1, 1961

A. Results of Soundex Duplication Check

Major Decision Group or Reference Code	"Positive" Linkages from Computer			"Possible" Linkages from Computer
	Total	Determined Not Correct	Net Positive Linkages	
A0000	357	0	357	6
A01XX	2	0	2	1
A02XX	220	2	218	2/ 918
A04XX	14	1/ 10	4	2
A05XX	0	0	0	0
A06XX	22	10	12	90
A08XX	11	0	11	0
A09XX	0	0	0	0
A12XX	1	1	0	0
Total	627	23	604	3/ 1,011

1/ These were determined to be twins.

2/ This substantial number of "possible" matches was due to the large number of records with missing month and day of birth.

3/ Of these possible linkages, 201 were determined to be positive. There are no counts by detailed reference code. However, the majority were in the A0232 group.

B. Type of Duplications Detected during Processing

Number of Psychiatric Cases on Rolls - July 1, 1961		22,869
Number of Positive Linkages Detected		805
State Hospital Leave and Clinic Care	615	
On Veterans Administration Hospital and Clinic Books	50	
On the Books of an Inpatient Facility and 2 Clinics	8	
On books of Two Inpatient Facilities	23	
In Private Hospital and on Clinic Books	7	
In State Hospital and on Clinic Books	65	
On Books of Two Clinics	37	
Net Number of Patients on Register		22,064

FIGURE 4

TAPE RECORDS USED IN LINKAGE CHECKS

MASTER RECORD USED IN LINKAGE CHECKS

REGISTER NUMBER	FACILITY CODE	PATIENT CASE NUMBER	CODE	SOUNDEX CODE	SEX	RACE	BIRTH			NAME- LAST - FIRST - MIDDLE
							YR	MO	DY	

MASTER RECORD (con't)

STREET NUMBER	STREET NAME	CITY OR TOWN	ZONE	STATE	MAIDEN NAME	SOCIAL SECURITY NUMBER	STATUS	CODE	CENSUS TRACT

DETAIL RECORD USED IN LINKAGE CHECK

FACILITY CODE	PATIENT CASE NUMBER	CODE	PSEUDO REGISTER NUMBER	LOCATED REGISTER NUMBER	SOUNDEX CODE	SEX	RACE	BIRTH			NAME - LAST - FIRST-
								YR	MO	DY	

DETAIL RECORD (con't)

MIDDLE	STREET NUMBER	STREET NAME	CITY OR TOWN	ZONE	STATE	MAIDEN NAME	SOCIAL SECURITY NUMBER

FIGURE 5

Soundex Check for Linkage

Fiscal 1962 Updating

		FIRST CHECK					SECOND CHECK			THIRD CHECK				
Reference code	Soundex code & sex	Surname	First name	Addr.	Birth year range	Decision	Soc. Sec. #	Maid. name	Decision	Race	Birth mo. & day	Birth year	Decision	COMPUTER COUNTS
A0000	0	0	0	0	0	Accept								00001187
A0100	0	0	0	0	1	Possible	0	0	Accept					None
A0110							0	1	Accept					00000001
A0120							1	0	Accept					00000002
A0130							1	1	Possible	0	0	1	Accept	00000011
A0131										0	1	1	Possible	00000020
A0132										1	0	1	Possible	None
A0133										1	1	1	Possible	None
A0200	0	0	0	1	0	Possible	0	0	Accept					00000008
A0210							0	1	Accept					00000096
A0220							1	0	Accept					00000192
A0230							1	1	Possible	0	0	0	Accept	00000894
A0231										0	0	1	Accept	00000093
A0232										0	1	0	Possible	00000227
A0233										0	1	1	Possible	00000690
A0234										1	0	0	Accept	00000017
A0235										1	0	1	Possible	00000003
A0236										1	1	0	Reject	00000030
A0237										1	1	1	Reject	00000358
A0300	0	0	0	1	1	Reject								00002956
A0400	0	0	1	0	0	Possible	0	0	Accept					00000002
A0410							0	1	Accept					00000003
A0420							1	0	Accept					00000002
A0430							1	1	Possible	0	0	0	Accept	00000045
A0431										0	0	1	Accept	None
A0432										0	1	0	Possible	00000008
A0433										0	1	1	Reject	00000034
A0434										1	0	0	Possible	None
A0435										1	0	1	Possible	None
A0436										1	1	0	Reject	None
A0437										1	1	1	Possible	None
A0500	0	0	1	0	1	Possible	0	0	Accept					None
A0510							0	1	Accept					None
A0520							1	0	Accept					None
A0530							1	1	Possible	0	0	1	Possible	00000001
A0531										0	1	1	Reject	00000092
A0532										1	0	1	Possible	None
A0533										1	1	1	Reject	None
A0600	0	0	1	1	0	Possible	0	0	Accept					None
A0610							0	1	Accept					00000005
A0620							1	0	Accept					00000027
A0630							1	1	Possible	0	0	0	Accept	00000077
A0631										0	0	1	Possible	00000095
A0632										0	1	0	Possible	00002127
A0633										0	1	1	Reject	00032516
A0634										1	0	0	Possible	00000002
A0635										1	0	1	Possible	00000044
A0636										1	1	0	Reject	00001391
A0637										1	1	1	Reject	00022685
A0700	0	0	1	1	1	Reject								00182702
A0800	0	1	0	0	0	Accept								00000023
A0900	0	1	0	0	1	Possible	0	0	Accept					None
A0910							0	1	Accept					None
A0920							1	0	Accept					None
A0930							1	1	Possible	0	0	1	Accept	None
A0931										0	1	1	Reject	None
A0932										1	0	1	Possible	None
A0933										1	1	1	Reject	None
A1000	0	1	0	1	0	Reject								00005283
A1100	0	1	0	1	1	Reject								00013624
A1200	0	1	1	0	0	Possible	0	0	Accept					None
A1210							0	1	Accept					None
A1220							1	0	Accept					None
A1230							1	1	Possible	0	0	0	Possible	00000002
A1231										0	0	1	Possible	None
A1232										0	1	0	Possible	None
A1233										0	1	1	Reject	00000003
A1234										1	0	0	Possible	None
A1235										1	0	1	Possible	None
A1236										1	1	0	Reject	None
A1237										1	1	1	Reject	None
A1300	0	1	1	0	1	Reject								00000015
A1400	0	1	1	1	0	Reject								00312449
A1500	0	1	1	1	1	Reject								00921648
													TOTAL	1,501,690

NB: 0 indicates agreement;
1 indicates discrepancy between the records

Tolerance Rules for Concordance:

Same as Figure 2.

FIGURE 6

"Positive" Decision Results
Soundex Check Program - 1962 Updating

Reference code	Soundex code & Sex	Surname	First name	Addr.	Birth year range	Soc. Sec.#	Maid. name	Race	Birth mo. & day	Birth year	Computer linkages	Clerical rejects	Net linkages	Clerk time hours
A0000	0	0	0	0	0						1187	0	1187	5.00
A0110	0	0	0	0	1	0	1				1	0	1	.02
A0120	0	0	0	0	1	1	0				2	0	2	.02
A0130	0	0	0	0	1	1	1	0	0	1	11	0	11	.25
A0200	0	0	0	1	0	0	0				8	0	8	.25
A0210	0	0	0	1	0	0	1				96	0	96	.35
A0220	0	0	0	1	0	1	0				192	0	192	.25
A0230	0	0	0	1	0	1	1	0	0	0	894	0	894	2.00
A0231	0	0	0	1	0	1	1	0	0	1	93	0	93	.35
A0234	0	0	0	1	0	1	1	1	0	0	17	0	17	.25
A0400	0	0	1	0	0	0	0				2	0	2	.02
A0410	0	0	1	0	0	0	1				3	0	3	.02
A0420	0	0	1	0	0	1	0				2	0	2	.02
A0430	0	0	1	0	0	1	1	0	0	0	45	0	45	.35
A0431	0	0	1	0	0	1	1	0	0	1	-	-	-	-
A0500	0	0	1	0	1	0	0				-	-	-	-
A0510	0	0	1	0	1	0	1				-	-	-	-
A0520	0	0	1	0	1	1	0				-	-	-	-
A0600	0	0	1	1	0	0	0				-	-	-	-
A0610	0	0	1	1	0	0	1				5	0	5	.03
A0620	0	0	1	1	0	1	0				27	15	12	.50
A0630	0	0	1	1	0	1	1	0	0	0	77	14	63	1.25
A0800	0	1	0	0	0						23	0	23	.20
A0900	0	1	0	0	1	0	0				-	-	-	-
A0910	0	1	0	0	1	0	1				-	-	-	-
A0920	0	1	0	0	1	1	0				-	-	-	-
A0930	0	1	0	0	1	1	1	0	0	1	-	-	-	-
A1200	0	1	1	0	0	0	0				-	-	-	-
A1210	0	1	1	0	0	0	1				-	-	-	-
A1220	0	1	1	0	0	1	0				-	-	-	-
										TOTAL	2685	29	2656	11.13

NB: 0 indicates agreement;
1 indicates discrepancy between the records

FIGURE 7

"Possible" Decision Results
Soundex Check Program - 1962 Updating

Reference code	Soundex code & Sex	Surname	First name	Addr.	Birth year range	Soc. Sec.#	Maid. name	Race	Birth mo. & day	Birth year	Computer possible linkages	Clerical non-linkages	Clerical Net linkages	Clerk Time hours
A0131	0	0	0	0	1	1	1	0	1	1	20	20	0	.25
A0132	0	0	0	0	1	1	1	1	0	1	-	-	-	-
A0133	0	0	0	0	1	1	1	1	1	1	-	-	-	-
A0232	0	0	0	1	0	1	1	0	1	0	227	35	192	5.50
A0233	0	0	0	1	0	1	1	0	1	1	690	622	68	3.50
A0235	0	0	0	1	0	1	1	1	0	1	3	0	3	.08
A0432	0	0	1	0	0	1	1	0	1	0	8	0	8	.25
A0434	0	0	1	0	0	1	1	1	0	0	-	-	-	-
A0435	0	0	1	0	0	1	1	1	0	1	-	-	-	-
A0437	0	0	1	0	0	1	1	1	1	1	-	-	-	-
A0530	0	0	1	0	1	1	1	0	0	1	1	1	0	.02
A0532	0	0	1	0	1	1	1	1	0	1	-	-	-	-
A0631	0	0	1	1	0	1	1	0	0	1	95	92	3	.50
A0632	0	0	1	1	0	1	1	0	1	0	2127	2115	12	8.75
A0634	0	0	1	1	0	1	1	1	0	0	2	2	0	.02
A0635	0	0	1	1	0	1	1	1	0	1	44	44	0	.20
A0932	0	1	0	0	1	1	1	1	0	1	-	-	-	-
A1230	0	1	1	0	0	1	1	0	0	0	2	0	2	.02
A1231	0	1	1	0	0	1	1	0	0	1	-	-	-	-
A1232	0	1	1	0	0	1	1	0	1	0	-	-	-	-
A1234	0	1	1	0	0	1	1	1	0	0	-	-	-	-
A1235	0	1	1	0	0	1	1	1	0	1	-	-	-	-
										TOTAL	3219	2931	288	19.09

NB: 0 indicates agreement;
1 indicates discrepancy between the records

FIGURE 8

DATE OF BIRTH CHECK FOR LINKAGE-CHECKING

1962 UPDATING

TO BE APPLIED TO ALL "REJECTS" AND
UNACCEPTED "POSSIBLES" FROM SOUNDEX CHECK

Reference code	Birth mo. & day; sex	Social Sec. #	Birth year	Maiden name	First name	Address	Birth year range	Decision	Computer Counts	Clerical Rejects	Net Linkages
B01	0	0						Accept	7	0	7
B02	0	1	0	0				Accept	16	0	16
B03	0	1	0	1	0			Accept	196	142	54
B04	0	1	0	1	1	0		Possible	9	1	8
B05	0	1	0	1	1	1		Reject			
B06	0	1	1	0	0			Accept	0	0	0
B07	0	1	1	0	1			Possible	30	30	0
B08	0	1	1	1	0	0	0	Accept			
B09	0	1	1	1	0	0	1	Possible	4	4	0
B10	0	1	1	1	0	1		Reject			
B11	0	1	1	1	1	0	0	Possible	38	38	0
B12	0	1	1	1	1	0	1	Reject			
B13	0	1	1	1	1	1		Reject			
								TOTAL	300	215	85

NB: 0 indicates agreement; 1 indicates discrepancy between the records

Tolerance Rules for Concordance:

Social Security Number - Complete agreement
Maiden name - Complete agreement in either maiden names or in cross-check with surname
Birth year, first name, address, birth year range - Same as in Figure 2

III

SOME RESULTS FROM TWO NEW DATA REDUCTION PROCESSES

Chairman, J. F. Daly, U. S. Bureau of the Census

SOME RESULTS FROM A NON-SYMMETRICAL BRANCHING PROCESS THAT LOOKS FOR INTERACTION EFFECTS [1]

James N. Morgan and John A. Sonquist
Survey Research Center
Institute for Social Research
University of Michigan

This paper presents some results of a data reduction process designed, programmed and operative on the IBM 7090. [2] It was designed with sample survey data in mind, that is, for data characterized by several thousand cases, a large number of explanatory variables or classifications, moderate intercorrelations among the predictors, and a continuous dependent variable, not badly skewed, but with a good deal of unexplained variation or noise.

Proponents of a new statistical procedure are always in danger of claiming too much for their method, both in terms of how original it is and in terms of how good it is. The program is original in scope but not in essence. It represents a simulation, with some added breadth and quantification, of what careful researchers have done for years by hand or using an IBM sorter and tabulator when investigating a new set of data.

Simulation of human behavior by a computer is not new. Ours is a particular kind of simulation not so much designed to gain insight into the behavior simulated, but to do a particular kind of job better than the human has the time or patience to do. We have systematically examined the behavior of a social science researcher tackling a particular kind of data analysis problem, making decisions, isolating interesting subgroups and computing statistics. We then stated the behavior explicitly and formally as a sequential set of decision rules and extended them to what the researcher might do if he had the time and patience. An examination of the preliminary results of the program indicates, among other things, that the decision rules of a researcher are more complicated than he realizes. We are now incorporating changes into our model and into the program to reflect these more sophisticated rules. But we have made a start in what we feel will be a fruitful line of development; the simulation of the researcher by machine. There are some unsolved problems of optimal strategy which we raise in the hope of stimulating further work along these lines.

So the basic idea is not new. What is new is the formalization of the analysis procedure and the capacity to apply it systematically and rigorously, so that unrealized compromises and arbitrary choices do not occur, and the results are completely reproducible.

There are some things this analysis technique will not do. It will not locate the best functional form in a set of data with a limited number of numerical predictors and a relatively low level of error or noise. That is what Professor Westervelt's process does well, and you will hear about that later on this program. Our procedure also will not answer the question whether a particular variable has a significant effect on the dependent variable, the other variables somehow controlled or "held constant". It was designed for a set of data full of interaction effects, and wherever there are interaction effects, by and large, it is not meaningful to ask about the direct effects of one variable at a time. It is difficult to give up the habit of testing for the effect of one variable after another. Yet in much behavioral research we measure not the theoretical factors in which our interest lies, but rather the measurable, proxy factors which, we hope, may reflect more basic characteristics. These often must interact (in the statistical sense) to be able to represent the theoretical construct adequately. For instance, "family size" may be used to represent "the amount of housing space needed," but in combination with "income" also may be used as a proxy for "ability to pay for housing." The procedure can, however, minimize the noise for a tight test of the effect of a single factor.

The procedure is not designed for a very large number of highly correlated predictors such as batteries of attitude questions. Also, since at present, it does not look ahead more than one step at a time, it will not locate certain completely offsetting symmetrical negative interaction effects where neither factor has any effect by itself. For example:

% Who go to Hospital

	Men	Women	
Young	2%	8	5
Old	8	2	5
	5	5	5%

But let's see what the program will do. We shall describe the process in terms of its purpose. (The formal algorithm is appended). The program divides the sample through a series of binary splits into a mutually exclusive set of subgroups. Every observation is a member of exactly one of these groups. These subgroups are chosen so that their means account for more of the total sum of squares (reduce the predictive error more) than the means of any other set of subgroups. The stopping point is subject to arbitrary decision and is set by parameters at the beginning of the computer run. (The present set of rules for stopping represents the best we have so far, but may not necessarily be optimal in terms of research strategy. Further work needs to be done in this area.)

At any stage in the branching process, the set of groups developed at that point represents, according to the criteria of the model, the best currently available scheme for predicting the dependent variable in that sample, from the information available. If the sample is representative, this is the best scheme for the population.

There are some minor qualifications to these claims, but for large samples and without certain symmetrical negative interactions, they seem to be valid.

In deciding which split to make, the rule is to scan all feasible splits and select the one which reduces the error sum of squares the most. This is a rule of importance, not significance. Subgroups which are significantly different, but which are so small that isolating them does not help in predicting a randomly selected individual (or group), are not split off. Why the rule of importance rather than significance?

Multivariate statistical techniques have been developed to the point where the argument can be made that significance tests are of doubtful value because of the large number of variables tried. With samples in the range of 1,000 to 3,000 observations many factors show up as statistically significant which are not important, in terms of their contribution to reducing predictive error. (They may, of course, have theoretical importance, for some reason.) The process we are using of scanning for all feasible splits at each stage vastly increases the number of things tried, so that the whole notion of degrees of freedom seems useless in this model. Formalizing the process makes this more obvious.

At each stage, the computer selects the group with the largest sum of squares within it, locates the best way of splitting it into two subgroups using each predicting classification, then takes the best of the best (on the basis of the largest between-groups sum of squares). We do not need to examine all possible combinations of classes of each predictor when separating a group into two parts, since it can be shown that after rearranging the classes into descending (or ascending) sequence according to the size of the class means of the dependent variable, it never pays to combine non-adjoining classes. [3/] Thus, with k classes for predictor X: only k-1 feasible splits exist. But a little algebra shows that with a dozen predicting classifications of, say, eight classes each (7 feasible splits each), and with this scanning repeated at each split, the number of different trees theoretically possible is, to put it mildly, somewhat larger than the sample size. There are clearly no degrees of freedom left.

The rules for stopping overlap. A safety precaution puts a maximum on the number of final, unsplit groups at, say, 20. No group containing only a small internal sum of squares (less than 2% of the original total sum of squares) is examined further. And no split is allowed unless it reduces the overall predictive error by at least a visible amount like ½%. (That is, the between sum of squares for the split must be more than ½% of the original total unexplained sum of squares for the whole sample).

We can summarize the rules then:

Take the group with the largest unexplained sum of squares within it, so long as it is more than 2% of the original sample's unexplained sum of squares.

Considering all predictors, find the best binary division of that group, in terms of the "between splits sum of squares", so long as it is greater than ½% of the original sample's unexplained sum of squares. (If no split on this group is worth while, try the group with the next largest internal sum of squares.)

If no group can be found worth examining, or if none of those which are found can be profitably split, then stop.

The details are in our article in the June Journal of the American Statistical Association.[4/]

Note that all predictors are treated as classifications, even if this means making classes out of a continuous variable. It has been felt that the small loss of information from grouping is offset by the flexibility in discovering non-linear relationships. Dummy-variable regression models are similarly attractive in this respect.[5/]

We turn now to some actual results. Chart 1 and Table 1 are a relatively simple analysis of a dichotomous dependent variable: whether or not the spending unit owns its own home. Comparisons between the tree and multiple regression findings are not easy, but one way is to ask which predictors appear most important. For this purpose we use from the new analysis the total reduction in error from splitting on that predictor, whether used once or more than once. For dummy variable regression we use a partial beta coefficient, squared to put it in the same dimensions. [6/] This can be thought of as the partial beta coefficients one would get if he took the dummy variable coefficients for each predictor to create a new scale or variable, and ran a multiple regression with these new scaled variables.

We give also the gross beta coefficient squared, which can be thought of as the square of a dummy variable multiple correlation coefficient using dummies to represent all the classes (except one) of that particular predictor. It tells how much of the unexplained sum of squares can be accounted for by using only that one predicting classification in all its detail.

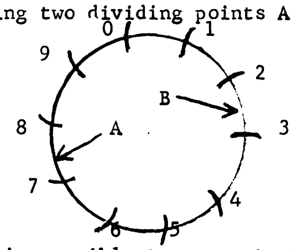
It is apparent that roughly the same variables appear important in both analyses. Does the tree really tell us anything that we didn't already know from the regression? Clearly it tells us that married people, and people with higher incomes become home owners earlier in life. In other words, the effects of age on home ownership depend on other things. Stated another way, the effect of income on home ownership depends on age and family status. There are interaction effects.

The total explanatory power of the two methods cannot be compared, since the regression uses all classes of each predictor, and since the newer analysis could always "explain" more if we allowed it to continue splitting. It is not even legitimate to argue that the new method explains more per unit of information utilized, since it has, in trying and discarding, used all the information the regression used, and with no restrictive additivity assemptions. But the tree provides an economical way to summarize a lot of data, and to make comparisons with other times or places.

Chart 2 and Table 2 move to a numerical dependent variable, annual medical expenses, based on a sample of individuals in Michigan.[7] Here again the same variables seem important, but while the regression indicated that good insurance coverage was associated with heavier utilization of medical services, the tree indicates that this is true to an important extent only for adult females. This makes sense to the analyst and certainly to someone concerned with keeping insurance costs down, this knowledge is of more use than the regression finding that insurance affects utilization.

Some of the problems with our original strategy can be seen in the tree illustrated in Chart 2. Why the particular combination of family sizes that puts 2.0 and 4.0 equivalent adults[8] together? It turns out that these are mostly people with 1 or 3 children. Perhaps they are the women who had a child born last year, but if so, the sub-group combinations appear partly fortuitous. We have allowed re-ordering of the subclasses of a predictor so that, for instance, the middle-aged could be separated from the old and young. The tendency for adult women with either low or high income to have higher medical bills may be real. But there may be places where we want to retain the order, if only for simplicity.

One way out of this impasse, which we are thinking of building into a new program, is to prevent the linear re-ordering of the classes of some predictors according to their means. Instead, an option would be provided to treat them as a circle, with identifying codes running from 1 to 9, thence to 0 and back to 1 again. Dividing a group into two sub-groups then requires determining two dividing points A and B. For example:



Thus, it is possible to separate the middle from both extremes without allowing erratic combinations. We shall surely add to the program a still more restrictive option that preserves the original linear scale ordering and allows only a split along that ordering. This option is being incorporated into the computer program in such a fashion that the researcher may specify which predictors are to have a split restriction, if any, placed on them.

Turn now to Table 3. The first use of the new program, with hourly earnings as the dependent variable, kept to the predicting classifications used in the regression analysis presented in the book Income and Welfare in the United States, by J. N. Morgan, M. David, W. Cohen and H. Brazer.[9] Here we have a relatively obvious example of the problem of variables at different stages in a causal process. Some of the explanatory variables used come earlier in time or are otherwise logically prior, and can help determine the levels of other explanatory variables, but cannot be affected by them. Age, race, sex are determined at birth, and will influence how much education a person gets. All these, including his education, will influence his occupation, and the cumulative set will help determine his present hourly earnings. In a case like this the analysis should be sequential if it is intended to explain the process, not merely predict. The clearly prior (exogenous) variables should be used first, and the residuals run against a more inclusive set of predictors. It is necessary to reintroduce such exogenous variables as race into the next stage because they may well mediate the effect of other things (education) on the dependent variable (earnings). The revisions being introduced into the program make such a sequential analysis easier.

At any rate, it is clear that if one puts both occupation and education in the same stage, occupation (being a more powerful predictor) "takes over" and, in the sequential design we use, education cannot fight its way back into the analysis as it does in multiple regression. This is a characteristic of sequential procedures when several predictors are intercorrelated, and the analysis problems result from the nature of the question being asked. If one wants only the best prediction, then even causal patterns among the predictors can be ignored. If one wants to unravel the causal mechanism then a several-stage analysis is called for. Even among correlated predictors at the same stage, the one that is best at an early split may explain enough of the variance so that the others may not show up at all.

Chart 3 shows the tree omitting two predictors, occupation and supervisory responsibility. Here education, age, and sex are clearly important. There are some other interesting findings. Achievement motivation, a variable interesting for theoretical reasons, turns out to be important in explaining hourly earnings only for middle aged college graduates. This is an interesting and meaningful finding, since these tend to be the people who are best able to affect their own hourly earnings by their attitudes and efforts. Most other people have to work longer hours or take a second job if they want to earn more money. This type of interpretation is, of course ex post facto, and should be validated by examining its implications and performing additional analyses to test the interpretation.

With regressions, it is possible to compute the sampling error of each dummy variable regression coefficient, or to make F-tests for each set. These tests are of doubtful validity when applied to multi-stage clustered samples, often weighted to adjust for sampling and response rate variations. There seems to be even less theoretical basis for computing sampling errors for one of our trees. Another sample might produce an entirely different sequence of splits. But the sampling stability of the branching process is, nevertheless, of some interest.

One way to investigate the stability of the branching process is to repeat it on split half samples. The results from three such subsamples are given in Table 3. The predictors which are seen to be important seem not to vary from sample to sample. On the other hand, the trees are different, sometimes even at the first split. The similarity occurs in terms of the groups which finally result, because one can isolate groups which are nearly the same by splitting in different orders. For instance, one can select first college graduates, then middle aged, then men, or first men, then college graduates, then middle aged, etc., arriving the same place by a different route.

Another way of investigating stability, which we are working on, is to take a tree derived from one sample and ask how well those final groups predict in a different sample.

Chart 3 still has predictors at several stages, sometimes combined into a single classification such as where the head of the unit grew up and where he lives now. This was done originally to build some interaction effects into the regression analysis, that is, to investigate mobility. How do we know that more basic things like race are not being pushed "out of the tree" by other things which are largely the result of race in the first place? To answer problems of this type, we must use a multistage analysis.

Chart 4 and Table 4 represent the first stage of a more detailed step-wise analysis where we use only the predictors that were determined in early childhood, (N/Ach is so determined, in theory). The smaller number of predictors with fewer subclasses leads to a tree that is easier to look at and interpret. It is clear that there are interaction effects, for the tree would be symmetrical if there were not. Indeed, one gets the impression that the interactions are of a particular kind which can be interpreted by saying that disadvantages are substitutes for one another while advantages are complements. Having one or two disadvantages is enough, and further splits on the others will not explain additional variation and hence will not appear. Being old, or young, or uneducated, or a woman, or from a southern or rural background, or nonwhite, are alternative barriers to high earnings. In this analysis those predictors which affect many people severely, tended to be used early in the tree. Those affecting a minority, like race, tend not to appear because we can explain enough by knowing the other things. This does not mean there is no prejudicial discrimination. Rather it reflects a characteristic of the analysis that it does not test each explanatory factor holding all the others constant, but asks whether a factor is needed more than any other factor, given the group currently under consideration.

The extreme case of such substitutability among predictors would be a tree where once a group was split off as having one disadvantage, it would never be split again. Whenever groups are split reflecting extreme disadvantages (being very old, very young, or a woman) they tend not to be split again in Chart 4. Similarly those with very low education tend not to split further.

Persons accustomed to one method of analysis always like to translate problems back into solutions with the familiar method. With regression analysis, such patterns would clearly require not simple cross-product terms, but a new set of dummy variables like:

(1) The man has one of the following disadvantages:
(2) He has two or more of them

If interactions behave like those described above, then such variables will take over and little credit will be left for each of the separate components.

In our original trees we tried to include the card counts and standard deviations of the means, but it was too much to look at. These data are all part of the output, however. The program also provides details of the subclass means just before each split is made, and the original subclass means for each predictor.

In looking for ways to present the data, we have tried one other method which is illustrated in Tables 5 and 6. Here we select one predictor and show the way the splits were made and the definitions of the subgroups just before they were split.

Table 5 shows the three ways in which education was used, once to divide the whole sample, and then on two different subgroups. Perhaps one reason why a second split on education was not made earlier, before looking at where the individual grew up, is that the critical educational level seems to differ depending on the person's background. For those without the disadvantage of a farm or southern background, it appears to be graduation from high school that matters, for the others whether they even learned to read.

Table 6 shows the four groups split on age, and how the splits were made. Clearly "middle-aged" people earn more in general, but for some groups (generally those with more education and fewer disadvantages) "middle age" starts at 35, while others reach it at 25. (The college graduate group contained one person 75 or older who was making $4.04 an hour and was grouped together with the "middle aged" by the computer.)

There is one problem which is not well illustrated in the data we present here: With a large number of predictors of many classes each, there is an increased possibility of fortuitous or untrustworthy splits. When the investigator looks at them, he immediately becomes aware that something is missing in a strategy which is willing to make <u>any</u> split that is important enough to reduce the error sum of squares by half a per cent. In other words, in arguing that it is importance that matters, not statistical significance, we tend to over-simplify the research problem. A better rule might be to proceed according to the importance of splits, but disregard any split, even if it appears important, if it is not significant, i.e., may be a fluke.

How can something be important but not significant? Whenever there are any extreme cases, or sufficient flexibility in combining codes on any of many variables and for smaller and smaller subgroups, this will allow the process to isolate some subgroup consisting of a few extreme cases. The formal relations between the number of cases, the between sum of squares, and the F-test need to be worked out, but it is clear from the fact that the square root of N appears in the denominator of the estimated sampling error, that subgroups of fewer than ten cases are unlikely to have means that differ <u>significantly</u>. Hence we are building into the program a side rule that no split is allowed if one of the resulting groups contains fewer than n (for example 10) cases. Further work clearly needs to be done in this area. Is it the size of the subgroup before splitting which should be above some minimal point? Should the smaller of the two subgroups split off? Answers are not yet forthcoming. One may also argue that small subgroups split off from large ones are deviant cases which <u>should</u> be removed from the main analysis and then examined in great detail.

Where does one go from here? Ideally the subgroups identified by the branching process should lead to the development of some new theoretical constructs, new variables that are combinations of the measured factors and have theoretical meaning as well as practical significance. Having defined these new variables, one could use them in an ordinary regression analysis for presentation purposes (and marvel at how well one explained things). The theoretical question raised is "why are these variables important". And, again, further analysis must provide the answers.

Some studies of the forecasting stability of this method compared with multiple regression would also be useful, as we have pointed out.

SUMMARY

It is clear, at least to us, that for purposes of discovering the structure of relations in a body of data, what really is related to a dependent variable, under what conditions, and through what intervening processes, this procedure offers some real advantages. Its strategy is to focus on what can be found out from the data with some assurance, rather than on testing the significance of effects of many factors and their cross-products, the results of many such tests being basically inconclusive rather than negative.

There are clearly some neat unsolved statistical problems of optimum strategy, or at least consistent strategy in setting the various arbitrary cut-off points. The 2% of total sum of squares before a group is examined, the ½% reduction in error before a split is allowed, and the minimal number of cases in a subgroup, should all depend on the sample size, the number of predictors, the constraints on rearranging scales, and the variance of the dependent variable, relative to its mean.

The trees look formidable at first, but are basically simpler than multiple regression results; the results to be presented being the definition of a subgroup and its mean on the dependend variable

We hope we have now started to come full circle to the point where the computer is doing what we want to do better, rather than doing incredible amounts of arithmetic, the results of which often do not meet the real needs of the analyst.

FOOTNOTES

1. A paper presented at the Meetings of the American Statistical Association, Case Institute of Technology and Western Reserve University, Cleveland, Ohio, September, 1963.

2. A great deal of credit is due to Kathleen Goode and Wen Chao Hsieh of the ISR programming section, who did the programming. The program is identified as the Automatic Interaction Detector (AID) Model 1. It is written in MAD for a 32K IBM 7090. It operates using the U. of M. Executive System Monitor. Our thanks are due to Dr. R. C. F. Bartels of the U. of M. Computing Center on whose equipment this experimentation took place.

3. A proof of this is due to Professor William Erickson of the U. of M., Computing Center and Mathematics Department.

4. Morgan, J. N., and Sonquist, J. A., Problems in the Analysis of Survey Data - and a Proposal, JASA, 58, (June, 1963), pp. 415-34.

5. See Suits, D. E., Use of Dummy Variables in Regression Equations, JASA, 52, (dec. 1957,) pp. 548-551.

6. See Andrews, F. M., The Revised Multiple Classification Analysis Program, Institute for Social Research, University of Michigan, August 1963, 13 pp. Multilith.

7. See Grover Wirick, Robin Barlow and James Morgan "Population Survey: Health Care and its Financing" in Walter J. McNerney et al, Hospital and Medical Economics (2 Vols.) Vol. 1, pp. 61-360, Chicago, Hospital Research and Educational Trust, 1962.

8. In the scale, the second adult, and children under 12 are counted as ½ each.

9. New York, McGraw Hill, 1962.

TABLE 1 --

WHETHER SPENDING UNIT OWNS ITS HOME

Predictors	Gross Beta Coefficients[2]	Multiple Classification Partial Beta Coefficients[2]	AID Analysis-Reduction in TSS(I)/TSS(T)
Age of heads	.111	.099	.107
Income	.088	.068	.040
Number of persons	.088	.062	.084
Unusual income last year	.039	.010	.000
Race	.014	.003	.000
Number of persons earning $600	.011	.003	.000
Education of heads	.001	.002	.000
R^2		.251	.231

CHART 1 --

AID ANALYSIS OF
WHETHER SPENDING UNIT OWNS ITS HOME

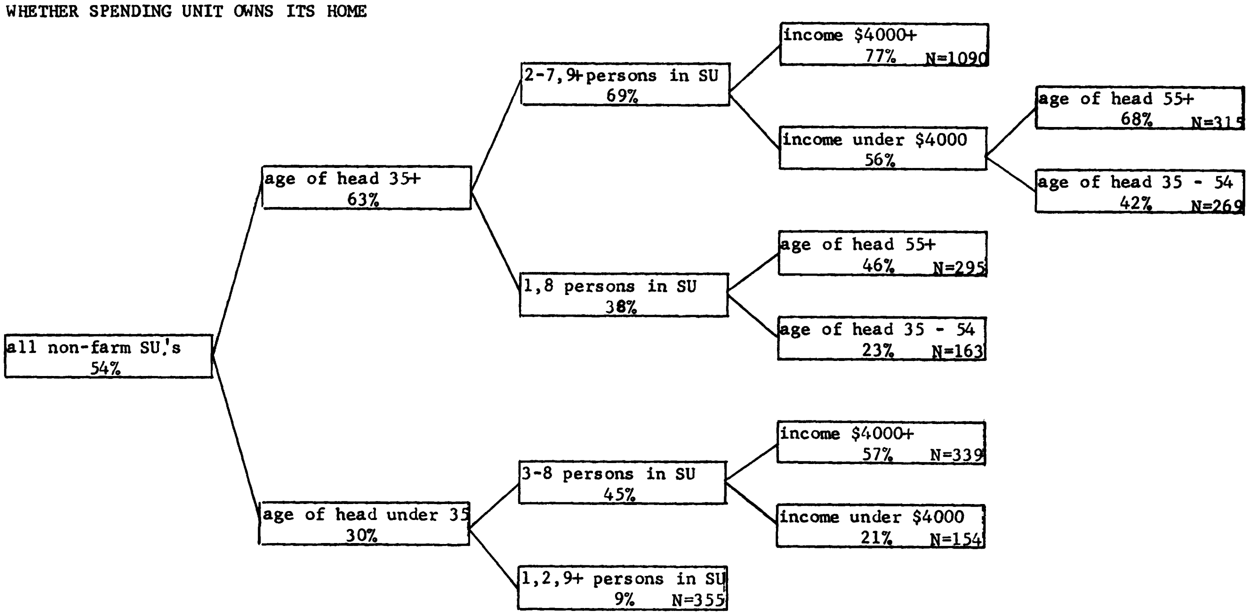


TABLE 2 --

INDIVIDUALS' MEDICAL EXPENSES

Predictors	Gross Beta Coefficients[2]	Multiple Classification Analysis- Partial Beta Coefficients[2]		AID Analysis- Reduction in TSS(I)/TSS(T)
		Males	Females	
Sex	.012	-----	-----	.016
Age	.041	.033	.069	.043
Health insurance coverage	.010	.011	.017	.007
Family income	.014	.006	.008	.005
Equivalent adults in family	.019	.004	.002	.007
Attitude toward early care	.002	.004	.004	.000
Education of head	.003	.003	.002	.000
Region where head grew up	.007	.001	.002	.000
Service level	.005	-----	-----	.000
R^2	----	.077	.089	.078

CHART 2 --
AID ANALYSIS
OF INDIVIDUALS' ANNUAL
MEDICAL EXPENSES

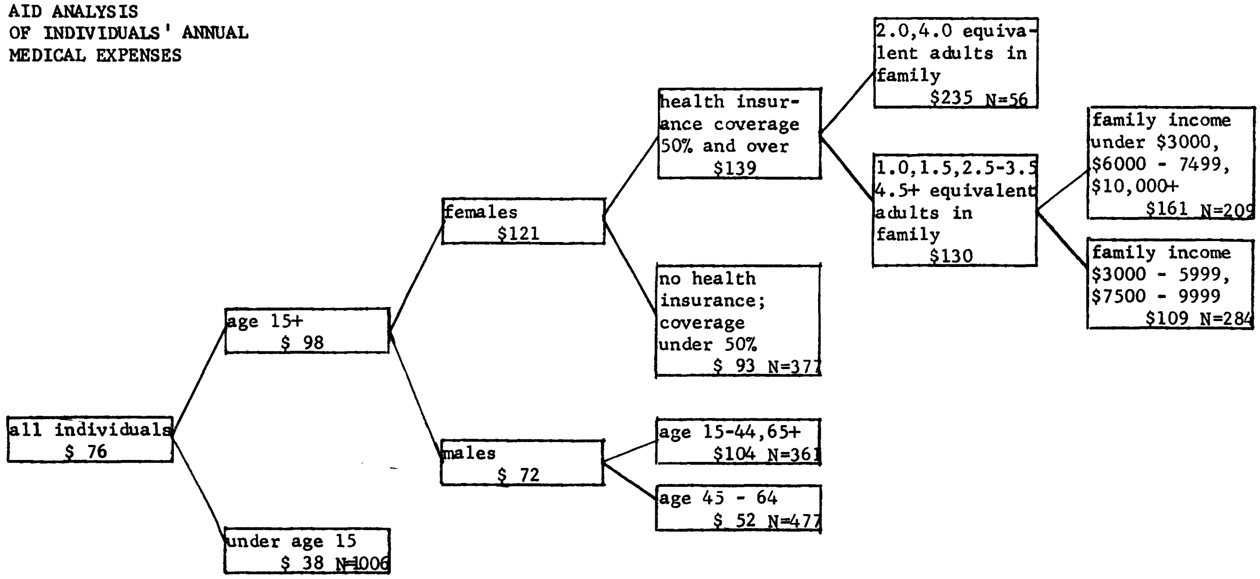


CHART 3 --
AID ANALYSIS
OF HOURLY EARNINGS
OF SPENDING UNIT HEADS

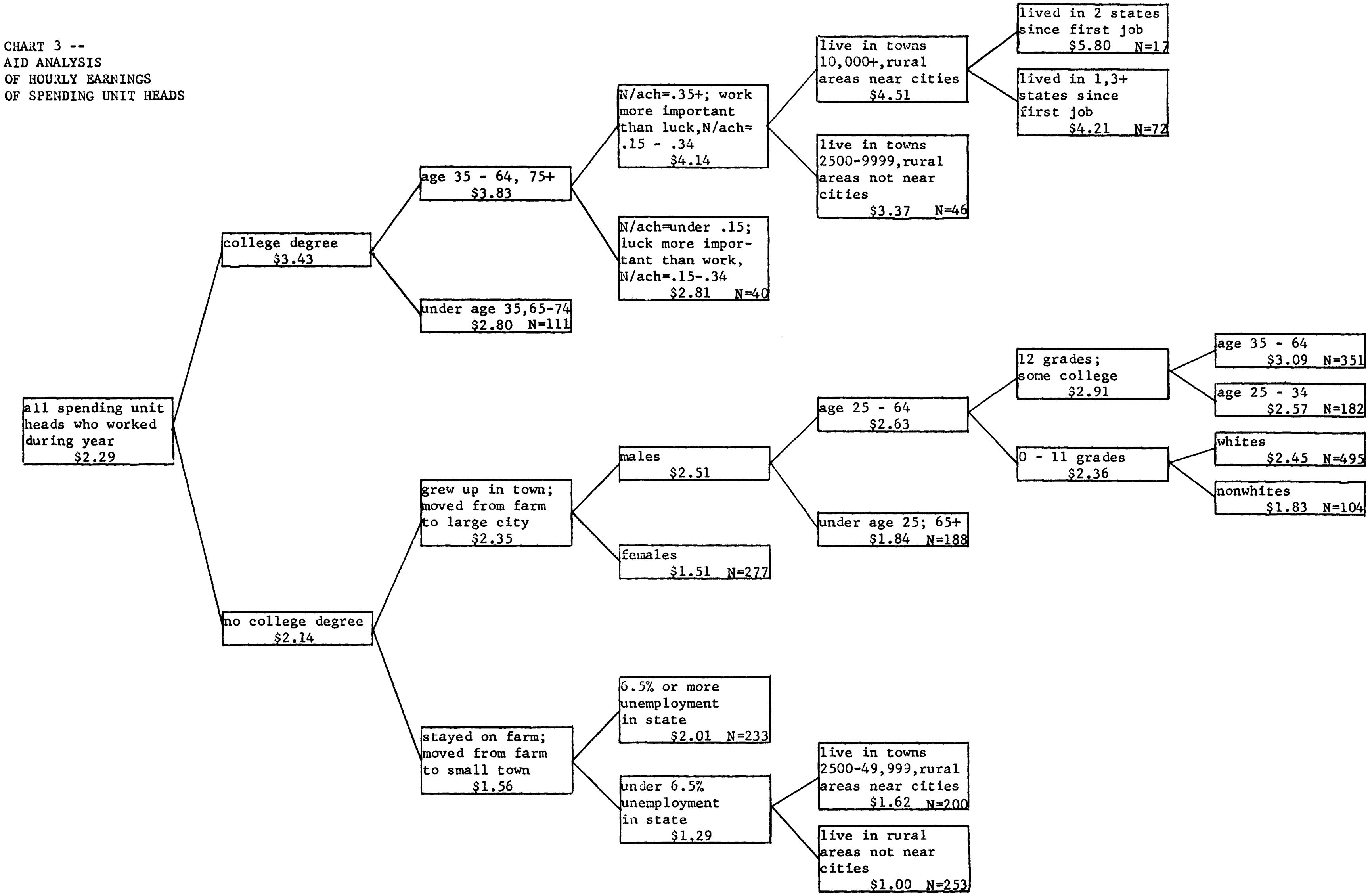


TABLE 3--

HOURLY EARNINGS OF SPENDING UNIT HEADS

Predictors	Gross Beta Coefficients^2	Multiple Classification Analysis-- Partial Beta Coefficients^2 With Occupation	AID Analysis-- Reduction in TSS(I)/TSS(T) With Occupation	AID Analysis-- Without Occupation Total	Split Half #1	Split Half #2	Split Half #3
Education	.133	.055	.014	.097	.073	.101	.083
Age	.039		.035	.040	.052	.039	.068
Sex	.045	.048	.039	.042	.062	.046	.057
Occupation	.159	.042	.211	----	----	----	----
Population of Cities	.063	.032	.000	.014	.028	.029	.051
Urban-rural migration	.079	.015	.025	.051	.088	.067	.101
Movement out of Deep South	.038	.010	.000	.000	.027	.017	.000
Unemployment in states	.024	.009	.000	.013	.011	.038	.042
Supervisory responsibility	.065	.007	.005	----	----	----	----
Attitude toward hard work, need-achievement score	.030	.005	.000	.011	.023	.013	.016
Race	.025	.004	.000	.005	.000	.000	.005
Ability to communicate	.032	.004	.000	.000	.000	.000	.000
Geographic mobility	.007	.003	.000	.007	.023	.013	.008
Physical Condition	.016	.003	.000	.000	.022	.009	.000
Rank and Progress in School	.052	.001	.000	.000	.000	.000	.000
R^2		.359	.329	.280	.409	.372	.431

TABLE 4 --

HOURLY EARNINGS OF SPENDING UNIT HEADS -- EXOGENOUS FACTORS ONLY

Predictors	Gross Beta Coefficients^2	AID Analysis- Reduction in TSS(I)/TSS(T)
Education	.133	.108
Age	.039	.039
Sex	.045	.046
Background	.069	.045
[illegible]	.020	.008
Race	.025	.000
Physical condition	.016	.000
Rank and progress in school	.052	.000
Religion	.040	.000
R^2	----	.246

CHART 4 --
AID ANALYSIS
OF HOURLY EARNINGS
OF SPENDING UNIT HEADS
(EXOGENOUS FACTORS ONLY)

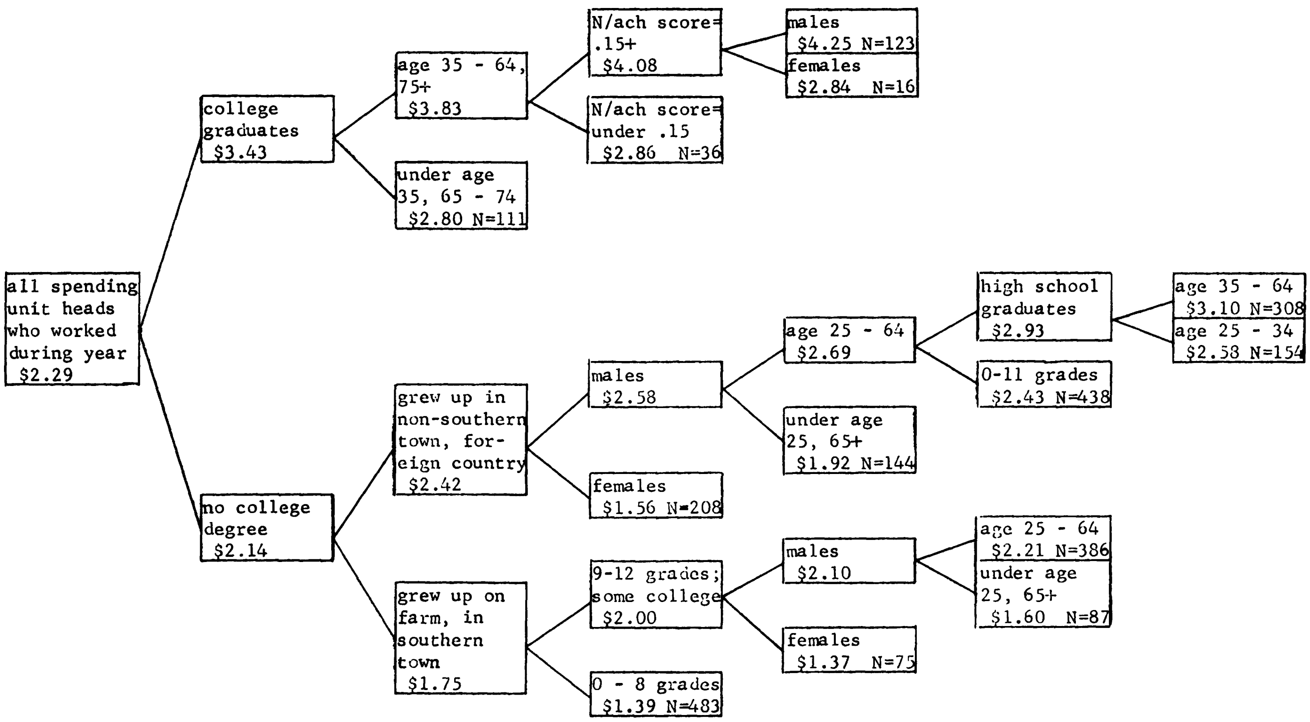


TABLE 5--

AID EDUCATION SPLITS ON HOURLY EARNINGS FOR SPENDING UNIT HEADS WHO WORKED IN 1959 (MEAN AMOUNTS FOR EACH SUBGROUP)

Education		All		Non-college graduates, Grew up outside Deep South in small town or city, male, aged 25-64		Non-college graduates, Grew up in Deep South or on a farm	
None		$1.04		$2.57		$.85	
1 - 8 grades		1.72	0 - 11 grades $\bar{Y}$=$2.43	2.32	0 - 8 grades $\bar{y}$=$1.39	1.42	
9 - 11 grades	Non-College graduates $\bar{y}$= $2.14	2.14		2.52		1.92	
12 grades		2.35		2.82		2.06	
12 grades plus non-academic		2.48		3.03	12 grades or more y=$2.93	1.99	9 grades or more $\bar{y}$=$2.00
College, no degree		2.51		3.01		2.17	
College, bachelor's degree		3.25					
College, advanced degree		3.98	College graduates $\bar{y}$=$3.43				
Number of cases		2569		900		1031	
Mean for group		$2.29		$2.69		$1.75	

TABLE 6 -- AID AGE SPLITS ON HOURLY EARNINGS FOR SPENDING UNIT HEADS WHO WORKED IN 1959 (MEAN AMOUNTS FOR EACH SUBGROUP)

Age	Mean Wage Rate	Non-college graduates, Grew up outside Deep South in town or city, who are male	College graduates	Non-college graduates, Grew up outside Deep South in town or city, who are male, aged 25-64 with 12+ grades of school	Non-college graduates, Grew up in Deep South South or farm, 9+ grades of school, male
Under 25	$1.68	$1.84	$2.08		$1.68
25 - 34	2.32	2.47	3.03	$2.58	2.15
35 - 44	2.52	2.72	3.77	2.94	2.34
45 - 54	2.41	2.78	4.13	3.20	2.05
55 - 64	2.34	2.85	3.55	3.33	2.22
65 - 74	1.67	2:.19	2.30		1.23
75 or older	1.02	.96	4.04		1.41
Number of cases	2569	1044	286	462	473
Mean for group	$2.29	$2.58	$3.43	$2.93	$2.10
Splits		25-64 $\bar{y}$=$2.69; Under 25, 65 or older $\bar{y}$=$1.92	Under 35, 65-74 $\bar{y}$=$2.80; 35-64 75 or older $\bar{y}$=$3.83	25-34 $\bar{y}$=$2.58; 35-64 $\bar{y}$=$3.10; Under 25, 65 or older $\bar{y}$ = $1.60	25-64 $\bar{y}$=$2.21

(A)utomatic (I)nteraction (D)etector

Model 1

Algorithm: Condensed Form.

PRELIMINARY READ IN. STEPS 0 AND 1.

0. Read in all parameters and all input observations, including all predictors and the dependent variable y.

1. To start, identify all input observations as belonging to group number one. Group number one is the current candidate group. Go to Step 5.

TEST FOR TERMINATION OF THE PROCEDURE. STEP 2.

2. Determine whether or not the current number of un-split groups is about to exceed the maximum permissible number; if so, go to Step 20, as the problem cannot proceed further.

DETERMINE WHICH GROUP SHOULD BE SELECTED FOR ATTEMPTED PARTITIONING. STEPS 3-5.

3. Considering all groups constructed so far, find one of them such that

   a. the total sum of squares (TSS_i) of that group is greater than or equal to 2 percent of the total sum of squares for the input observations (TSS_t);

   b. the group has not already been split up into two other groups;

   c. there has been no previous failure to split up the group;

   * d. the total sum of squares of that group is not smaller than the sum of squares for any other group that meets the above three criteria.

4. If there is no such group, go to Step 21; the problem is complete.

5. The group selected is the current candidate group, which will be the subject of an attempted split. Identify it with its group number (i) and, by option, print out N_i $\sum y_i \sum y_i^2$ $\bar{Y}_i$ and TSS_i. These statistics are always printed out if the group number one is the current candidate group.

PARTITION SCAN OVER ALL PREDICTORS. STEPS 6-17

6. Set j = 1 and go to Step 8.

7. Increment j by 1. If j is larger than the number of predictors being used in the analysis, the partition scan is complete; go to Step 18.

8. Compute N_{ijc}, Σy_{ijc}, Σy_{ijc}, $\bar{Y}_{ijc}$ for each class c of predictor j over group i.

9. Determine whether or not there exist two or more classes c, such that $N_{ijc} \neq 0$. If not, predictor j is a constant over group i; print an appropriate comment and go to step 7.

10. Sort the statistics produced in Step 8, together with the class identifiers for predictor j, into descending sequence using $\bar{Y}_{ijc}$ as a key.

PARTITION SCAN OVER THE c CLASSES OF PREDICTOR j.
STEPS 11 - 15

11. Set p = 1 and go to Step 13.

12. Increase p by 1. If p is larger than $(c_j - 1)$, where c_j is the number of classes in the jth predictor, then go to Step 16 as all possible feasible splits have been examined.

13. If $\Sigma N_k = N_1 = 0$ for k = 1...p, or if $(N_i - N_1) = N_2 = 0$, go to Step 12 as this split cannot be made because of empty classes in this group for predictor j. Otherwise, compute BSS_p, the between-groups sum of squares for the attempted binary split of group i on predictor j between the sorted classes (1,..,p) and the adjacent sorted classes (p+1,..,c).

*14. If this BSS_p is not larger than any BSS_p previosuly computed for this predictor over this group, go to Step 12.

15. This is the largest BSS_p encountered so far for this predictor. Remember BSS_p and the partition number p; then go to Step 12.

DETERMINATION OF BEST PREDICTOR. STEPS 16-17.

*16. Was the maximum BSS_p for predictor j larger than the largest BSS_p obtained from any of the other predictors previously tested over group i? If not, go to step 7.

17. This is the best BSS_p produced by any of the predictors tested so far over group i. Remember this partition and this predictor and then go to Step 7.

IS THE BEST PREDICTOR WORTH USING? STEPS 18-19

*18. Was the maximum BSS retained after the scan of all predictors over group i equal to at least 1/2 percent of the total sum of squares? If not, mark group i as having failed in a split attempt and then go to step 3.

19. Group i is to be split into two new groups and destroyed. Using the class identifiers and the partition rule remembered from Step 17, split the observations in group i into two parts. Identify the two new groups as having been created. Identify group i as having been split. Print the statistics from the successful partition attempt. Increase the total number of groups created so far by the quantity 2. Increase the current number of un-split groups by one. Then go to Step number 2.

TERMINATION OF THE ALGORITHM. STEPS 20-21

20. The maximum number of permissible un-split groups has been reached. Print an appropriate comment and go to Step 22.

21. There are no more groups eligible for further splitting. Print an appropriate comment and go to Step 22.

22. Print out a summary record of all groups created in the process of splitting, including the group number, its parent group, the values of the predictor class identifiers that were used in the partition which constructed the group, the predictor number used in this partition, an indication of whether or not this present group was ever split, and N_i, Σy_i Σy_i^2, and Tss_i. Stop.

FORMULAS

$$\bar{Y} = \Sigma y/N$$

$$TSS = \Sigma y^2 - \frac{(\Sigma y)^2}{N}$$

$$BSS = \frac{(\Sigma y_1)^2}{N_1} + \frac{(\Sigma Y2)^2}{N_2} - \frac{(\Sigma y)^2}{N}$$

$$WSS = TSS - BSS$$

* These decision rules constitute the crucial steps in the process which may be described in more global terms as follows.

1. Select that sample subgroup which has the largest total sum of squares, $TSS_i \geq .02\ (TSS_T)$

$$TSS_i = \Sigma X_i^2 - \frac{(\Sigma X_j)^2}{N_i}$$

The total sample is considered the first (and indeed, only) such group at the start.

2. Find the division of the classes of any single characteristic such that the partition p of this group into two subgroups on this basis provides the largest reduction in the unexplained sum of squares. Choose a division so as to maximize $(N_1\bar{X}_1^2 + N_2\bar{X}_2^2)$ with the restrictions that (1) the classes are ordered in descending sequence using their means as a key and (2) observations belonging to classes which are not contiguous after sorting are not placed together in one of the new groups to be formed.

3. For a partition p on variable k over group i to take place after the completion of (2), it is required that:

$$(N_1\bar{X}_1^2 + N_2\bar{X}_2^2) - N_i\bar{X}_i^2 \geq .005\ (\Sigma X_T^2 - N\bar{X}_T^2)$$

Otherwise group i is not capable of being split. No variable is "useful" over this group. The next most promising group (TSS_i = max) is selected.

4. If there are no more groups such that $TSS_i \geq .02\ (TSS_T)$ of if for the groups that meet this criterion there is no "useful" variable, or if the number of un-split groups exceeds a specified number, the process terminates.

Note 1: Eligibility criteria (2% for trying, ½% for an acceptable split) can be changed and should be for varying sample sizes and numbers of predictors. The program handles weighted data; the formulas being easily derivable.

Note 2: For an extended explanation of this algorithm see J. N. Morgan and J. A. Sonquist, Problems in the Analysis of Survey Data -- and a Proposal, JASA, Vol. 58 No. 302, June 1963, pp 415 - 434.

IV

ACCURACY OF CENSUS RESULTS

Chairman, Dudley Kirk, The Population Council

A PRELIMINARY EVALUATION OF THE 1960 CENSUSES OF POPULATION AND HOUSING

By Conrad Taeuber and Morris H. Hansen
Bureau of the Census

INTRODUCTION

The purpose of this paper is to summarize the findings to date of the work on the evaluation of the quality of the 1960 Censuses of Population and Housing. We are bringing together some of the material that has been presented by various staff members on other occasions and are adding some new material that has become available. A general familiarity with the traditional functions of the decennial census in the United States will be assumed. However, some notes about innovations in the 1960 Censuses are offered as background against which we may consider to what extent we achieved our goals.

On the whole, our intention was to collect much the same kind of population data in 1960 as in 1950. Although we planned essentially the same census as in 1950 with respect to content, we tried to achieve this goal with improved accuracy of data and with considerably earlier publication of results, at lower cost than would have been involved by using 1950 methods in 1960. The major changes were therefore procedural.

Among the principal changes in methods in 1960 were the following: (1) Extension of sampling, limiting the complete count to a few basic items, and collecting most of the information from a 25-percent sample of households; (2) extension in the use of electronic equipment, including an auxiliary device (FOSDIC), which eliminated manual card punching; (3) new enumeration methods including the taking of the census in two stages for most of the country, and making use of the Post Office for the distribution of the first-stage forms covering 100-percent data, to be completed and held for the enumerator on his regular door-to-door canvass when he also left the sample data forms at every fourth household for completion and mail-in; and (4) new and more widely applied methods of quality control in data collecting, editing, coding and all the steps in processing leading to the production of final results.

The evaluation work is only partially completed. We tentatively conclude, from the results now available, that in general, our 1960 methods have succeeded in producing a better census than the 1950 Census. More data on a small-area basis have become available (reference 1). We are presenting some evaluation information as a basis for discussion of the completeness of enumeration and accuracy of data on characteristics.

RELATIVE COMPLETENESS OF COVERAGE OF 1960 AND 1950 CENSUSES

This analysis discusses the accuracy of the 1960 Census count of the total resident population of the United States relative to that of the 1950 Census. The discussion relates to the net change in the total resident population of the United States as indicated by the 1960 and 1950 Censuses and to the independent estimates of the components of change during the decade. The comparison of the independent estimates of change based on the components of change with the net change implied by the 1960 and 1950 Census counts provides some evidence as to whether the 1960 level of net undercount differs appreciably from the 1950 level and in which direction.

Estimates of population change and of the components are as follows:

(In thousands)

Population April 1, 1960	179,323
Population April 1, 1950	151,326
Net increase	+27,997
Components of change:	
Births (corrected for under-registration)	40,963
Deaths	15,608
Net movement of aliens and citizens	+2,695
Net movement of Armed Forces abroad	-330
Expected net increase	+27,720
Estimated increase in coverage based on above	+277

From the demographic components of population change indicated above, it appears that the absolute level of net undercount in 1960 was a little lower than in 1950; however, each of these components is subject to varying degrees of error, which have an important impact on the estimated net undercount. This estimate is subject to an error of 100 percent or more, and the impact of this and some alternative estimates will be indicated. Furthermore, 1960 computer processing may have introduced a new element in the intercensal balancing equation and its effect should be considered in any coverage evaluation. These topics are dealt with in more detail below.

First, however, we may speculate about the rate of net underenumeration in the two censuses on the basis of this estimate. For 1950, we may take three rates for illustrative purposes from table B of The Post-Enumeration Survey: 1950 (Technical Paper No. 4):

Source	Estimated net undercount in 1950 Census	
	Amount (000's)	Percent of estimated totals
PES estimate	2,091	1.4
"Minimum reasonable" estimate	3,715	2.4
Coale estimate	5,429	3.6

If now, we subtract 277,000, the estimated increase in coverage, we obtain the following estimates for 1960:

1950 benchmark	Estimated net undercount in 1960 Census	
	Amount (000's)	Percent of estimated totals
PES estimate	1,814	1.0
"Minimum reasonable" estimates	3,438	1.9
Coale estimate	5,152	2.8

Births

The estimate of births shown above was adjusted for incomplete registration on the basis of the results of the Birth Registration Test of 1950 and extension of these results to 1960. The 1950 Test indicated that 2.1 percent of the births occurring in the January-March 1950 period were not registered. The overall correction for the 1950-1960 decade is about 600,000, or 1.45 percent. If, for example, birth under-registration were 50 percent greater, the intercensal change would be about 300,000 greater and the estimated 1960 net undercount would be larger by that amount. If, on the other hand, birth under-registration were only half as great as estimated (as might be the case if completeness improved within the stratum of births occurring out of hospitals), then the estimated net undercount in 1960 would be about 300,000 smaller.

Deaths

No evidence as to the extent of under-registration of deaths is available. In these computations, however, the data for infant deaths have been adjusted by the factor used for births; approximately 21,000 deaths were added for the entire decade. If, for example, we were to assume that all deaths were incompletely registered to the same extent as births--a very unlikely situation--the total adjustment would be 233,000, and the estimated undercount in 1960 would be that much less.

Net civilian immigration

Of the three demographic components of change, the estimate of net immigration to the United States is subject to the greatest margin of error. Although the size of the component (net) must be relatively small compared with that of the other components (about 3 million for the decade versus about 40 million births and 15 million deaths), the uncertainty involved in estimating the size of some of the elements that make up net immigration from abroad is very large.

The immigration statistics cover six principal classes of migrants, as defined by the Immigration and Naturalization Service (INS): Immigrant aliens, nonimmigrant aliens, and citizens arriving; emigrant aliens, nonemigrant aliens, and citizens departing. In addition, it is necessary to allow for movement between the United States and Puerto Rico and other outlying areas.

The estimates of the gross components that make up net immigration used in preparing the final intercensal population estimates are shown in table 1.

Table 1.--MIGRATION: 1950 TO 1960

(In thousands. Figures may not add to totals because of rounding)

Class of migrant	Net movement	Total	
		Arrivals	Departures
Alien immigrants and emigrants	+2,249	[1]2,500	[2]251
Alien nonimmigrants and nonemigrants	-10	[3]3,459	[3]3,469
Citizen passengers arriving and departing	-	NA	NA
Movement to and from Puerto Rico and other outlying areas	+455	3,725	3,269
Total	+2,695	NA	NA

[1]Since July 1957, includes alien nonimmigrants whose status was later converted to immigrants.
[2]Based on reported data to July 1957. The INS believes that the number of departures has been quite small in recent years.
[3]Data to July 1956 only.

The net movement of U.S. citizens (exclusive of those moving between Puerto Rico and the mainland) was estimated from two different sets of data: (1) INS figures on sea and air travel; (2) the census counts of Americans abroad in 1950 and 1960. The first yielded estimates of arrivals and departures whereas the second gives an estimate of net movement only.

The first approach shows 12,264,000 arrivals and 11,984,000 departures, or a net in-movement of 280,000. In this approach, it is implicitly assumed that the large gross volume of movement over our land borders had a zero balance. Thus, the figure +280,000 is subject to a very large margin of error.

The second approach may be summarized as follows:

Net emigration of civilian citizens, 1950-1960	172,000
= Civilians enumerated abroad, 1960	765,000
- Alien dependents abroad, 1960	41,000
- Civilians abroad, 1950	239,000
+ Alien dependents abroad, 1950	9,000
- Registered births to citizens abroad, 1950-1960	332,000
+ Estimated deaths to civilians enumerated abroad, 1950-1960	10,000

There are several uncertainties in this estimate: (a) The 1950 Census did not count Americans abroad such as businessmen, students, retired persons, etc. It was assumed that this group, 188,000 in 1960, increased over the decade at the same rate as Federal employees and their dependents overseas; (b) the estimated number of alien dependents of Americans overseas had to be subtracted--41,000 in 1960 and 9,000 in 1950; and (c) births to U. S. citizens abroad are reported voluntarily to U. S. consuls, and the number is probably incomplete.

In view of the net in-movement shown by one estimate and the net out-movement by the other and of the uncertainties concerning both, it was simply assumed for purposes of this paper that there was no net movement of citizens in the 1950-1960 decade. Had the estimate of +280,000 been used, the estimated decrease in coverage would have been 3,000. On the other hand, had the -172,000 been used, the estimated increase in coverage would have been 449,000.

Net movement of armed forces abroad

The estimate for this component is based on data from the Department of Defense and is believed to be subject to little error. The estimate includes allowance for deaths to Armed Forces overseas during the period and for overseas inductions. (With respect to the enumeration of military personnel overseas, there was about the same numerical difference in 1950 as in 1960 between the census data and data from the Department of Defense. In 1960 the census deficit in the overseas population was about 50,000 out of 700,000, compared with a 1950 deficit of 61,000 out of 400,000.)

Computer processing

As is mentioned in the United States summary of general population characteristics (reference 2), 776,655 persons were included in the 1960 count through computer imputation of population to housing units for which there was some evidence of occupancy. Part of this evidence came from an indication of an occupied unit on the housing schedule but with no corresponding FOSDIC-readable persons on the population schedule, and part from a reenumeration of a sample of field "close outs" that were so reported in which it was found that many such units were occupied. In any census there is always a marginal group from the standpoint of whether they were literally "counted." Granted that the 1960 procedure for computer imputation of population was a necessary final stage of the enumerative process, there is, nevertheless, some evidence that the computer may have "overimputed" persons. The amount of this overcount has not been closely determined, but its range could reasonably be from 100,000 to 400,000.

Summary

It is evident from the prior discussion that errors in the intercensal estimates of births, deaths, and of military movement are not likely to be of sufficient magnitude to affect the general picture regarding the relative accuracy of the 1960 and 1950 counts. Errors in the immigration data, on the other hand, may be fairly large; and, as stated earlier, it is not possible to determine the direction or approximate size of the net error. The nature of computer imputation for housing units with no occupants listed further clouds the picture.

In summary, if we make no allowance for any overimputation of persons in the special computer procedure, and if we assume no net immigration of citizens in the 1950-1960 decade, the estimated amount of net underenumeration is lower in 1960 than in 1950. If we allow 250,000 for computer overcount then the amount of net underenumeration would be about the same in 1960 as in 1950. Under either assumption the estimated rate of underenumeration would be lower in 1960, and sufficiently lower to indicate a gain of the order of 500,000 to 1,500,000 in the coverage in the 1960 census as compared with what it would have been if the 1950 rate of undercoverage had continued.

ABSOLUTE ESTIMATES OF COVERAGE AND OF GROSS COVERAGE ERRORS

The previous discussion is concerned primarily with coverage of the 1960 Census as compared to the 1950 Census, and is based on 1950 evaluation study results, and on estimates of population changes between 1950 and 1960. Although the evaluative studies for the 1960 Census have not as yet been brought to final conclusion, many of the results are available, and preliminary findings can be given. Comparisons of independent estimates of coverage errors between censuses are difficult because the evaluation studies themselves differ in effectiveness.

One major method for studying coverage in both 1950 and 1960 was made through a re-enumerative procedure. Omissions and errors in the counting of occupied living quarters are analyzed, as a source of omission and duplicated enumeration of persons. Within properly enumerated living quarters, there can be omissions or erroneous inclusions of occupants. Table 2 shows estimates of coverage errors for 1960 and 1950, as estimated from re-enumerative surveys.

Table 2.--ESTIMATES OF POPULATION COVERAGE ERROR

(Percent of Census total)

Enumeration errors	1960	1950
Omissions of persons...............	3.0	2.3
In missed living quarters.........	1.6	1.6
In enumerated living quarters....	1.4	0.6
Erroneous inclusions of persons....	1.3	0.9
Net undercoverage of persons.......	1.7	1.4

The estimates in table 2 are not entirely comparable for 1960 and 1950. In 1960, the check for missed persons included "housing units" and all "group quarters." In 1950, the check included "dwelling units" (comparable to "housing units" in 1960) and "quasihouseholds" where less than 35 persons had been enumerated. In addition, there was a difference in the timing and effectiveness of the re-enumeration procedures.

On the basis of other studies and evidence, it was concluded that the net undercoverage in the 1950 Census was substantially underestimated by the 1950 Post-Enumeration Survey (PES) (reference 3), especially for persons not readily identified with a regular place of residence. As a consequence of weaknesses detected in 1950 re-enumerative procedures, steps were taken to strengthen corresponding procedures in 1960. Therefore, for 1960, there are higher, and perhaps more reasonable, estimates of net undercoverage than in 1950.

The results of analytical methods previously described may be combined with the evidence from the re-enumerative studies to give some overall estimates of net undercount. Through resurvey methods a net undercount of population of about 1.7 percent may be estimated, and through analytical methods between 1.0 and 2.9 percent. Considering the evidence now available, Steinberg (reference 4), et al have indicated that a reasonable estimate of the level of net undercount in 1960 seems to be in the range of 1.7 to 2.0 percent of the total as compared to the "minimum reasonable" estimate in the 1950 Census of 2.4 percent. In absolute terms, this amounts to a net undercount in 1960 of 3 million to 3½ million people.

A composite of analytic methods, reported on by Akers (reference 5), gives us some estimates of net undercounts for 1960 by sex, age and color (see table 3). The estimates are fairly reliable for ages under 25 but are quite rough for the older ages. For ages under 25, the estimates are based on survivors of births. Coale's iterative method was used to estimate ages 25 to 64, as well as ages 65 to 74 for whites. Another iterative method using mortality data was used for the older ages and for nonwhites 65 to 74.

Table 3.--ESTIMATES OF NET CENSUS UNDERCOUNT BY BROAD AGE GROUPS, SEX, AND COLOR: 1960

(Composite of preferred methods. A plus sign indicates a net overcount.)

(Percent)

Age	All classes	White male	White female	Non-white male	Non-white female
All ages....	-2.3	-1.1	-1.7	-10.3	-7.1
Under 5 yrs.	-2.6	-2.1	-1.4	-7.9	-6.4
5-14 yrs....	-2.1	-2.3	-1.3	-4.8	-3.8
15-24 yrs...	-4.0	-3.3	-2.3	-13.9	-9.5
25-44 yrs...	-2.6	-2.2	-0.7	-16.0	-6.2
45-64 yrs...	-2.3	-0.2	-1.8	-13.0	-12.8
65 yrs. and over.......	+0.9	+8.1	-4.5	+7.9	-2.6

Entirely different estimates of net census undercount of the population 65 and over are suggested by a comparison of estimates of the population 65 years and over in 1960. The 1960 estimate of this population based on survivors of the 1950 population and other factors fell short of the census count by about 900,000, or 5.5 percent. A similar discrepancy was also noted in 1950 and 1940. In the interpretation of this discrepancy for earlier censuses, greatest weight was given to erroneous reporting of age at the latest census. However, an examination of the population count for ages 55 to 64 in 1960 and of age misreporting as indicated in the reenumerative procedure tends to discount this factor as a major source of bias. Nor can a deficiency in the reporting of immigration contribute much to the explanation. In view of the lack of evidence regarding the relative weights of various factors, a fuller explanation of the discrepancy must await the results of the other studies in the evaluation program.

Table 4 shows a comparison of estimates for 1960 and 1950 for undercounts of children under one year of age and under five years for whites and nonwhites. These results reflect the accuracy of age reporting as well as completeness of coverage. The figures are based on estimates of survivors of births, using birth registration data and estimates of underregistration made by the National Vital Statistics Division. Each of the age and color groups shown, except 10 to 14, indicates considerable improvement in 1960.

Table 4.--ESTIMATED PERCENTAGE OF NET CENSUS UNDERCOUNT OF CHILDREN

Color and year	Under one year	Under five years old	5 to 9 years old	10 to 14 years old
TOTAL				
1960.....	2.0	2.6	2.4	1.9
1950.....	11.0	4.7	3.6	1.8
WHITE				
1960.....	1.2	1.8	2.0	1.7
1950.....	10.1	4.0	2.7	1.0
NONWHITE				
1960.....	5.8	7.2	4.8	3.7
1950.....	16.6	9.7	9.6	6.3

Further evaluation of the 1960 Census will be possible as final results of a series of studies become available. Of interest is a de facto enumeration of persons present in a sample of enumerated living quarters and a number of "reverse record checks" that are being carried through, involving samples from other sources being matched against the census.

NONRESPONSE RATES

Table 5 provides a comparison of nonresponse rates in the 1960 and 1950 Censuses for a few characteristics. Most of the nonresponse rates compared are higher in 1960 than in 1950.

In 1950, after an enumerator had made reasonable but unsuccessful efforts to obtain census information about persons from the usual acceptable respondents, he was instructed to make inquiries from neighbors. This procedure was followed in 1960 only for the 100-percent items for which neighbors might provide reasonably acceptable information.

In 1960, for the sample items, there was a close-out procedure instructing the enumerator to obtain information about persons from acceptable respondents only, and to terminate his efforts after three calls. This procedure was patterned after CPS policy which does not permit the enumerator to obtain information about a person from any source other than a responsible member of the household. In the 1960 Census, when a nonresponse rate in an enumerator's assignment was found to be unacceptable on review, further follow-up work was to be done by hourly rate enumerators.

The 1960 procedures were based on the assumption that allowing information to be obtained from neighbors and other unqualified respondents encouraged poor standards and loose work, and that with a reasonably low nonresponse rate, mechanical imputation yielded data that are more reliable than inquiry of neighbors or informal imputation by the enumerator. For some items (such as place of birth and mother tongue, occupation, place of work and means of transportation) nonresponses were not imputed in the computer, but were tabulated as NAs. The alternative methods for dealing with nonresponse still need to be evaluated. At present, it is difficult to appraise whether the higher nonresponse rates in 1960 (table 5) represented a deterioration or improvement in quality.

The increase in nonresponse for age, which is a 100-percent item, presumably results from the collection of information by date of birth, instead of age in years as of the last birthday.

Table 5.--PERCENT OF NONRESPONSE FOR SELECTED CHARACTERISTICS: 1960 AND 1950

Selected characteristics	Percent nonresponse	
	1960	1950
Age................................	[1]1.7	0.2
State of birth......................	2.7	1.0
School enrollment (persons 5-34 yrs. old)............	8.3	[2]5.9
Highest grade completed (persons 25 and over)..............	4.9	4.6
Employment status (persons 14 & over)	3.1	1.0
Occupation (employed persons)........	4.9	1.3
Children ever born (to women ever married)............	6.0	9.0
Income (persons 14 & over)...........	6.2	6.7

[1]Year or decade of age not reported. The 1.7 figure is obtained on the basis of Stage I or 100-percent enumeration. In Stage II, the corresponding nonresponse figure was 1.0 percent.

[2]Enrollment data available only for persons 5 to 29 years old in 1950.

In considering the NA rates for occupation, it should be noted that employment status was computer-allocated in 1960, but not in 1950 when NA's for employment were placed in the category, "Not in the labor force." The NA rates with respect to "occupation" are taken relative to the number of employed persons. The 1960 allocation procedure for employment tends to increase the NA's for occupation. Hence, the difference in procedure probably accounts for a part of the higher NA rate for occupation in 1960 compared with 1950.

The income NA rates shown here for 1960 and 1950 are also not strictly comparable because the additional editing performed in 1960 and the revised questions tended to reduce the number of NA's that had to be imputed by the computer. For example, in 1950, all persons who failed to report on work experience and on earnings were counted as income nonrespondents; in 1960, persons who failed to report on work experience and on earnings, but were subsequently assigned by the computer to the category "Did not work in 1959" were assigned "None" codes in "Wage or salary" and in self-employment income and were not counted as NA's on earnings. Also, the modification in the 1960 questions eliminated from consideration those persons who did not have work experience during the preceding 10 years. In 1950, however, the income questions were to be asked of all persons 14 years of age and over. Both these changes tended to reduce the NA rates for 1960.

The Census in 1960 was taken by a two-stage procedure for about 82 percent of the population, and by the usual "one-stage" procedure elsewhere (reference 6), including, for the most part, the more sparsely settled areas. Some comparisons of nonresponse rates for occupation for 1960 and 1950 were made for single-stage and two-stage cities of similar size by Shryock and Greene (reference 7). Sixteen of the largest single-stage cities were selected and matched to cities of similar size enumerated by the two-stage procedure. The entire 1950 Census was taken in one stage. Comparisons were made with the 1950 nonresponse rates for the same 32 cities selected for 1960. The results are summarized in table 6.

Table 6.--EMPLOYED PERSONS NOT REPORTING OCCUPATION, PERCENTAGES FOR SELECTED CITIES ENUMERATED BY SINGLE-STAGE AND TWO-STAGE METHODS IN 1960 BY COLOR: 1960 AND 1950

For selected cities (employed persons)	Percent of nonresponse for occupation	
	Single-stage cities	Two-stage cities
All persons, 1960........	4.0	6.5
All persons, 1950........	1.0	1.1
White persons, 1960......	3.9	5.8
Nonwhite persons, 1960...	4.4	10.1

Thirteen of the two-stage cities had a nonresponse rate in excess of their "comparable" single-stage cities. For 1950, however, the nonresponse rates do not indicate any pattern of difference between the two groups of cities.

An inspection of nonresponse rates for 100-percent items (see table 7) can be made for 1960 by size of place (reference 8). The central cities of urbanized areas, where there are special problems of enumeration, had a relatively large number of imputations compared with percentages for the United States as a whole. Corresponding figures for rural areas were relatively low, and those for urban areas outside central cities of urbanized areas were generally near those for rural areas.

Table 7.--ALLOCATION FOR NONRESPONSE FOR THE UNITED STATES BY SIZE OF PLACE: 1960

Size of place	Number of persons	Persons substituted (percent)	Persons with one or more allocations (percent)
United States........	179,323,175	0.4	3.0
Urban total.........	125,268,750	0.5	3.2
Central cities....	57,975,132	0.6	3.9
Urban fringe......	37,873,355	0.3	2.6
Other urban:			
Places of 10,000 or more.......	16,172,839	0.3	2.7
Places of 2,500 to 10,000.....	13,247,424	0.4	2.5
Rural total:			
Places of 1,000 to 2,500........	6,496,788	0.3	2.4
Other rural.......	47,557,637	0.4	2.4

This type of allocation shown in the column "Persons substituted," not included in the other percentages in this table, consists of cases where persons, and all their characteristics, were substituted for an estimated number of persons in households for which the enumerator obtained no population data.

As indicated earlier, moderate NA rates, of themselves, are not in general a satisfactory indicator of quality of the census measurements for characteristics. The NA rate for CPS, which we regard as of high quality relative to the census, is approximately four or five percent for most items, and considerably higher for some. In part, this NA rate results from the fact that interviewing must be completed within a very short time.

On the other hand, where NAs run to 25 or 50 percent or more for a particular area, it can only be interpreted as providing poor data for that area. This, for example, was the situation in the city of Chicago where among census tracts having 1,000 inhabitants or more, there were three tracts, each containing one or more EDs, in which allocation rates were 50 percent or more for a majority of the sample items.

ERRORS IN PROGRAMMING AND TABULATING

It is not possible, of course, to give a meaningful quantitative summary of errors in programming and tabulating. Some errors are important; others are trivial. Minor differences between tables resulted from imperfections in the processing of the data, but these are much smaller in magnitude for the 1960 Census than for earlier censuses. There is at least one major tabulation error, however, which was not caught until after publication. The figure for Westchester County (N.Y.) residents born in the United Kingdom was shown as being larger than the county's total population.

Programming errors affected all publication areas, except that sometimes an error was found and corrected after several States had already been tabulated and retabulation of the first States was not deemed feasible. Sometimes corrected figures are obtained and shown separately with the errata. At other times, the reader is simply told the nature of the error (and usually the direction of the bias), but the corrected figures are not available.

RESPONSE BIAS

Introduction

Since NA rates, when they are relatively low, do not necessarily provide an adequate measure of quality, a more valid test of the results is made after imputation for the NAs. Such measures of quality are estimated by comparing census results with other data regarded for this purpose, as of better quality or as "preferred" for particular items of information. In this context, preferred data may be information independently obtained through record keeping or reporting systems, other independent surveys, or reinterview surveys. When the census and the "preferred" data collection method yield comparable summary measurements of the same characteristics, the differences between corresponding summary statistics, taking sampling error into account, can be considered as a measurement of <u>net response error</u>, or <u>response bias</u>. Gross error, or response variance, will be discussed later.

For purposes of this discussion of response bias, census results will be compared with results from the Current Population Survey, with information independently obtained by other governmental agencies for education and income, and with results obtained through intensive reinterviews of a sample of 1960 Census respondents. We take the difference between a census result (C) and a comparable result from another source regarded for the present as a standard (S), and compute: (C-S)/S as a measure of <u>relative response bias</u>.

Comparison of Census results with CPS results

Labor force data were not collected in the intensive reinterview study; the Current Population Survey is used as the standard against which census statistics of the employment status of the population are measured (see table 8 and reference 9). The April CPS data for 1950 and 1960 are used for comparison although there are differences in the time reference. The census enumeration was spread over time and the data do not relate to a single week, but they relate mostly to a week in April.

Without exception, for all population groups shown, there is evidence, according to CPS results, of census undercount of persons in the labor force. However, except for persons employed in agriculture, the differences relative to CPS results are smaller for 1960 than for 1950. Differences in employment in agriculture may be especially affected by the variations in time reference.

(It should be noted that the usual sampling error estimates apply to the CPS data, and also to the differences between Census and CPS data, regarding sampling error for the census results as trivial.)

The response biases, as previously defined, are indicated in columns 4 and 8 of table 8.

Response bias in 1960 Census education statistics

Two basic items of information on education were collected in the census: school enrollment and educational attainment.

Table 8.—COMPARISON OF EMPLOYMENT STATUS, CENSUS AND CPS

(Percent)

Employment status	1960				1950			
	Census	April CPS	Difference	Difference rel. to CPS	Census	April CPS	Difference	Difference rel. to CPS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All persons, 14 years old and over								
In labor force	55.5	57.0	-1.5	-2.7	53.7	56.9	-3.2	-5.6
Employed	52.6	54.0	-1.4	-2.6	51.2	53.7	-2.6	-4.8
In agriculture	3.5	4.4	-.9	-21.3	6.3	6.6	-.3	-5.0
In non-agricultural industries	49.2	49.6	-.5	-.9	44.9	47.1	-2.2	-4.8
Unemployed	2.9	3.0	-.1	-4.7	2.6	3.2	-.6	-19.9
Not in labor force	44.5	43.0	1.5	3.6	46.3	43.1	3.2	7.4
White								
In labor force	55.3	56.6	-1.3	-2.3	53.4	56.5	-3.0	-5.4
Employed	52.7	54.0	-1.3	-2.4	51.1	53.5	-2.4	-4.6
Unemployed	2.6	2.6	-.02	-.8	2.4	3.0	-.6	-19.9
Not in labor force	44.7	43.4	1.3	3.0	46.6	43.5	3.0	7.0
Nonwhite								
In labor force	57.0	60.7	-3.7	-6.0	56.6	61.2	-4.6	-7.5
Employed	52.1	54.6	-2.5	-4.5	52.1	55.6	-3.5	-6.3
Unemployed	5.0	6.1	-1.2	-19.4	4.4	5.6	-1.1	-20.3
Not in labor force	43.0	39.3	3.7	9.3	43.4	38.8	4.6	11.9
Male								
In labor force	78.1	80.0	-1.8	-2.2	79.7	83.3	-3.6	-4.4
Employed	74.2	75.7	-1.5	-2.0	75.8	78.3	-2.6	-3.3
In agriculture	6.6	7.8	-1.3	-16.3	11.8	11.8	-.1	-.6
In non-agricultural industries	67.6	67.9	-.2	-.3	64.0	66.5	-2.5	-3.8
Unemployed	3.9	4.2	-.2	-6.0	3.9	5.0	-1.1	-21.6
Not in labor force	21.9	20.1	1.8	8.7	20.3	16.7	3.6	21.7
Female								
In labor force	34.9	36.3	-1.4	-3.9	29.2	32.1	-2.9	-9.2
Employed	33.0	34.3	-1.4	-4.0	27.8	30.5	-2.7	-8.8
In agriculture	.6	1.3	-.6	-50.0	1.0	1.6	-.6	-37.2
In non-agricultural industries	32.3	33.1	-.7	-2.2	26.8	28.9	-2.1	-7.3
Unemployed	1.9	1.9	-.04	-2.1	1.3	1.6	-.2	-15.8
Not in labor force	65.1	63.7	1.4	2.2	70.8	67.9	2.9	4.3

In a paper by Nam (reference 10) for overall figures on enrollment in the public schools, two comparisons were made from information now available (see table 9). First, in accordance with its annual practice, the U. S. Office of Education issued an estimate of total enrollment, from kindergarten through the twelfth grade, based on a brief mail questionnaire sent to all offices of State school systems in the fall of 1959 requesting enrollment figures as of October 1, 1959, or as close to that date as enrollment stabilization was believed to have occurred. The census information referred to school enrollment between February 1, 1960 and the census date. Secondly, the CPS had a school enrollment question in October 1959. This information was gathered in the usual CPS way, and is subject to CPS sampling error.

Table 9

Source	Enrollment (millions)
1960 Census..........................	35.3
1959 Office of Education Survey......	35.2
1959 CPS (October)...................	34.9

The census figure might be expected to be slightly lower than the 1959 OE fall figure because the Census data refer to the spring semester and some slight attrition at these grades takes place between the fall and spring terms. However, some shifting from parochial or other private schools to public schools probably also takes place during that time, and this shifting would tend to compensate for the attrition effect. At any rate, the figures are quite close. Also, when considering comparable figures on a State basis (reference 10), in only 13 States is the difference between Census and the OE figures as great as 3 percent and in only one State (Alaska) is the deviation extreme. Most of these differences by States, moreover, can probably be attributed, in great part, to varying definitions of residence or to transfers or residential mobility between States in the interval from fall to spring.

College enrollment data from the 1960 Census can also be compared with the October 1959 CPS and with the 1959-60 biennial Office of Education Survey (see table 10). The OE information is obtained from college and university officials. Although the CPS and OE figures are in close agreement, the Census differs from both to a marked degree.

Table 10

Source	Enrollment (millions)
1960 Census..........................	2.9
1959 CPS (Oct.)......................	3.3
1959-1960 OE Survey..................	3.4

A special inquiry concerning fall and spring enrollment, sent to the largest universities in six States where the Census figure fell appreciably below the OE figure, showed that the attrition rate was very close to the difference between the two sets of figures.

Response bias in 1960 Census income data

For evaluation of census income data with respect to response bias, two sources of information are now available: CPS data and estimates of the Office of Business Economics (OBE). In the present section, some findings reported by Miller (reference 11), about CPS and OBE estimates compared with Census data (see table 11) are made. All relate to aggregate income for the year 1959 for 1960 comparisons and the year 1949 for 1950 comparisons.

Table 11.--CENSUS, CPS AND OBE ESTIMATES OF AGGREGATE INCOME, BY TYPE OF INCOME, FOR THE UNITED STATES: 1959 AND 1949

Year and type of income	Census[1]		CPS[2]		
	Dollar amount	Percent difference from OBE	Dollar amount	Percent difference from OBE	OBE[3]
1959					
Total income....	331.7	-5.6	306.7	-12.7	351.4
Wages & salaries	246.5	-1.0	233.5	-6.3	249.1
Self-employment.	47.9	13.5	38.3	-9.2	42.2
Income other than earnings..	37.3	-37.9	32.7	-45.6	60.1
1949					
Total income....	173.2	-9.3	159.8	-16.3	191.0
Wages & salaries	[4]124.3	-3.5	120.0	-6.8	128.8
Self-employment.	[4]31.1	-0.6	26.5	-15.3	31.3
Income other than earnings..	[4]16.6	-46.3	13.3	-57.0	30.9

[1]Total population 14 years old and over.
[2]Persons 14 years old and over, excluding inmates and members of Armed Forces living on base.
[3]Total population, all ages.
[4]Estimates on preliminary sample tabulations because final data do not contain distribution of income by type.

A comparison of the Census and OBE estimates by type of income shows that in 1949 and 1959 there was very close agreement for wages and salaries but evidence of substantial underreporting of income other than earnings in the Census. In each case the Census estimate was in closer agreement with OBE than were the CPS figures and there was also substantial reduction in underreporting in 1960 compared to 1950.

Census and OBE estimates for regions and States were in close agreement for 1959. Census estimates were less than 90 percent of OBE estimates for only four States.

Age heaping

An additional point of some interest in relation to response bias, approached analytically, is a decline in "age heaping", as shown in a recent study (reference 12).

In each census in which data on single years of age have been collected, there have been overstatements of ages ending in certain digits and understatements for other digits. In 1960, further reduction in the overall age heaping has occurred, according to an index used by Myers (reference 13) based on a percentage distribution of ages by final digit for the population aged 23 to 99 years (see table 12 for 1960).

Table 12

Ending digit of age	Percent in digit group in 1960
All digits	100.0
0	9.9
1	9.9
2	9.9
3	9.8
4	10.1
5	10.3
6	9.9
7	10.1
8	9.8
9	10.3

The index is one-half the sum of the deviations from 10.0 percent, each taken without regard to sign. For 1960, this index is 0.8. Comparisons are made with other census years back to 1880 in table 13.

Table 13

Census years	Index of age-heaping
1880	10.4
1890	7.8
1900	4.7
1910	5.6
1920	4.5
1930	4.3
1940	3.0
1950	2.2
1960	0.8

In 1960 and in 1900, "date of birth" rather than "age as of last birthday" was the inquiry, and the improved question together with the extension of the trend, and opportunity for self-enumeration in 1960 may be largely responsible for less age heaping. The reduction in 1960 compared with 1950 occurred for both males and females in the white and nonwhite classifications (see table 14).

Table 14.--SUMMARY INDEX OF AGE HEAPING

(Population 23 to 82 years)

Total	White		Nonwhite	
	1960	1950	1960	1950
Male	0.5	1.6	2.6	5.6
Female	0.6	2.2	2.2	6.4

However, when the age heaping indexes are combined for ages ending in 0-4 and 5-9, to measure the effect of heaping on age statistics tabulated in the conventional 5-year groups (table 14a), no improvement is noted between the 1950 and 1960 censuses and no particular trend appears over the years.

Table 14a.--INDEX OF AGE HEAPING

(Combining 5-year age groups)

Census year	Index
1880	0.3
1890	0.3
1900	0.1
1910	0.3
1920	0.6
1930	0.4
1940	0.2
1950	0.3
1960	0.4

Response bias for various characteristics as measured by the intensive reinterview survey

Following the 1960 Census as in 1950, a "population content evaluation study" was conducted to obtain measures of response error with respect to selected items of information. In 1960, an intensive reinterview approach was used (Study EP-10) comparable in some ways with the Post Enumeration Study (PES) (reference 3) of 1950. In both cases, the interviewer was to obtain responses before consulting previously obtained census responses. Following the reinterview procedure for the samples reported on here, the interviewer was to compare the new response with the corresponding census entries, and where there were differences, an effort was made to determine the more accurate response or an improved response ("reconciliation").

As reported by Pritzker and Hanson (reference 14) study EP-10 used a more intensive interview procedure than the PES, interviewer training was more thorough, and the study was conducted after a shorter time following the census. For these and other reasons, EP-10 results (1960 are believed to be more accurate than PES results (1950). Thus the net differences observed in tables 19 through 23 at the end of this paper must be interpreted in the light of procedural differences in the two studies.

In considering the results shown in this series of tables, note that a larger bias for 1960 than 1950 could result from an improved reinterview study in 1960, a reduced accuracy in the census in 1960, or both. The data in tables 19-23 are consistent with the view that, on the whole, both census and reinterview procedures produced more accurate results in 1960 than in 1950. However, the findings of Study EP-10 are difficult to interpret, and the tables are offered for consideration and discussion. The results are based on person-to-person comparisons of responses and for any item include only cases for which responses were obtained in both the census and in the evaluation program.

In 1960, EP-10 is regarded as giving "preferred" results; in 1950, the PES gives the "preferred" results. For each classification of persons, the percentage in the classification according to the reinterview study is subtracted from the corresponding census result. A negative sign for "bias" indicates a lower census value. Where both 1950 and 1960 results are available, the differences in the absolute values of the relative biases are given, where a plus sign indicates a higher level of error in 1950 than in 1960.

The population characteristics for which bias measures are shown in some detail are in the following tables at the end of this paper:

Table 19.--Sex and color
Table 20.--Age by five-year classes
Table 21.--Mobility status
Table 22.--Educational attainment
Table 23.--School enrollment

EVIDENCE ON RESPONSE VARIANCE

Because of time and space limitations, this presentation does not discuss the technical development of measurement devices. Persons who are especially interested in theoretical considerations may consult references 14, 15, and 17 listed at the end of this paper.

We approach the estimation of response variance in two ways, the first of which is associated with repetitions of some defined phase of data collection or processing. This condition is approximated by the familiar concept of reinterview, or by "matching" information from two or more sources for identical individuals, or by the repetition of a processing operation such as the coding of the same interview data by different persons. This basic trial-to-trial average variability results in gross differences.

The intensive reinterview survey (EP-10) gave us two measurements for each person reinterviewed, for selected items of information. The diagram below illustrates the approach.

Reinterview survey	Census		
	In age class 40-44 years	Not in age class 40-44 years	Total
In age class 40-44 years...	a	b	a + b
Not in age class 40-44 years.........	c	d	c + d
Total.......	a + c	b + d	n = a + b + c + d

Where paired responses were identical, the "zero difference" made no contribution to the response variance (cells a and d). When the responses were different, the effect was to remove a person from one classification, and place him in another (cells b and c). For example, a person whose age in the census was reported as 43 years and as 45 years in the reinterview survey would be classified in the first case in the "age class 40-44 years," and in the second case "not in age class 40-44 years." This difference in response would result in an entry in cell c. Note that a minor change in response not resulting in a change in classification (or cell) would have, for present purposes, no statistical effect.

The sum (b + c) relative to the number of persons n, in both the original and the reinterview survey, (b + c)/n, is called the "gross difference rate," identified as "g."

The estimated "index of inconsistency," shown in the last columns of tables 19-23, and represented by I, makes use of the measure of "gross difference rate" in such a way as to make these estimates of inconsistency or response more comparable from one item to another. This is accomplished by dividing the gross difference rate by 2pq, where p is the proportion of the specified population that has the characteristic under consideration, and q = 1 - p.

That is
$$I = \frac{g}{2pq}$$

The denominator 2pq approximates the maximum value that the gross difference rate g can have under independent repetitions of a survey.

Some average values of I for sex, race, five-year age classes, mobility-status and educational attainment classes are shown in table 15. As for tables 19-23, table 15 makes use of parallel information from the PES of 1950 and Study EP-10 for 1960.

Table 15.--PRELIMINARY ESTIMATES OF THE AVERAGE INDEX OF INCONSISTENCY (I) FOR SELECTED POPULATION STATISTICS: 1960 AND 1950 CENSUSES OF POPULATION

Characteristic	Estimated unweighted average index per class		
	1960 Census (1)	1950 Census (2)	Difference (2)-(1) (3)
Sex........................	.018	n.a.	n.a.
Race.......................	.045	n.a.	n.a.
Five-year age classes.......	.054	.070	.016
Mobility-status classes[1]....	.120	.335	.215
Educational-attainment classes (population 25 years old and over only)...	.256	.394	.138

[1]Mobility during a one-year period in 1950 and during a five-year period in 1960.

In general, comparisons of the indexes of inconsistency show lower indexes for 1960 than for 1950. At first glance, this might seem to be a clear indication of a reduction in response variance in the 1960 Census. Actually the situation is not so simple; an improved evaluation program in 1960 would reduce the indexes of inconsistency, even with no changes in the quality of the census. It follows that reductions in the indexes of inconsistency might have occurred as a result of improvement of the intensive reinterview survey in 1960 as compared with 1950, or as a result of improvement in the 1960 Census as compared with 1950, or both. Efforts were made to achieve improvements in both the census and the evaluation program, and on the basis of this fact and other evidence there are reasons to believe that both of these factors influenced the reduction in the indexes of incosistency.

Another study common to 1960 and 1950, offering comparable estimates of indexes of inconsistency, is the Current Population Survey results matched in the census results (CPS-Census Match) (reference 14). As noted earlier, the CPS is regarded as the "preferred" method, compared to the census, for the collection of labor force information. In the CPS-Census Match, Stanley Greene brought together the data for a sample of about 8,000 households enumerated in both the April 1960 CPS and in the 1960 Census. Estimates of indexes of inconsistency for 1960, and the comparable CPS-Census match data for 1950, are shown in table 16.

Again it should be noted that often the census data have a different time reference from the CPS data. However, there is some reason to believe that the CPS was of about equivalent quality in 1950 and 1960. Column 3 of table 16 shows differences in the respective indexes of inconsistency, most of which are slightly favorable to the 1960 Census.

Although CPS reinterview in relation to original CPS results are somewhat irrelevant in a discussion of census results, it may be of interest to have a measure of the reliability of CPS data since they are used to evaluate census data (see table 17). Columns 7 and 8 show indexes of inconsistency for both reconciled and unreconciled results (reference 15). Indexes for the unreconciled results, which are more directly comparable to the CPS-Census match in which there was no reconciliation, are roughly half the values for I for 1960 shown in table 17.

The coding variance study (reference 16) was largely a by-product of the quality control scheme used in the 1950 census, using a sample of 1 in 40 households from the 25 percent sample for whom occupation and industry data were collected. Three different coding clerks with approximately the same training and coding experience all coded independently from the census schedule, but only one person, the "Census Coder," entered his code on the census schedule. The coded results were then matched. An index of inconsistency, analagous to the one previously described, made use of all three codes.

Table 16.--ESTIMATES OF THE "INDEX OF INCONSISTENCY" FOR LABOR-FORCE CLASSIFICATIONS IN THE 1960 AND 1950 CENSUSES OF POPULATION, FOR THE "IDENTICAL POPULATION" FOURTEEN YEARS OLD AND OVER, BY SEX

Sex and labor-force status	Index of inconsistency, I		Difference[1] (2) - (1)
	1960 Census	1950 Census	
	(1)	(2)	(3)
Males			
In the civilian labor force	.177	.205	+.028
Employed	.170	.196	+.026
(1) In agriculture	.224	.144	-.080
(2) In nonagricultural industries	.132	.140	+.008
Unemployed	.500	.513	+.013
Not in the civilian labor force	.177	.205	+.028
Females			
In the civilian labor force	.192	.195	+.003
Employed	.175	.180	+.005
(1) In agriculture	.593	.957	+.364
(2) In nonagricultural industries	.156	.145	-.011
Unemployed	.720	.751	+.031
Not in the civilian labor force	.192	.195	+.003

[1]Minus sign indicates greater unreliability in 1960 Census than in 1950 Census; plus sign indicates greater unreliability in 1950 Census.

Table 17.—PROPORTIONS OF PERSONS IN INDICATED CLASSIFICATIONS IN ORIGINAL AND REINTERVIEW SURVEYS (FOR IDENTICAL PERSONS) AND GROSS DIFFERENCE RATES RELATIVE TO 2 pq FOR RECONCILED AND UNRECONCILED RESULTS AS SPECIFIED

Survey and item classification	p_1: proportion in class in original survey (percent)		p_2: proportion in class in reinterview survey (percent)		g (percent)		g relative to 2pq[1] (I)	
	Unreconciled	Reconciled	Unreconciled	Reconciled	Unreconciled	Reconciled	Unreconciled	Reconciled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CPS, 1958-1961, status in labor force—								
In labor force	55.8	56.2	56.0	56.9	4.1	1.6	.08	.03
Employed	52.3	52.8	52.5	53.2	3.9	1.4	.08	.03
Agriculture	4.8	4.6	4.8	4.7	1.2	.4	.13	.05
Nonagriculture	47.6	48.1	47.7	48.5	3.5	1.4	.07	.03
Full time	36.1	36.7	36.4	36.5	4.4	1.5	.09	.03
Part time	9.1	9.0	9.2	9.6	4.6	1.7	.28	.10
With a job, not at work	2.4	2.4	2.2	2.4	1.4	.5	.32	.11
Not employed	3.4	3.5	3.5	3.6	1.9	.8	.29	.11
Not in labor force	44.2	43.8	44.0	43.2	4.1	1.6	.08	.03

[1]Here, p is defined as $(p_1 + p_2)/2$, and q is 1-p.

Table 18.—NUMBER OF INDUSTRY AND OCCUPATION CODES BY INDEX OF INCONSISTENCY

Index of inconsistency	Industry codes: Number of codes	Industry codes: Percent of codes	Industry codes: Estimated percent of labor force[1]	Occupation codes: Number of codes	Occupation codes: Percent of codes	Occupation codes: Estimated percent of labor force[1]
.001 - .100	59	39.3	73.6	142	48.0	74.4
.101 - .200	56	37.3	18.8	93	31.4	22.4
.201 - .300	23	15.3	6.6	41	13.9	2.6
.301 - .400	6	4.1	0.8	11	3.7	0.2
.401 - .500	2	1.3	0.1	4	1.3	0.3
More than .500	4	2.7	0.1	5	1.7	(2)
Total codes[3]	150	100.0	100.0	296[4]	100.0	100.0

[1]Estimates based on a sample of 420,000.
[2]Less than .05 percent.
[3]Includes the codes for "not report."
[4]Excludes Code 000 "Accountants and auditors" because of programming error.

Table 18 gives the distributions of the 149 industry codes and 296 occupation codes by size of the inconsistency index. A substantially larger proportion of occupation codes have low indexes of inconsistency than do industry codes. Forty-eight percent of the occupation codes as compared with 39 percent of the industry codes had indexes between .001 and .100. For both types of coding, the .001 to .100 class accounted for about 74 percent of the experienced civilian labor force.

A further study, made of the 20 percent or so codes having the highest indexes indicated that industry codes for wholesale trades were particularly troublesome.

For all classifications of the data discussed, it should be noted that gross differences which may be compensating in substantiating degree for simple means may have more significant effect when measuring relationships between classes of the population. The effect of gross differences on relationships should be a subject for further study.

Enumerator variance study

The special concern of this evaluation study was to develop estimates of the variance in census results attributable to enumerators and their immediate supervisors. It was anticipated that the heavy use of self-enumeration on sample items, and the consequent use of enumerators only in the follow-up program in 1960 would reduce the effect of the variability attributable to enumerators well below that of 1950. The preliminary results available from 24 out of 50 studies designed to measure enumerator variance support this conclusion. More definitive answers will be available on the completion of the 50 studies.

SUMMARY REMARKS AND CONCLUSIONS

Evaluation of the completeness of coverage and quality of measurement of a census is difficult. Attempts have been made to compare the relative quality of the 1960 and the 1950 Censuses. In addition, an effort has been made to compare each of these censuses with other sources of data, but at best such comparisons can provide only limited evidence. There is reason to believe that for some elements in the population it is becoming increasingly difficult to conduct a satisfactory enumeration. Substantial efforts to develop improved procedures are needed to keep up with changing conditions and to avoid deterioration from one census to the next.

On the whole, we conclude that there were improvements in the quality of the 1960 Census as compared with 1950. Some of the indicated improvements have been substantial and some have been minor. Publication of results has been much earlier than in the comparable period following 1950; costs were less than would have been indicated by the changes in price levels and the growth of the population, and there was an increase in the amount of information that was made available.

Users generally have been given more information concerning the quality of the published data than ever before. Not all of the gains that were hoped for were achieved and hindsight reveals a number of errors that one wishes had been avoided. For some areas and for some topics the rates of non-response are troublesome to the users of the data.

We are by no means complacent that we have achieved the quality of results that are needed or can be achieved. It is obvious that much remains to be desired in improved quality of censuses in both coverage and content. In the discussion with this and other groups, a great deal of attention should be concentrated on the effect of errors on the various uses to which census results are put, and on methods and studies that may result in increased accuracy of census results. We welcome the discussion and consideration that is offered by this meeting.

ACKNOWLEDGMENTS

Most of the evaluation program of the 1960 Censuses has been carried out under the direction of Joseph Steinberg, Chief, Statistical Methods Division. The planning has been the joint responsibility of Leon Pritzker, Chief, Response Research Branch, Statistical Research Division, Joseph Steinberg, and numerous other staff members, under the general direction of William N. Hurwitz, Chief, Statistical Research Division. Many persons in the Bureau played important parts in collecting the data and reducing it to manageable form. Their contributions are too many and varied to acknowledge here. The principal work of drafting this paper was done by Marie D. Wann, Mathematical Statistician. Henry S. Shryock, Jr., Assistant Chief (Program Development), Population Division, drafted the section on comparative coverage of the 1960 and 1950 Censuses, and has provided helpful comments and advice, as have many others who have participated in the various aspects of the evaluation program.

Table 19.--ESTIMATES OF BIAS IN THE STATISTICS AND OF THE "INDEX OF INCONSISTENCY" FOR SEX AND COLOR IN THE 1960 CENSUS OF POPULATION, FOR THE "IDENTICAL POPULATION"

Characteristic and category	1960 Census (percent)	Bias[1] x 100	Relative bias x 100	I
	(1)	(2)	(3)	(4)
Sex				
Male	49.3	+0.2	+0.4	.018
Female	50.7	-0.2	-0.4	.018
Color				
White	88.6	+0.2	+0.2	.045
Nonwhite	11.4	-0.2	-1.7	.045

NOTE: See section VII A of text for explanation of "Index of Inconsistency".
[1]Minus sign indicates understatement in Census; plus sign indicates overstatement.

Table 20.--ESTIMATES OF BIAS IN THE STATISTICS AND OF THE "INDEX OF INCONSISTENCY" FOR FIVE-YEAR AGE CLASSES IN THE 1960 AND 1950 CENSUSES OF POPULATION, FOR THE "IDENTICAL POPULATION"

Age class	Percentage distribution		Bias x 100[1]		Relative bias x 100			Index of inconsistency, I		
	1960 Census	1950 Census	1960 Census	1950 Census	1960 Census	1950 Census	Difference[2] |(6)|-|(5)|	1960 Census	1950 Census	Difference[2] (9) - (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0-4	11.3	10.7	+.01	-.18	+.06	-1.63	+1.57	.020	.025	+.005
5-9	10.4	8.8	+.02	+.08	+.16	+.92	+.76	.029	.028	-.001
10-14	9.4	7.4	+.05	+.01	+.47	+.11	-.36	.024	.034	+.010
15-19	7.4	7.1	-.07	+.11	-1.00	+1.64	+.64	.029	.040	+.011
20-24	6.0	7.6	-.04	+.02	-.79	+.26	-.53	.037	.051	+.014
25-29	6.1	8.1	+.08	...	+1.53	-.03	-1.50	.036	.062	+.026
30-34	6.7	7.6	-.03	+.04	-.49	+.48	-.01	.043	.076	+.033
35-39	7.0	7.5	+.12	+.06	+1.85	+.78	-1.07	.058	.075	+.017
40-44	6.5	6.8	+.03	+.09	+.44	+1.38	+.94	.078	.088	+.010
45-49	6.1	6.0	-.12	...	-1.85	-.07	-1.78	.071	.101	+.030
50-54	5.4	5.5	+.03	+.02	+.59	+.30	-.29	.078	.112	+.034
55-59	4.7	4.8	+.10	-.16	+2.11	-3.11	+1.00	.063	.103	+.040
60-64	4.0	4.0	-.10	-.04	-2.77	-1.04	-1.73	.098	.084	-.014
65-69	3.5	3.3	+.09	-.02	+2.63	-.52	-2.11	.078	.090	+.012
70-74	2.6	2.3	-.11	...	-.40	+.12	-.28	.095	.095	...
75 and over	3.1	2.6	-.05	-.03	-1.80	-1.07	-.73	.032	.051	+.019

NOTE: See section VII A of text for explanation of "Index of Inconsistency".
[1]Minus sign indicates understatement in Census; plus sign indicates overstatement.
[2]Minus sign indicates higher level of error in 1960 Census than in 1950 Census; plus sign indicates higher level of error in 1950 Census.

Table 21.--ESTIMATES OF THE BIAS IN THE STATISTICS AND OF THE "INDEX OF INCONSISTENCY" FOR MOBILITY-STATUS CLASSES IN THE 1960 AND 1950 CENSUSES OF POPULATION, FOR THE "IDENTICAL POPULATION" FIVE YEARS OLD AND OVER IN 1960 AND ONE YEAR OLD AND OVER IN 1950

Mobility-status classes[1]	Percent in class		Bias x 100[2]		Relative bias[3] x 100			Index of inconsistency, I		
	1960 Census	1950 Census	1960 Census	1950 Census	1960 Census	1950 Census	Difference[3] |(6)|-|(5)|	1960 Census	1950 Census	Difference[3] (9) - (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Same house............	50.7	82.6	+1.4	+0.3	+2.6	+0.3	-2.3	.072	.223	+.151
Different house, same county...............	30.3	11.4	+0.3	-0.7	+1.2	-5.8	+4.6	.125	.260	+.135
Different county, same State................	8.7	3.0	-0.7	+0.1	-7.6	+2.3	-5.3	.108	.274	+.166
Different State........	9.0	2.7	-0.9	+0.2	-11.2	+8.3	-2.9	.107	.336	+.229
Abroad.................	1.3	0.4	-0.2	+0.2	-13.3	+278.3	+265.0	.187	.584	+.397

NOTE: See section VII A of text for explanation of "Index of Inconsistency."
[1]Residence five years prior to the Census data for the 1960 Census; residence one year prior to the Census data for the 1950 Census.
[2]Minus sign indicates understatement in Census; plus sign indicates overstatement.
[3]Minus sign indicates higher level of error in 1960 Census than in 1950 Census; plus sign indicates higher level of error in 1950 Census.

Table 22.--ESTIMATES OF BIAS IN THE STATISTICS AND OF THE "INDEX OF INCONSISTENCY" FOR EDUCATIONAL ATTAINMENT CLASSES IN THE 1960 AND 1950 CENSUSES OF POPULATION, FOR THE "IDENTICAL POPULATION" TWENTY-FIVE YEARS OLD AND OVER

Educational attainment class	Percentage distribution		Bias x 100[1]		Relative bias x 100			Index of inconsistency, I		
	1960 Census	1950 Census	1960 Census	1950 Census	1960 Census	1950 Census	Difference[2] |(6)|-|(5)|	1960 Census	1950 Census	Difference[2] (9) - (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
None..............	2.3	2.6	-0.01	-1.0	-0.7	-29.0	+28.3	.238	.554	+.316
Elementary, 1-4 years....	6.1	8.5	-0.5	+0.3	+8.5	+3.6	-4.9	.309	.360	+.051
Elementary, 5-6 years....	7.5	9.4	-0.8	-0.6	-11.0	-5.7	-5.3	.333	.479	+.146
Elementary, 7 years......	6.4	7.0	-0.8	-1.2	-11.1	-14.3	+3.2	.399	.604	+.205
Elementary, 8 years......	17.5	20.8	+0.7	+1.4	+4.6	+7.3	+2.7	.300	.400	+.100
High school, 1-3 years...	19.2	17.4	+0.7	-0.7	+3.6	-4.0	+0.4	.240	.375	+.135
High school, 4 years.....	24.6	20.7	-0.5	+0.3	-1.9	+1.3	-0.7	.186	.263	+.077
College, 1-3 years.......	8.8	7.3	+1.0	+1.0	+11.4	+15.0	+3.6	.224	.339	+.115
College, 4 or more years.	7.7	6.2	+0.2	+0.5	+3.1	+8.7	+5.6	.074	.170	+.096

NOTE: See section VII A of text for explanation of "Index of Inconsistency".
[1]Minus sign indicates understatement in Census; plus sign indicates overstatement.
[2]Minus sign indicates higher level of error in 1960 Census than in 1950 Census; plus sign indicates higher level of error in 1950 Census.

Table 23.--ESTIMATES OF BIAS IN THE STATISTICS AND OF THE "INDEX OF INCONSISTENCY" FOR SCHOOL ENROLLMENT CLASS IN THE 1960 CENSUS OF POPULATION, FOR THE "IDENTICAL POPULATION" 5 TO 34 YEARS OLD

School enrollment class	Percent of total population in class	Bias x 100[1]	Relative bias x 100	Index of inconsistency, I
Kindergarten and elementary..	37.8	-1.0	-2.2	.038
High school.................	11.8	-0.1	-1.1	.096
College.....................	3.6	+0.4	+14.9	.158
Not enrolled................	46.9	+0.7	+1.9	.041

NOTE: See section VII A of text for explanation of "Index of Inconsistency".
[1]Minus sign indicates understatement in Census; plus sign indicates overstatement.

References

(1) Brunsman, Howard G., "Processing and Editing the Data from the 1960 Census of Population," Presented at the May 1960 meetings of the Population Association of America in Washington, D. C.

(2) U. S. Bureau of the Census, U.S. Census of Population: 1960, General Population Characteristics, United States Summary, Final Report PC(1)-1B, p. XVII, U. S. Government Printing Office, Washington, 1961.

(3) U. S. Bureau of the Census, The Post-Enumeration Survey: 1950, Technical Paper No. 4, Washington, D. C., 1960.

(4) Steinberg, Joseph; Margaret Gurney; and Walter Perkins, "The Accuracy of the 1960 Census Count," In: American Statistical Association, Proceedings of the Social Statistics Section: 1962, pp. 76-79, Washington, D. C. (Note that later figures are used in the present paper.)

(5) Akers, Donald S., "Estimating Net Census Undercount in 1960 Using Analytical Techniques," Presented at the Annual Population Association of America meetings on May 5, 1962.

(6) U. S. Bureau of the Census, U.S. Census of Population: 1960, Number of Inhabitants, United States Summary, Final Report PC(1)-1A, p. 11, U.S. Government Printing Office, Washington, D. C., 1961.

(7) Shryock, Henry S., Jr. and Stanley Greene, "Not Reported Rate for 'Occupation' by Single- and Two-Stage Cities," U. S. Bureau of the Census, memorandum to Dr. Conrad Taeuber, December 18, 1962.

(8) U. S. Bureau of the Census, U.S. Census of Population: 1960, General Population Characteristics, United States Summary, Final Report PC(1)-1B, Table B-1, U. S. Government Printing Office, Washington, D. C., 1961.

(9) President's Committee to Appraise Employment and Unemployment Statistics, "Measuring Employment and Unemployment," Table J.1, U. S. Government Printing Office, 1962.

(10) Nam, Charles B., "Some Comparisons of Office of Education and Census Bureau Statistics on Education," In: American Statistical Association, Proceedings of the Social Statistics Section: 1962, pp. 258-269, Washington, D. C.

(11) Miller, Herman, unpublished memorandum (to be published as a monograph), U. S. Bureau of the Census.

(12) Bradshaw, Benjamin S. and Donald S. Akers, "Age Heaping in the 1960 Census of Population," unpublished report, U. S. Bureau of the Census.

(13) Myers, Robert J., "Accuracy of Age Reporting in the 1950 United States Census," Journal of the American Statistical Association, Vol. 49, No. 268, p. 828, December 1954.

(14) Pritzker, Leon and Robert Hanson, "Measurement Errors in the 1960 Census of Population," In: American Statistical Association, Proceedings of the Social Statistics Section: 1962, pp. 80-90, Washington, D. C.

(15) U. S. Bureau of the Census, The Current Population Survey Reinterview Program: Some Notes and Discussion, Technical Paper No. 6, U. S. Government Printing Office, Washington, D. C., 1963.

(16) Fasteau, Herman H.; J. Jack Ingram; and Ruth H. Mills, "Study of the Reliability of Coding of Census Returns," In: American Statistical Association, Proceedings of the Social Statistics Section: 1962, pp. 104-115, Washington, D. C.

V

BAYESIAN INFERENCE: AN EXPOSITORY SESSION

Chairman, Howard L. Jones, University of Chicago

APPLICATION OF BAYESIAN INFERENCE
(Abstract)

George C. Tiao, University of Wisconsin

Frequently the distribution of observations $\underline{y}$ depends not only upon a set of parameters $\underline{\xi}_1$ of interest, but also on a set of nuisance parameters $\underline{\xi}_2$. In judging the sensitivity of inference about the parameters of interest relative to assumptions about the model such as normality and independence, the nuisance parameters can be measures of departure from normality and independence. From a Bayesian point of view, the posterior distribution $p(\underline{\xi}_1 | \underline{\xi}_2 = \underline{\xi}_{20}, \underline{y})$ indicates the nature of inference about $\underline{\xi}_1$ if the corresponding assumptions $\underline{\xi}_2 = \underline{\xi}_{20}$ about the model are made, while the posterior density $p(\underline{\xi}_2 = \underline{\xi}_{20} | \underline{y})$ reflects the plausibility of such assumptions. The marginal posterior distribution of $\underline{\xi}_1$ obtained by integrating out $\underline{\xi}_2$,

$$p(\underline{\xi}_1 | \underline{y}) = \int_R p(\underline{\xi}_1 | \underline{\xi}_2, \underline{y}) \; p(\underline{\xi}_2 | \underline{y}) \; d\underline{\xi}_2, \text{ thus}$$

indicate the overall inference about $\underline{\xi}_1$ when proper consideration is given to the various possible assumptions. This approach is illustrated in detail by two examples; one concerns the comparison of two variances where the parent distributions are assumed to be members of a class of non-normal distributions, the other concerns making inferences about the coefficients in a regression model where the disturbances are autocorrelated.

BAYESIAN INFERENCE*

Harry V. Roberts, University of Chicago

Bayesian inference or Bayesian statistics is an approach to statistical inference based on the theory of subjective probability. A formal Bayesian analysis leads to probabilistic assessments of the object of uncertainty. For example, a Bayesian inference might be: "The probability is .95 that the mean of a normal distribution lies between 12.1 and 23.7." The number .95 represents a degree of belief, either in the sense of "subjective probability consistent" or "subjective probability rational" (see PROBABILITY: PHILOSOPHY AND INTERPRETATION, which should be read in conjunction with the present article); .95 need not and typically does not correspond to any "objective" long-run relative frequency. Very roughly, a degree of belief of .95 can be interpreted as betting odds of 95 to 5 or 19 to 1. A degree of belief is always potentially a basis for action; for example, it may be combined with utilities by the principle of maximization of expected utility (see STATISTICAL DECISION THEORY).

By contrast, the traditional or "classical" approach to inference leads to probabilistic statements about the method by which a particular inference is obtained. Thus a classical inference might be: "A .95 confidence interval for the mean of a normal distribution extends from 12.1 to 23.7." The number .95 here represents a long-run relative frequency, namely the frequency with which intervals obtained by the method that resulted in the present interval would in fact include the unknown mean. (It is not to be inferred from the fact that we used the same numbers, .95, 12.1, and 23.7, in both illustrations that there will necessarily be a numerical coincidence between the two approaches.)

The term "Bayesian" arises from an elementary theorem of probability theory, named after the Rev. Thomas Bayes, an English clergyman of the 18th century, who first enunciated it and proposed its use in inference. Bayes' theorem is typically used in the process of making Bayesian inferences, as will be explained below. For a number of historical reasons, however, current interest in Bayesian inference is quite recent--dating, say, from the 1950's. Hence the term "neo-Bayesian" is sometimes used instead of "Bayesian."

An Illustration of Bayesian Inference

For a simple illustration of the Bayesian approach, consider the problem of making inferences about a Bernoulli process with parameter p. A Bernoulli process can be visualized in terms of repeated independent tosses of a not-necessarily "fair" coin. It generates "heads" and "tails" in such a way that the conditional probability of heads on a single trial is always equal to a parameter p regardless of the previous history of heads and tails.

Suppose first that we have no direct sample evidence from the process. Based on experience with similar processes, introspection, general knowledge, etc., we may be willing to translate our judgments about the process into probabilistic terms. For example, we might assess a (subjective) probability distribution for $\tilde{p}$ (the tilde "$\sim$" indicates that we are now thinking of the parameter p as a random variable). Such a distribution is called a prior distribution because it is usually assessed prior to sample evidence. Purely for illustration, suppose that the prior distribution of $\tilde{p}$ is uniform on the interval from 0 to 1: the probability that $\tilde{p}$ lies in any subinterval is that subinterval's length, no matter where the subinterval is located between 0 and 1. Now suppose that we observe heads, heads, and tails on three tosses of a coin. The probability of observing this sample, conditional on p, is

$$p^2(1-p) \ .$$

If we regard this expression as a function of p, it is called the <u>likelihood function</u> of the sample. Bayes' theorem shows how to use the likelihood function in conjunction with the prior distribution to obtain a revised or <u>posterior</u> distribution of $\tilde{p}$. "Posterior" means after the sample evidence, and the posterior distribution represents a reconciliation of sample evidence and prior judgment. In terms of inferences about $\tilde{p}$, we may write Bayes' theorem in words as

Posterior probability (density) at p, given the observed sample =

$$\frac{\text{Prior probability (density) at } p \times \text{likelihood}}{\text{Prior probability of the observed sample}}$$

Expressed mathematically,

$$f''(p|r,n) = \frac{f'(p)\ p^r(1-p)^{n-r}}{\int_0^1 f'(p)\ p^r(1-p)^{n-r}dp} \ ,$$

where $f'(p)$ denotes the prior density of $\tilde{p}$, $p^r(1-p)^{n-r}$ denotes the likelihood if r heads are observed in n trials, and $f''(p|r,n)$ denotes the posterior density of $\tilde{p}$ given the sample evidence.

*A revised version of this paper will appear in the (forthcoming) International Encyclopedia of the Social Sciences.

In our example, $f'(p) = 1$, $(0 \le p \le 1)$, $r = 2$, $n = 3$, and

$$\int_0^1 f'(p)\ p^r(1-p)^{n-r}dp = \int_0^1 p^2(1-p)dp$$
$$= 1/12\ ,$$

so

$$f''(p|r = 2,\ n = 3) = 12\ p^2(1-p),\quad 0 \le p \le 1$$
$$= 0 \quad \text{otherwise.}$$

Thus we emerge from the analysis with an explicit probability distribution for $\tilde{p}$. This distribution characterizes fully our judgments about $\tilde{p}$. It could be applied in a formal decision-theoretic analysis in which utilities of alternative acts are functions of p. For example, we might make a Bayesian point estimate of p (each possible point estimate is regarded as an act), and the seriousness of an estimation error ("loss") might be proportional to the square of the error. The best point estimate can then be shown to be the mean of the posterior distribution; in our example, this would be .6. Or, we might wish to describe certain aspects of the posterior distribution for summary purposes; it can be shown, for example, that

$$P(\tilde{p}<.194) = .025 \quad \text{and} \quad P(\tilde{p}>.932) = .025\ ,$$

so a .95 "credible interval" for $\tilde{p}$ extends from .194 to .932. Again, it can be shown that $P(\tilde{p}>.5) = .688$: the posterior probability that the coin is "biased" in favor of heads is a little over 2/3.

The Likelihood Principle

In our example, the effect of the sample evidence was wholly transmitted by the likelihood function. All we needed to know from the sample was $p^r(1-p)^{n-r}$; the actual sequence of individual observations was irrelevant so long as we believed the assumption of a Bernoulli process. In general, a full Bayesian analysis requires as inputs for Bayes' theorem only the likelihood function and the prior distribution. Thus the import of the sample evidence is fully reflected in the likelihood function, a principle known as the likelihood principle (see also LIKELIHOOD). Alternatively, given that the sample is drawn from a Bernoulli process, the import of the sample is fully reflected in the numbers r and n, which are called sufficient statistics (see SUFFICIENCY).

The likelihood principle implies certain consequences that do not accord with traditional ideas. Here are examples: (1) Once the data are in, there is no distinction between sequential analysis and analysis for fixed sample size. In the Bernoulli example, successive samples of n_1 and n_2 with r_1 and r_2 successes could be analyzed as one pooled sample of $n_1 + n_2$ trials with $r_1 + r_2$ successes. Alternatively, a posterior distribution could be computed after the first sample of n_1; this distribution could then serve as a prior distribution for the second sample; finally, a second posterior distribution could be computed after the second sample of n_2. By either route the posterior distribution after $n_1 + n_2$ observations would be the same. Under almost any situation that is likely to arise in practice, the "stopping rule" by which sampling is terminated is irrelevant to the analysis of the sample. For example, it would not matter whether r successes in n trials were obtained by fixing r in advance and observing the r^{th} success on the n^{th} trial, or by fixing n in advance and counting r successes in the n trials. (2) For the purpose of statistical reporting, the likelihood function is the important information to be conveyed. If a reader wants to perform his own Bayesian analysis, he needs the likelihood function, not a posterior distribution based on someone else's prior, nor traditional analyses such as significance tests, from which it may be difficult or impossible to recover the likelihood function.

Vagueness about Prior Probabilities

In our example we assessed the prior distribution of $\tilde{p}$ as a uniform distribution from 0 to 1. It is sometimes thought that such an assessment means that we "know" $\tilde{p}$ is so distributed, and that our claim to knowledge might be verified or refuted in some way. It is indeed possible to imagine situations in which the distribution of $\tilde{p}$ might be known, as when one coin is to be drawn at random from a number of coins, each of which has a "known" p determined by a very large number of tosses. The frequency distribution of these p's would then serve as a prior distribution, and all statisticians would apply Bayes' theorem in analyzing sample evidence. But such an example would be unusual. Typically, in making an inference about $\tilde{p}$ for a particular coin, the prior distribution of $\tilde{p}$ is not a description of some distribution of p's but rather a tool for expressing judgments about $\tilde{p}$ based on evidence other than the evidence of the particular sample to be analyzed.

Not only do we rarely "know" the prior distribution of $\tilde{p}$, but we are typically more or less vague when we try to assess it. This vagueness is comparable to the vagueness that surrounds many decisions in everyday life. For example, a person may decide to offer $21,250 for a house he wishes to buy, even though he may be quite vague about what amount he "should" offer. Similarly, in statistical inference we may assess a prior distribution in the face of a certain amount of vagueness. If we are not willing to do so, we cannot pursue a formal Bayesian analysis and must evaluate sample evidence intuitively, perhaps aided by the tools of descriptive statistics and classical inference.

Vagueness about prior probabilities is not the only kind of vagueness to be faced in statistical analysis, and the other kinds of vagueness are equally troublesome for approaches to statistics that do not use prior probabilities. Vagueness about the likelihood function, that is, the process generating the data, is typically substantial and hard to deal with. Moreover, both classical and Bayesian decision theory bring in the idea of utility, and utilities often are vague.

In assessing prior probabilities, skillful self-interrogation is needed in order to mitigate vagueness. Self-interrogation may be made more systematic and illuminating in several ways. (1) Direct judgmental assessment. In assessing the prior distribution of $\tilde{p}$, for example, we might ask, "For what p would we be indifferent to an even money bet that $\tilde{p}$ is above or below this value?" (Answer is the .50-fractile or median.) Then, "If we were told that $\tilde{p}$ is above the .50-fractile just assessed, but nothing more, for what value of p would we now be indifferent in such a bet?" (Answer is the .75-fractile.) Similarly we might locate other key fractiles, or key relative heights on the density function. (2) Translation to equivalent but hypothetical prior sample evidence. For example, we might feel that our prior opinion about $\tilde{p}$ is roughly what it would have been if we had initially held a uniform prior, and then seen r heads in n hypothetical trials from the process. The implied posterior distribution would serve as the prior. (3) Contemplation of possible sample outcomes. Sometimes we may find it easy to decide directly what our posterior distribution would be if a certain hypothetical sample outcome were to materialize. We can then work backwards to see the prior distribution thereby implied. Of course, this approach is likely to be helpful only if the hypothetical sample outcomes are easy to assimilate. For example, if we make a certain technical assumption about the general shape of the prior distribution (beta distribution), the answers to the following two simply-stated questions imply a prior distribution of $\tilde{p}$: (a) How do we assess the probability of heads on a single trial? (b) If we were to observe a head on a single trial (this is the hypothetical future outcome) how would we assess the probability of heads on a second trial?

These approaches are intended only to be suggestive. If several approaches to self-interrogation lead to substantially different prior distributions, we must either try to remove the internal inconsistency or be content with an intuitive analysis. Actually, from the point of view of "subjective probability consistent," the discovery of internal inconsistency in one's judgments is the only route toward more "rational" decisions. The danger is not that internal inconsistencies will be revealed but that they will be suppressed by self-deception or glossed over by lethargy.

It may happen that vagueness affects only unimportant aspects of the prior distribution: theoretical or empirical analysis may show that the posterior distribution is insensitive to these aspects of the distribution. For example, we may be vague about many aspects of the prior distribution, yet feel that it is nearly uniform over all values of the parameter for which the likelihood function is not essentially zero. This has been called a diffuse, informationless, or locally-uniform prior distribution. These terms are to be interpreted relative to the spread of the likelihood function, which depends on the sample size; a prior that is diffuse relative to a large sample may not be diffuse relative to a small one. If the prior distribution is diffuse, the posterior distribution can be easily approximated from the assumption of a strictly uniform prior distribution. The latter assumption, known historically as Bayes' postulate (not to be confused with Bayes' theorem), is regarded mainly as a device that leads to good approximations in certain circumstances, although supporters of "subjective probability rational" sometimes regard it as more than that in their approach to Bayesian inference. The uniform prior is also useful for statistical reporting, since it leads to posterior distributions from which the likelihood is easily recovered and presents the results in a form readily usable to any reader whose prior distribution is diffuse.

Probabilistic Prediction

A distribution, prior or posterior, of the parameter $\tilde{p}$ of a Bernoulli process implies a probabilistic prediction for any future sample to be drawn from the process, assuming that the stopping rule is given. For example, the denominator in the right hand side of the Bayes' formula for Bernoulli sampling (p. 3) can be interpreted as the probability of obtaining the particular sample actually observed, given the prior distribution of $\tilde{p}$. While a person's subjective probability distribution of $\tilde{p}$ cannot be said to be "right" or "wrong," there are better and worse subjective distributions, and the test is predictive accuracy. Thus if Mr. A and Mr. B each has a distribution for $\tilde{p}$, and a new sample is then observed, we can calculate the probability of the sample in the light of each prior distribution. The ratio of these probabilities, technically a marginal likelihood ratio, measures the extent to which the data favor A over B or vice-versa. This idea has important consequences for evaluating judgments and selecting statistical models.

In connection with the previous paragraph a separate point is worth making. The posterior distributions of A and B are bound to grow closer together as sample evidence piles up, so long as neither of the priors was dogmatic. An example of a dogmatic prior would be the opinion that $\tilde{p}$ is exactly .5.

Multivariate Inference and Nuisance Parameters

Thus far we have used one basic example, inferences about a Bernoulli process. To introduce some additional concepts, we now turn to inferences about the mean μ of a normal distribution with unknown variance σ^2. In this case we begin with a joint prior distribution for $\tilde{\mu}$ and $\tilde{\sigma}^2$. The likelihood function is now a function of two variables, μ and σ^2. An inspection of the likelihood function will show not only that the sequence of observations is irrelevant to inference, but also that the magnitudes are irrelevant except insofar as they help determine the sample mean $\bar{x}$ and variance s^2, which, along with the sample size n, are the sufficient statistics of this example (see SUFFICIENCY). The prior distribution combines with the likelihood essentially as before except that a double integration (or double summation) is needed instead of a single integration (or summation). The result is a joint posterior distribution of $\tilde{\mu}$ and $\tilde{\sigma}^2$.

If we are interested only in $\tilde{\mu}$, then σ^2 is said to be a nuisance parameter. In principle it is simple to deal with a nuisance parameter: we "integrate it out" of the posterior distribution. In our example this means that we must find the marginal distribution of $\tilde{\mu}$ from the joint posterior distribution of $\tilde{\mu}$ and $\tilde{\sigma}^2$.

Multivariate problems and nuisance parameters can always be dealt with by the approach just described. The integrations required may demand heavy computation, but the task is straightforward. A more difficult problem is that of assessing multivariate prior distributions, and research is needed to find better techniques for overcoming the problems presented by vagueness in such assessments.

Design of Experiments and Surveys

So far we have talked only about problems of analysis of samples, without saying anything about what kind of sample evidence, and how much, should be sought. This kind of problem is known as a problem of design. A formal Bayesian solution of a design problem requires that we look beyond the posterior distribution to the ultimate decisions that will be made in the light of this distribution: the best design depends on the purposes to be served by collecting the data. Given the specific purpose and the principle of maximization of expected utility, it is possible to calculate the expected utility of the best act for any particular sample outcome. We can repeat this for each possible sample outcome for a given sample design. Next, we can weight each such utility by the probability of the corresponding outcome in the light of the prior distribution. This gives an overall expected utility for any proposed design. Finally, we pick the sample design with the highest expected utility. For two-action problems--e.g., deciding whether a new medical treatment is better or worse than a standard treatment--this procedure is in no conflict with the traditional approach of selecting designs by comparing operating characteristics, although it formalizes certain things--prior probabilities and utilities--that often are treated intuitively in the traditional approach.

Comparison of Bayesian and Classical Inference

Certain common statistical practices are subject to criticism either from the point of view of Bayesian or of classical theory: for example, estimation problems are frequently regarded as tests of null hypotheses, and .05 or .01 significance levels are used inflexibly. Bayesian and classical theory are in many respects closer to each other than either is to everyday practice. In comparing the two approaches, therefore, we shall confine the discussion to the level of underlying theory. In one sense the basic difference is the acceptance of subjective probability judgment as a formal component of Bayesian inference. This does not mean that classical theorists would disavow judgment, only that they would apply it informally after the "purely statistical" analysis is finished: judgment is the "second span in the bridge of inference." Building on subjective probability, Bayesian theory is a unified theory, whereas classical theory is diverse and ad hoc. In this sense

Bayesian theory is simpler. But in another sense Bayesian theory is more complex because it incorporates more into the formal analysis. Consider a famous controversy of classical statistics, the problem of comparing the means of two normal distributions with possibly unequal and unknown variances (the so-called "Behrens-Fisher" problem). Conceptually this problem poses major difficulties for some classical theories (not Fisher's fiducial inference; see FIDUCIAL INFERENCE), but none for Bayesian theory. In application, however, the Bayesian approach faces the problem of assessing a prior distribution involving four random variables. Moreover, there may be messy computational work after the prior distribution has been assessed.

In many applications, however, a credible interval emerging from the assumption of a diffuse prior distribution is identical or nearly identical to the corresponding confidence interval. There is a difference of interpretation, illustrated in the opening two paragraphs of this article, but in practice many people interpret the classical result in the Bayesian way. There often are numerical similarities between the results of Bayesian and classical analyses of the same data, but there can also be substantial differences, for example, when the prior distribution is non-diffuse and when a genuine null hypothesis is to be tested.

Often it may happen that the problem of vagueness, discussed at some length above, makes a formal Bayesian analysis seem unwise. In this event Bayesian theory may still be of some value in selecting a descriptive analysis or a classical technique that conforms well to the general Bayesian approach, and perhaps in modifying the classical technique. For example, many of the classical developments in sample surveys and analysis of experiments can be given rough Bayesian interpretations when vagueness about the likelihood (as opposed to prior probabilities) prevents a full Bayesian analysis. Moreover, even an abortive Bayesian analysis may contribute insight into a problem.

Bayesian inference has as yet received much less theoretical study than has classical inference. It is hard at this writing to predict how far Bayesian theory will lead in modification and reinterpretation of classical theory. Before a fully Bayesian replacement is available there is certainly no need to discard those classical techniques that seem roughly compatible with the Bayesian approach; indeed, many classical techniques are, under certain conditions, good approximations to fully Bayesian ones. In the meanwhile, the interaction between the two approaches promises to lead to fruitful developments in statistical inference, and the Bayesian approach promises to illuminate a number of problems--such as allowance for selectivity--that are otherwise hard to cope with.

A Few Suggestions for Further Reading

The first book-length development of Bayesian inference, which emphasizes heavily the decision-theoretic foundations of the subject, is Robert Schlaifer, Probability and Statistics for Business Decisions (New York: McGraw-Hill Book Company, Inc., 1959). A more technical development of the subject is given by Howard Raiffa and Robert Schlaifer, Applied Statistical Decision Theory (Boston: Division of Research, Graduate School of Business Administration, Harvard University, 1961). An excellent short introduction with an extensive bibliography is Leonard J. Savage, "Bayesian Statistics," in Robert E. Machol and Paul Gray, eds., Recent Developments in Information and Decision Processes (New York: The Macmillan Company, 1962). An interesting application of Bayesian inference is given, along with a penetrating discussion of underlying philosophy and a comparison with the corresponding classical analysis, in Frederick Mosteller and David L. Wallace, "Inference in an Authorship Problem," Journal of the American Statistical Association (Vol. 58, 1963), pp. 275-310. A fuller description of this study will be found in Mosteller and Wallace, Inference and Disputed Authorship: the Federalist Papers. (Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., in press.) This study gives a specific example of how one might cope with vagueness about the likelihood function. Another example is to be found in George E. P. Box and George C. Tiao, "A Further Look at Robustness via Bayes' Theorem," Biometrika (Vol. 49, 1962), pp. 419-432. A thorough development of Bayesian inference from the viewpoint of "subjective probability rational" is to be found in Harold Jeffreys, Theory of Probability (Oxford: Clarendon Press, 3rd edition, 1961).

Discussion

Arthur Schleifer Jr., Dartmouth College

I should first like to congratulate Professors Roberts and Tiao for their excellent papers. Bayesian statistics owes a great debt to expositors like Harry Roberts whose paper is a clear, concise and well-balanced summary of many difficult and controversial problems; and George Tiao's paper, along with some of his previous publications, should appeal even to those who have not been attracted to Bayesian statistics by idealogical arguments, for it dramatically shows the power of the Bayesian approach in handling departures from the idealized assumptions of classical statistics. Although I am in basic agreement with almost every point made by Professors Roberts and Tiao, I shall devote this discussion to a difference of opinion regarding emphasis which in no way dampens my high regard for both papers.

In a session on Bayesian inference, I find it a bit surprising that neither paper attempts to distinguish formally between inference and decision theory. I must confess a certain uneasiness in trying to ascertain where the dividing line between these two areas lies, but I think the task is worthwhile because it forces us to consider what sorts of assessments of the object of uncertainty are useful.

In particular I should like to question the role in Bayesian inference of certain summary measures of posterior probability distributions to which Professor Roberts makes reference--for example, Bayesian point estimates, credible intervals, and the Bayesian analogue of significance tests. I shall argue that these measures either sacrifice experimental information or fail to process that information sufficiently, and that a tendency to summarize experimental results in terms of these measures can result in reports, such as one given in Professor Tiao's paper, which fail to serve the needs of the decision-maker as well as they might. Professor Tiao's example was, I am sure, intended to illustrate the application of a technique, and not to indicate how inferences should be made in the context of a real problem; but the possibility of misinterpretation by the casual reader is very real, and it is against this possible misinterpretation that I wish to argue.

Problems to which Bayesian analysis is applicable can be partitioned into an <u>experimental</u> phase and a <u>decision-making</u> phase, the output of the experimenter serving as input for the decision maker. No matter how "purely scientific" an experimenter may be, he must recognize that the report of his experimental results may some day be used in a decision context. At the other end of the spectrum, a rational decision maker, when acting under uncertainty, must take account of whatever objective experimental evidence is available and whatever subjective or informal evidence he may have acquired elsewhere.

Given these two phases of problem analysis, we may specify that the output of Bayesian decision theory is a <u>course of action</u>, while the output of Bayesian inference is a <u>report</u> which serves as input for the decision-maker. It is generally agreed that such a report should <u>summarize</u> the experimental data and <u>process</u> the data in the form most useful to the decision-maker, but that <u>under no circumstances should the report sacrifice information</u>.

The prescription for fulfilling these goals depends on the gap between experimenter and decision-maker. If the experimenter cannot even anticipate the decision-maker's data-generating model, then it is best simply to report the raw experimental data. If the experimenter <u>can</u> specify a data-generating model, but is unable to anticipate the decision-maker's prior distribution and the economic consequences of each possible course of action, then the raw data, sufficient statistics, likelihood function, or distribution posterior to a diffuse prior will each convey all of the experimental information. A strong case can be made for choosing the posterior distribution as the best summary of the experimental information: it is easily combined via Bayes' theorem with the decision-maker's prior distribution provided the latter can be expressed as a pseudo-sample, and it may convey sufficient information by itself to permit the decision-maker to forego the sometimes painful process of assessing his prior.

In this case any summary measures of the reported posterior distribution run the risk of losing information by preventing the decision-maker from incorporating his own prior information. By way of illustration, consider Professor Tiao's example for the comparison of two variances. The data happen to consist of measurements made by an inexperienced analyst A_1 and an experienced analyst A_2 in performing a chemical assay. As the man responsible for making decisions about A_1's competence, I might be very interested in <u>a</u> posterior probability that $V > 1$ -- but not in <u>the</u> posterior probability that $V > 1$ given in Professor Tiao's paper. I would certainly want <u>my</u> probability to reflect what I already know of the relative abilities of A_1 and A_2, of the amount of experience A_1 has already had, of my knowledge of learning rates, etc., but the posterior probability that $V > 1$ given by Professor Tiao, whether conditional on β or marginal, does not permit me to incorporate this information. What is really needed for my purpose is the marginal distribution of V posterior to a diffuse prior.

Having considered the case in which the experimenter knows only the data-generating model which the decision-maker will use, let us turn to the case in which the experimenter, before conducting the experiment, possesses all of the prior and economic information of the decision-maker. In this case he may indeed calculate the distribution of the decision parameter posterior to both the experiment and the decision-maker's prior distribution, and, with knowledge of the economic structure of the problem, may even obtain summary measures of this distribution without sacrificing information. But I maintain that in this case a report of such measures does not constitute sufficient processing of the experimental data: a report of the posterior expected utility of each act is in keeping with our goal of processing the experimental results as far as is possible without sacrificing relevant information.

Summary measures of the posterior distribution may be relevant when the experimenter knows the prior distribution of the decision-maker and the functional form of the utility of each act with respect to the decision variable, but does not know the parameters of these functions. Such cases, however, are too rare to justify these measures.

Surely these summary measures are important because of their relation to the analogous classical point estimates, confidence intervals, and significance tests; they are also interesting conceptually and useful in the teaching of Bayesian inference and decision theory. But I think that their importance among the tools of the trade has been overemphasized.

VI

INFORMATION STORAGE AND RETRIEVAL - I and II

Chairman, Ezra Glaser, U. S. Patent Office

INTRODUCTORY REMARKS ON PATENT OFFICE RESEARCH ON INFORMATION RETRIEVAL

Ezra Glaser, U. S. Patent Office

The four papers which comprise this session present a unified description of a research program undertaken by the U.S. Patent Office on information search and retrieval. There is an undeniable importance to this topic, in its own right, because it affects the work of the Patent Office and many other scientific and technical organizations, both public and private.

Nevertheless, the inherent scope of the meeting is broader than information retrieval or the use of scientific documentation. You will observe, in the papers that follow, that we are dealing with the properties of large organizations and complex adjudication processes. The present embodiment of the study is the Patent Office and the work of the patent examiner. Many of the same problems will occur in the regulatory and benefit granting agencies of government and in a substantial range of private organizations.

Mr. Spencer's presentation will give the context in which the Patent Office and the patent examiner must be observed. You will note the exceptionally demanding definition of literature search and the imposing difficulty of the patent examiner's job.

With this realistic statement of the problem facing the Patent Office and its personnel, you will be equipped to understand the purpose and approach of the paper by Dr. Bryant in which he constructs a general model of the examining procedure. The proposed model provides a framework for the analysis of the performance of patent examiners. In addition, it provides the specifications for reporting systems and experimental designs for the gathering of data essential to useful applications of the model. You will note that Dr. Bryant's model has already been fitted out with preliminary data for a summary characterization of the examining activity at the Patent Office and to illustrate some of the kinds of analysis which become possible with this approach.

Donald King addresses himself to some of the technical problems in experimental design which arose in the course of this work. He will present some early experimental results.

Dr. Cornog, in the final paper in this session, will discuss the particular aspects of the research problem that required her services as an experimental psychologist. In particular, the experimentation involved the observation of highly trained and extremely sophisticated personnel, who could not help but be aware of the implications of the experiments and studies of which they were subjects.

Mr. Richard A. Spencer is Staff Director of Research, and is responsible for the Office of Research and Development of the U. S. Patent Office. His bachelor's degree is from Catholic University in Washington, D. C. and he has a law degree from Georgetown University.

Dr. Edward C. Bryant has a bachelor's degree and a master's degree from the University of Wyoming and a doctorate from Iowa State University.

Mr. Donald W. King received his bachelor's and master's degrees from the University of Wyoming.

Dr. June R. Cornog has a bachelor's degree from Pennsylvania State University, master's degrees from the University of Pittsburgh and George Washington University and a doctorate from Pennsylvania State University.

Westat Research Analysts, Inc., is a contractor to the U. S. Patent Office for the purpose of providing technical services in mathematical statistics and operations research.

PATENT EXAMINATION AND THE NEED FOR RESEARCH

Richard A. Spencer, U. S. Patent Office

The patent system, as it exists today, is designed to implement the provision of the Constitution that empowers Congress "to promote the progress of science and the useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." The philosophy of granting monopoly rights as a means of fostering progress and providing an incentive to invent and disseminate new knowledge was inherited from the English common law. Without question the patent system has contributed profoundly to the growth of our country and our present way of life.

Throughout our history various laws have been passed by Congress to carry out the patent provision of the Constitution. Although these laws have differed there are certain unifying principles. The granting of a patent has been made a matter of right. An inventor who complies with the provisions of the statutes must be granted a patent; it is not a discretionary matter with the Patent Office. Therefore, the Office must make a determination, in each case, as to whether or not the applicable statutory provisions have been complied with. This involves many technical and legal considerations on the part of the patent examiners who constitute the heart of the Patent Office.

Of prime importance is the determination of whether or not a patentable invention has been made. The law provides that a patent may be obtained for any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof. Thus the invention must be new and the law provides that a person is not entitled to a patent if the invention was patented or described in a printed publication in this or a foreign country before his invention thereof or more than one year prior to his application for patent. This imposes a staggering searching burden on the examiner for a description of the invention anywhere in the literature in any language can preclude the granting of a patent. The problem is further complicated by the fact that a patent may not be obtained even though the invention is not identically described in the prior literature if the differences between the subject matter of the invention and the prior art are such that the subject matter as a whole would have been obvious to a person having ordinary skill in the field to which it pertains. In effect an inventor is charged with full knowledge of every pertinent publication. Patents are not granted for normal developments in a field - there must be, as some have termed it, a flash of genius.

In order to obtain a patent an inventor must file an application. This is typically done through a patent attorney. The application must include (a) a written description of the invention in such full, clear, concise and exact terms as will enable any one skilled in the field to which it pertains to make and use the invention, (b) one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant considers to be his invention and (c) where appropriate, a drawing of the invention. The claims define the scope of the patent protection the inventor is seeking and generally there will be a number of claims; some claiming the invention broadly and others claiming it more specifically.

At the present time we are receiving more than 80,000 applications a year. When an application is received it is checked for completeness, given a filing date and serial number and assigned to an examiner. The examining corps is organized into four operations; chemical, electrical, mechanical and general engineering. Each operation is divided into groups which are responsible for more specialized portions of these broad technical areas. In each group examiners are assigned even more specialized areas of technology. Thus, for example, one examiner might only deal with steroid chemistry and another might only deal with analog to digital converters. The subject matter for which he is responsible is called the examiner's art. In fields where there is sufficient activity several examiners might be assigned the same art. The assignment of a new application to an examiner is governed by the subject matter of the invention. By the very nature of invention applications do not always fit neatly into preconceived pigeon holes and such applications will be assigned to the examiner having the art to which it most nearly pertains.

Sometime after receipt and assignment of the application the examiner will take it up for consideration. He must read and thoroughly understand the description and claims. He determines the scope of the search which frequently will extend beyond his art. He searches the patents and other documents that are classified in his art and any others that are pertinent to find the most closely related documents. He develops a position with respect to each claim. His search may be generic or specific and includes a search for equivalent disclosures and disclosures that can be combined to meet the claims. Obviously the broader the claim the more likely that the search will produce an anticipatory reference. Then he writes an "office action" in which he clearly states his position with respect to each claim; either rejection or allowance. When a rejection of a claim is based on prior art, the examiner must distinctly point out how he has applied the prior art to the claim. The office action will also include a discussion of the statutes, court decisions and formal requirements that are germane to the case.

A copy of the office action is forwarded to the applicant who, under normal circumstances, has six months to reply. The typical reply will agree with the examiner in part and disagree in part, arguments will be pressed as to why certain claims should be allowed and the application will be amended by clarifying portions of the description and deleting, modifying or adding claims.

The case now is categorized as an amended application and is taken up in due course for further consideration by the examiner. The examiner considers the arguments of the applicant and the amendments and may conduct a further search of the prior art before preparing another office action. Again the applicant must reply in six months which may be followed by additional office actions and replies.

Through this process the examiner and applicant, or his attorney, negotiate until a resolution of the case is arrived at. This resolution can take several forms; the applicant may abandon his attempt to obtain a patent, or he may put it in a condition where the examiner allows all remaining claims and issues a patent or the two parties may reach an impasse. In this latter case there will normally be some claims that are allowed and some that the examiner is not willing to allow. The examiner makes a final rejection of the claims he is not willing to allow and the applicant may then appeal to our Board of Appeals. No new issues can be raised on appeal and the Board will render its decision, based on the record, affirming, modifying or reversing the position of the examiner. The Board of Appeals is the final adjudicating authority in the Patent Office on matters going to the merits of the application. If, after appeal, the applicant still disagrees with the position of the Patent Office, he can appeal to the courts. Of course the final court decision, which may be one by the Supreme Court, is binding on both parties.

In processing applications many other problems arise and procedures exist for handling them. As an example it is not uncommon to have two or more applications which claim essentially the same subject matter. As technology progresses it is to be expected that the same solution to problems in a field will be arrived at independently by two or more people. In such cases, if claims are found patentable, the Patent Office institutes an Interference Proceeding and a Board of Interference Examiners determines priority based upon evidence given by the parties as to actual dates of invention, reduction to practice, etc. Again a decision by the Office can be appealed to the courts. Thus an application can be shunted along many different paths before a final disposition is made of the case.

The laws charge the Patent Office with maintaining a system which most will agree is one of the foundations of our economy. In view of this alone, it behooves us from time to time to consider how well we are performing our function. Unfortunately, in looking into our performance, we are forced to conclude that we can not be satisfied. We are in the midst of a crisis and the system is in jeopardy. Advancing technology has created problems which, if not solved, may well force abandonment of the system as we know it.

In the early days the Office had only a handful of examiners. It has been stated that it was not at all unusual then for an examiner to wait around for an application to come in so he would have something to do. Today we have more than 1000 examiners and over 200,000 applications pending in the Office. Every examiner is deluged with a backlog of work and, in many arts, applications are not taken up for consideration until almost two years after filing or amendment. On the average it now takes more than three years to arrive at a final disposition of an application and it is not unusual for ten years to elapse between the date of filing and final disposition. The adverse effects of such delays are many; not the least significant of which are the denial of the proper protection of the patent laws to the inventor and the withholding from the public of new knowledge for an undo length of time. The dissemination of new knowledge is the quid pro quo on which the granting of the seventeen year patent monopoly is based. There are many causes of this backlog such as the increasing complexity of technology, the increasing volume of prior art that must be searched in each case. The net effect which causes us considerable concern is that the rate of disposals has been steadily decreasing - in the mid thirties the average examiner disposed of 160 cases per year; today his disposal rate is 80 cases per year.

When we look at the quality of our performance we are equally disturbed. Every issued patent is presumed to be valid and most are never challenged. However, a certain number are challenged in the courts - normally when someone other than the patent holder produces the item and is sued for infringement by the patent holder. In such cases, the percentage of patents held invalid, because of prior art brought forth by the alleged infringer, which the examiner apparently did not consider, has been increasing at an alarming rate. Admittedly patents and other issues do not usually get to the stage of court litigation unless there is a reasonable difference of opinion between reasonably competent groups of attorneys. However, our concern over our quality of performance is further strengthened as a result of recently instituted quality control procedures. Quality control of an operation such as patent examination is, at best, extremely difficult but we have attempted to set up means for rating the quality of performance on cases that have gone to final disposition. The review and rating of cases is done by our most highly skilled examiners and the factors rated extend from routine records and housekeeping detail to fundamental points such as the proper application of legal decisions and prior art to the claims. A sample of cases was drawn and rated to determine the existing quality level for one of the operations. The results showed that, in a substantial percentage of the cases, at least one office action was defective enough as to raise doubts regarding the validity of the patent that subsequently issued. Unfortunately we have no true measure of quality in the past but, even if we had, we could not be satisfied with the present apparent level.

The problems of productivity have been with us for some time. The problems of quality are just now emerging. In the past productivity problems have, in general, been attacked by adding to the examining corps. This may have helped to keep the backlog from rising to even higher levels than it is today but it has not solved the problem. Much of the advantage that seemed inherent in expanding the examining corps proved to be illusory because of higher attrition, the need for training and the lowering of the average experience level.

It has been evident for some time that more drastic measures are necessary if we are to preserve the examining system as it exists today. Obviously, things can be done within the present framework of operations to improve our effectiveness and our present Commissioner has been conducting a dynamic program to wring the utmost from our current methods of conducting business. A complete reorganization of the Office has been implemented, more authority has been given to examiners of proven merit, compact prosecution practices have been introducted which are designed to reduce the number of office actions and applicant replies per application, quality and quantity norms are being established and a promotion policy based essentially on performance rather than seniority has been introduced. These steps have had and will have substantial effects on our productivity and quality but they can not, in themselves, solve the overall problem. The crux of our problem is the exponential proliferation of technical literature.

Any one, today, in any technical field can not help but be effected by the avalanche of technical literature. One can minimize this problem by increasing the specialization of his area of interest. One who does this can rightfully say there is no information problem - if he is faced with a problem he can call one to a half dozen or so scientists who are working in his specialized field and quickly determine if there is a solution. Unfortunately the patent examiner can not do this. He is constantly faced with the problem of someone who claims he has found a unique solution to a problem. Much of the function of the examiner is a determination of whether the claimed unique solution is in fact new or whether it is anticipated by the prior art.

In theory, at least, the totality of the world's literature is subject to being searched for each application, for an anticipation may be found in many obscure places. Further nothing ever becomes too old to be of potential interest. Citations of patents more than 100 years old are not uncommon and the Old Testament has even been cited to reject claims. In practice, of course, we can not search everything. We do, however, maintain a technical file of over 3 million U. S. patents, more than 5 million foreign patents and uncounted books, journals, periodicals, etc. This file is constantly increasing in size and we are slowly being buried under the avalanche.

The problem of dealing with large amounts of technical information is not new; we have been living with it for many years. In order to deal with it we have devised a classification system which is probably the most sophisticated in the world. We have, at present, over 300 classes subdivided into some 62,000 subclasses into which documents are classified. A subclass may contain from a few to several thousand patents. There is a need for constant reclassification as technology develops and additional documents are added. For many years this system worked well in directing the examiner to a subset of documents which contained the most pertinent references. However with the growth in magnitude of the total file and the necessity of finer and finer distinctions in classification it is becoming less and less effective. We are facing a losing battle in trying to maintain a classification system for manual search of documents. Continued subdivision of files can in many cases only force the examiner to search in more subclasses - for the claimed concept may overlap many subclasses. Documents characteristically contain information about many things; equivalent or analogous disclosures can exist in seemingly unrelated fields; two or more documents can be combined to anticipate an idea. For example claims drawn to as simple a device as the poppet beads worn by many women were searched in 15 subclasses covering such diverse arts as Ornamentation; Education and Amusement Devices; and Chain, Staple and Horse Shoe Making.

As the store of information grows, the time required to conduct a proper search grows despite anything that can be done through classification. Coupled with the growing search load there have been continual pressures to maintain or increase production because of the backlog and delays. In such a situation something has to give. It is characteristic of the system that the individual examiners can control production within wide limits by the time they devote to each case. We are concerned that the system is giving in the area of quality. If quality is suffering we are doing a disservice to the inventor and to the public which can cause substantial economic effects.

In view of the importance of the search aspects of examination and the fact that searching has been estimated to occupy the major portion of an examiner's time, it is evident that something must be done to make the searching time more effective and to minimize the necessity of examiners looking at numerous documents that have no bearing on the issue in question. Ideally the examiner should be able to quickly identify the documents which should be considered in order to properly develop his position with respect to the application at hand.

This need has been recognized for some time. In the late 40's a pilot project was undertaken to see if the advantages of automation could be brought to bear on the search problem. A group of chemical patents was selected and analyzed in detail to identify each bit of disclosed information that would be of potential interest to an examiner. Techniques for coding different kinds of information were developed and a card file containing these codes was established. Machine searches of typical examiner inquiries, using a simple sorting machine were demonstrated. This small project definitely showed the feasibility of mechanized searching as an aid to the examiner. By asking a question in code form he could quickly identify a small subset of patents which contained the most pertinent disclosures.

For many reasons further effort was not conducted along this line until 1955. At that time a group was set up to again investigate the use of mechanization for searching. This group subsequently became the Office of Research and Development. Arrangements were made to conduct research jointly with the National Bureau of Standards because of its outstanding competence in computers and automation. That arrangement continues to this day.

Two courses of research have been followed. One took the path of the earlier project and was directed to the development of coding means for selected arts. The other was directed toward more basic research problems which have to be solved if the benefits of mechanization are to be extended to all the arts and all the search problems. It became evident at an early stage that existing technology would only permit extraction and coding of well organized and identifiable concepts in documents; for example the structure of a chemical compound. However documents are full of unorganized, vague and ambiguous concepts as well as those that only convey the intended meaning when read in context. So it was apparent that basic research had to be conducted in linguistics, self organizing and adaptive systems, machine analysis of documents, automatic derivation of concepts and the relationships between concepts in context, etc. A large part of our research in these areas is done jointly with the National Bureau of Standards. We feel that the eventual solution can only be achieved through basic research in the fundamental problems of information storage, manipulation and retrieval, but any further discussion of this would not be germane to this session.

The efforts along the pioneering path established in the earlier project were designed to give more immediate assistance to the examiners than could be expected from the basic research. In view of the inherent organization of certain aspects of chemical information most of these projects have been directed to various chemical arts. Arts in which the examiner's search is directed toward the structure of a compound were identified and methods for analyzing and coding the documents in the appropriate subclasses were developed. The first such system to be put on an operational basis was for the field of steroid chemistry. Steroid compounds are all related through having a common structural nucleus. They differ from one another by the elements or functional chemical groups that are attached to this nucleus. The coding system is specifically linked to these features of steroids and is, therefore, not applicable to compounds in general. A more generalized coding scheme was then developed and applied to a group of patents in the organo phosphorus art. Later an even more generally applicable system was developed and applied to the organo metallic art. This latter system is the most recent to be put on an operational basis in the Office. Some work has also been done in fields other than chemistry.

In the organo-metallic system, as in the other chemical systems, the analysis and coding is directed to compounds only. Each patent is analyzed by a skilled chemist who determines the structure of each compound disclosed or implied. Each structure is decomposed into fragments which may be single elements or groups of elements and which are generally accepted by chemists as building blocks of compounds. The coding consists essentially of recording the fragments that are connected to each other and descriptive information about these fragments. It includes specific and generic information. This information is transferred to punched cards for searching with a multi-column sorter. For searching, the examiner specifies what fragments he wants connected together and what descriptive characteristics the fragments must have. The search produces a print out of the appropriate patent numbers. If there is an answer to his question it should be contained in this set of patents. However patents that do not satisfy his purpose may also be retrieved. For example, if an examiner wants fragments A, B, and C connected in sequence he can only ask for A connected to B and B connected to C. The search will retrieve, in addition to patents having the A,B,C connection, patents disclosing the A,B connection in one compound and the B,C connection in another compound and also patents disclosing a compound having two B fragments, one of which is connected to an A and the other to a C. So the examiner must still carefully review the retrieved patents to determine pertinency.

We now have over 12,000 documents in our various mechanized files. Similar systems could undoubtedly be derived for other arts, particularly chemical. There has been good acceptance on the part of most examiners associated with these mechanized systems. There are drawbacks such as the fact that a search for a chemical process can only be conducted by searching for compounds that might have been made by similar processes. Despite the drawbacks, some are of the opinion that more effort should be directed to setting up such files. However, we are not content to proceed on the basis of opinions. We feel it is essential that we properly evaluate such systems before embarking on an extensive program of applying these techniques to other arts. Every mechanized file represents a capital investment - our direct labor costs for analysis have varied from $5 to over $100 per patent. Since we can not now code everything that is in a document, we must also continue the classification effort. Is the capital investment justified? Can we gain benefits in increased production or better quality? We know that after the introduction of the steroid system productivity rose markedly but it has since declined to levels comparable to the pre-mechanization days. How long can we expect a mechanized file to operate effectively? Technical developments necessitated a revision in the steroid coding which effected some of the patents previously coded. It seems that the more specific a coding system is the more susceptible the file is to becoming obsolete as a result of developments. Unfortunately the difficulty and cost of analysis increases as the coding systems become more general. What is the optimum balance? With human analysis there are human errors. What effect does this have on retrieval? What accuracy or consistency of analysis can we expect or require? What is the effect of retrieving false answers? Do they obscure the true answers? These and many other similar questions, we feel, must be answered before we will know best how to proceed. A large part of our current program is directed towards experiments and tests which are designed to answer such questions. Some of the work that has been done in this area will be discussed in the subsequent papers.

A STOCHASTIC MODEL FOR THE PATENT EXAMINING PROCESS*

Edward C. Bryant, Westat Research Analysts, Inc.

A. Introduction

The United States Patent Office is a complex operating system in which patent applications are received and processed according to strict rules of practice. The heart of the examining process is the patent examiner. While his primary function is that of ruling on admissibility of patent claims, the importance of his communication with the applicant in narrowing and molding the scope of claims cannot be ignored. The two-way communication between examiner and applicant (usually through the applicant's attorney) is accomplished in an environment of laws, rules, operating procedures, boards, other examiners, clerical operations and files, and these all taken together constitute the patent examining system.

Rulings of the examiner are based upon his knowledge of the art within which he is working, augmented by references he finds in the files, so that his ability to extract pertinent information from the files is one of the keys to successful operation of the system, and constitutes the search which he is required by law to make. The complexity of the search has led to interest in information storage and retrieval, the establishment of the Research and Development Office, and the initiation of research now in progress.

It is worth noting, however, that the storage and retrieval of information is only a part (admittedly an important one) of the entire system. Improvements in this activity will be reflected in the operation of other components of the system. Thus it is important to understand the operation of the entire system in order to evaluate and possibly to make conditional predictions as to the effect of changes in any of the components. One approach to describing the entire system is to develop a mathematical model of it. Such a model must be "stochastic", in the sense that one cannot predict with certainty the length of time required to perform any of the examining functions, nor whether the claims of the application will be denied, approved, or modified. However, all of the operations of the system might be well described by probability distributions so that, even though one cannot describe the course of a particular application with precision, he can, nevertheless, treat with considerable accuracy the behavior of the system under a large input of applications.

If successfully constructed, a stochastic model of the Patent Office should: 1. Serve as a basis for understanding the structure of the system. 2. Reveal where there is potential for major gains in efficiency. 3. Show the effect of proposed changes in one component on other components of the system due to interdependence of the process elements. 4. Indicate the direction of needed experimentation. 5. Tie together the pieces of the research program.

B. Construction of the Transition Model

It will be convenient to think of the system as composed of stations and activities which are performed at the stations. The stations represent locations at which functions related to a patent application take place. The application does not always pass from station to station, but the responsibility for the next action does. For example, the application is not physically returned to the applicant for amendment, but the examiner's action on the application is sent and this shifts the responsibility for the next action to the applicant.

A simplified flow diagram for actions related to patent examinations is given in Figure 1. Note that the "examiner" station is actually a collection of stations (examiners, divisions, groups or operations) and the "applicant" station can be similarly subdivided. The degree to which these stations will be subdivided will depend on the use to be made of the model. Once an application has been assigned and examined it is presumed that all future examiner actions are accomplished within the same examining organization (division or group). There may be exceptions, of course, but their importance will be determined in the process of

*The research leading to this paper was done under contract to the U. S. Patent Office with the cooperation and assistance of the staff of the Office of Research and Development. Data presented here were collected by Patent Office chemical analysts under the direction of Mr. Leibowitz of the Office of Research and Development.

collecting data for the parameters of the model.

When the application is received it is given a filing date (an important step with regard to patent rights) and is microfilmed. Certain other preliminary things are done to it and it is assigned to an examining division. Any controversy over what division should handle the application is settled by the Classification Division. Upon reaching an examining division it must await its turn among all original applications. That is, it must go to the end of the "queue" unless priority is given to it by the Commissioner of Patents. It should be noted that there are separate queues for initial applications and amended applications and no fixed rules for determining which of the two shall be given preference.

After examining the application the examiner may "allow" all or part of the claims or "reject" them. If the claims are allowed, a notice of allowance is sent to the applicant and, upon receipt of a fee by the Patent Office and completion of other mechanical details, the patent is issued. It is printed and copies are placed in the files as well as made available for sale.

If the claims are rejected the applicant is so notified and the basis for the rejection is given in the examiner's action. The applicant has a fixed time period within which to amend the application, otherwise it goes to the abandoned file. Assuming that he amends his claims, his amended application is placed at the end of the examiner's queue of amendments and is acted on after everything ahead of it (unless priority is given).

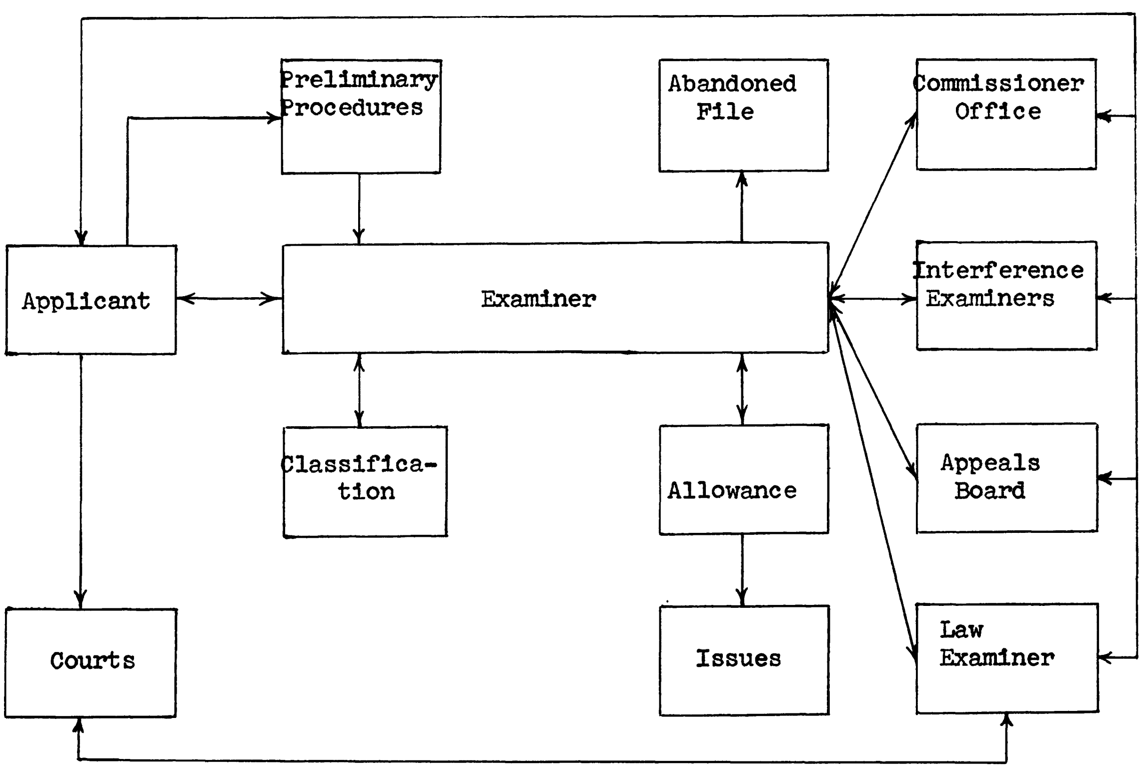


Figure 1. Flow diagram for actions in patent examinations.

After a second rejection the applicant may appeal to the Board of Appeals, he may abandon the application, or he may amend it again. Ultimately, every application is either abandoned or issued (in case any part of it is allowed), but actions on some applications follow a circuitous route before final disposition.

When an application is judged allowable by the examiner it is his responsibility to effect an "interference search" to determine whether there are other applications in process or recently issued patents whose claims are so similar as to raise questions concerning priority of invention. If, in the examiner's judgement, there are such possible interferences he will take the necessary action to set up "interference proceedings" in an effort to determine priority of invention. He may notify the law examiner who asks all parties except the one with earliest filing date for affidavits relating to evidence of date of invention. If, in the law examiner's opinion, there are still grounds for controversy the patent examiner holds an interference hearing to try to decide the matter. The examiner may also take action directly with the parties, suggesting the wording of claims so that there is no doubt that the same invention is being claimed. He may hold hearings with the parties, but if the issue cannot be resolved by him, an interference is formally declared, and the case goes to the interference examiners who make the final decision for the Patent Office. Appeal from this action can be taken only through the courts.

The function of the Appeals Board is to determine patentability controversies which arise between applicant and examiner. Again, its decision is final with respect to the Patent Office, and recourse is only through the courts.

Another action which may be taken is for the applicant to petition the Commissioner of Patents for reconsideration of an action by an examiner.

This brief outline of the examining process is vastly oversimplified. It is not intended to be comprehensive, but simply to indicate that there is a flow of actions through a network of stations, at each of which some time is required to service the application and at most of which there is a queue awaiting service. Since the flow of actions can be described in terms of probabilities, the system is called "stochastic" and it should be possible to construct a mathematical model of it.

One has a great deal of flexibility in choosing the amount of detail to be specified by the model. For example, one might wish to consider a model consisting only of applicant, examiner, issue and abandonment stations. Or he might restrict his attention to the examining process for chemical patents. Here it should be noted that the examining system is, in fact, a collection of subsystems, one for each grouping of the arts. They are tied together at certain common stations through which all may pass, such as issue, abandonment, appeal, classification, and so forth. Thus it is possible to aggregate arts in any manner which is meaningful with regard to the problem under investigation. The model which is contemplated here is general enough that it can encompass any of the specialized models which may be needed.

We first consider a transition model which attaches probabilities to the passage of actions from one station to another. It is independent of time and hence has serious limitations. A time-related model will be presented later.

For any particular form of the model we will assume a set of stations S_i, $i = 1, 2, \ldots, n$, with probabilities governing the passage of an action from S_i to S_j. These probabilities are called "transition probabilities" and a matrix of such probabilities is called a "transition matrix". An illustration is given in Figure 2. Since the entries are probabilities we have $0 \leq p_{ij} \leq 1$ for all i and j. Also, it is clear that $\sum_j p_{ij} = 1$ for all i. A matrix exhibiting these properties is called a "stochastic matrix". In case $p_{ii} = 1$ we say that state S_i is an "absorbing state". That is, once the application has reached this state it will remain there.

We will consider here only the class of stochastic matrices having a finite number of entries, although for certain applications a system with infinitely many states may prove helpful.

In the Patent Office model the probability that action on an application will pass from S_i to S_j is dependent on the history of the document prior to its arrival at S_i. Thus the model does not exhibit the independence property of a Markov chain. However, it is possible to construct a matrix of transition probabilities with the Markov property by adding stations. For example, we can let

S_1 = source of the original application

Present station	Next Station S_1	S_2	...	S_j	...	S_n
S_1	p_{11}	p_{12}	...	p_{1j}	...	p_{1n}
S_2	p_{21}	p_{22}	...	p_{2j}	...	p_{2n}
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
S_i	p_{i1}	p_{i2}	...	p_{ij}	...	p_{in}
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
S_n	p_{n1}	p_{n2}	...	p_{nj}	...	p_{nn}

Figure 2. Transition probabilities governing passage of actions through the system.

S_2 = original examination by the examiner
S_3 = first amendment by applicant
S_4 = review of first amendment by examiner
S_5 = second amendment by applicant

and so forth.

To avoid an infinite set of states we can choose some arbitrarily large value representing the maximum number of times a document will be amended and require that the application either be abandoned or go to appeal at that point. A reasonable value can be determined by survey, and little damage will be done to the model by imposing this kind of restriction.

C. Mathematical Notes on the Transition Model.

Consider a system in which there are absorbing states A_1, A_2, ..., A_r, and non-absorbing states T_1, T_2, ..., T_s. The total number of states is then $r + s = n$. The stochastic matrix of transition probabilities $P = \|p_{ij}\|$ may be partitioned as follows:

$$P = \left[\begin{array}{c|c} I & 0 \\ \hline R & Q \end{array}\right]$$

where I is the r x r identity matrix representing the absorbing states, R is an s x r matrix of probabilities of transition from nonabsorbing to absorbing states, Q is an s x s matrix of probabilities of transition from nonabsorbing to nonabsorbing states, and 0 is an r x s matrix of zeros. We assume that states can be chosen in such a way that the transition probabilities are dependent only on the state i, so that the stochastic matrix is a representation of a Markov chain with absorbing states. Ways of selecting states so as to achieve this result are discussed in the following section.

Let $p_{ij}^{(m)}$ be an entry of the matrix found by multiplying the matrix P by itself m times. That is $P^m = \|p_{ij}^{(m)}\|$. The entires $p_{ij}^{(m)}$ represent the probabilities of transition from state i to state j in exactly m steps. Consider the powers of the partitioned matrix, P.

$$P^2 = \left[\begin{array}{c|c} I_{rxr} & 0 \\ \hline (I_{sxs} + Q)R & Q^2 \end{array}\right]$$

In general,

$$P^m = \left[\begin{array}{c|c} I_{r\times r} & 0 \\ \hline (I_{s\times s}+Q+Q^2+\ldots+Q^{m-1})R & Q^m \end{array}\right]$$

If for each transition state i there exists an integer k such that $p_{ij}^{(k)} > 0$, where j is an absorbing state, Q^m will converge to the null matrix and $(I + Q + Q^2 + \ldots + Q^{m-1})$ will converge to the inverse* $(I - Q)^{-1}$. The matrix $(I-Q)^{-1}$ is called the "fundamental matrix" of the Markov chain.

Consider a nonabsorbing state i and an absorbing state j. Let q_{ik} denote an entry of the matrix Q and r_{ij} denote an entry of the matrix R. Then the probability of transition from i to j in one step is r_{ij}, in two steps is $\sum_k q_{ik}r_{kj}$, in three steps is $\sum_m\sum_k q_{ik}q_{km}r_{mj}$, and so forth. These are seen to be the entries of the matrices R, QR and Q^2R. Thus, the partitioned form of P^m, above, shows that the probability of ultimate transition from i to j can be found by the entries of the matrix B, where

$$B = (I - Q)^{-1}R.$$

Hence, one can readily determine probability of absorption in any absorbing state from any nonabsorbing state.**

Also, the expected number of times of being in state k (a nonabsorbing state), starting from state i (another nonabsorbing state), is the ik^{th} entry of $(I-Q)^{-1}$. We are primarily interested in the first row of the inverse which shows the expected number of times a patent application will be in each state, starting from its receipt by the Patent Office. By adding together the entries for all examiner actions we can determine the expected number of examiner actions required.

*David Rosenblatt, "On linear models and the graphs of Minkowski-Leontief matrices", Econometrica, Vol. 25, No. 2, April, 1957, pp. 325-28, shows that the inverse will exist for substochastic matrices Q if such Q contain no stochastic closed cyclic nets.

**These results may all be found in Kemeny and Snell, Finite Markov Chains, D. Van Nostrand Co., Inc., Princeton, 1960, Chapter III.

D. Illustration of the Transition Model with Actual Data from the Patent Office.

The data. A random sample of 500 cases relating to chemical patents was drawn from those on which action had been completed in 1962. The sample was allocated proportionately between abandoned and allowed cases so that results could be combined without weighting. The study was restricted to chemical patents because the Chemical Operation* was the only one of the four Patent Office examining operations which had been reorganized at the time the sample was drawn. Also, other research was in progress in the Chemical Operation which made it possible to use the sample for multiple purposes.

Each case is represented by a file containing the original application, the examiners' actions, amendments to the application, notices of appeals, briefs, arguments, and all ancillary papers associated with the process of obtaining (or being denied) a patent. All papers are dated so that elapsed time from one step to another can be computed. Examination of the sequence of actions identifies the "states" which must be considered in the stochastic matrix of transition probabilities.

The data collection forms, as well as a discussion of the various contemplated models, have been presented in an informal report** to the Patent Office and will not be reproduced here. It is sufficient to say that the required data were extracted from the files by trained chemical analysts in the Office of Research and Development, and were then punched on IBM cards for flexibility in analysis.

The states and transition probabilities.***

*The word "Operation" refers to a major grouping of the arts for administration purposes.

**Bryant, E. C., "On stochastic models of the Patent Office examining system," WRA PO 8, April, 1963, Office of Research and Development, United States Patent Office.

***All of the numerical data in this paper must be considered to be preliminary, since the limited time available for the preparation of this paper has made it necessary to use data which have not been checked.

The preliminary model contains the following states:

000 - original application

001 - first amendment prior to final rejection

002 - second amendment prior to final rejection, etc.

010 - first amendment after final rejection

011 - second amendment after final rejection, etc.

021 - first amendment after notice of appeal

022 - second amendment after notice of appeal, etc.

100 - examiner action on original application

101 - examiner action in first amendment, etc.

111 - examiner action on first amendment after final rejection

112 - examiner action on second amendment after final rejection, etc.

312 - amendment after allowance

313 - request for amendment after allowance denied

320 - certificate of correction (after issue)

401 - first examiner amendment

402 - second examiner amendment

411 - first examiner amendment after final rejection

412 - second examiner amendment after final rejection

500 - final rejection after original application (can only occur when application is a continuation in part of a previous action)

501 - final rejection after first amendment, etc.

700 - notice of appeal

701 - appealed case settled by examiner

702 - appealed case decided by Board of Appeals

703 - appealed case decided by Board, with subsequent action by examiner (on amended claims)

751 - interference proceedings

752 - interference case settled by examiner

753 - interference case settled by Board of Interference Examiners

754 - interference case settled by Board, with subsequent action by examiner (on amended claims, or claims not in interference)

760 - both interference and appeal

800 - allowance (subject to amendment)

860 - final issue

900 - abandonment

The matrix of transition probabilities is a 48 x 48 matrix and hence is too extensive to reproduce here except in abbreviated form. Since most of the entries are zeros it is possible to express the matrix as shown in Table 1. The decision to use the particular states shown is an arbitrary one. It is possible, for example, to construct a simplified model in which one does not distinguish between the various kinds of amendments or examiner actions. The states chosen are believed to be adequate to develop the kind of analysis needed for initial management decisions.

It should be pointed out that a single state of this model may, in fact, represent a sequence of operations which can be expressed as another stochastic model. As an example, consider the actual process of acting on an application. One might consider the states of (1) preliminary reading and study, (2) literature search, (3) detailed study, (4) writing the action, (5) typing, (6) reading and signature, and (7) mailing. Undoubtedly, other states would reveal themselves during the analysis of the data. Some research is planned in the development of such detailed model if the larger model, presented here, appears to yield useful results.

The states in the matrix of Table 1 have been arranged in such a manner that the matrix is triangular, that is, no entries appear below the main diagonal. This was accomplished by careful choice of the transition states. A model in which one does not distinguish between a first amendment and a second amendment or between a first and second action would not have this property.

The matrix of Table 1 was partitioned and the fundamental matrix $(I-Q)^{-1}$ computed, as well as the product of this matrix and the matrix R (transition from nonabsorbing to absorbing states). Some of the pertinent results are shown in Table 2. The first column shows probability of issue of a patent, given that the application is in the state indicated in the margin. Note that these are probabilities of issuing _a_ patent, not necessarily on all of the claims currently being made. For example, when a case is appealed (state 700) the probability that a patent will be granted is .662, yet the annual report of the Commissioner of Patents for FY 1962 shows that the examiners' decisions were reversed in whole or in part in only 21 out of 81 cases. This apparent discrepancy is due to two principal factors: (1) most cases for

Table 1. Matrix of transition probabilities, Chemical Operation examining model.

Origin	No. of obsns.	Destination states and transition probabilities (given in parenthesis)			
000	491	100 (.917)	001 (.065)	500 (.002)	401 (.004)
		800 (.008)	900 (.004)		
100	452	001 (.910)	101 (.002)	751 (.004)	900 (.084)
001	443	101 (.679)	002 (.065)	501 (.171)	751 (.006)
		401 (.025)	800 (.052)	900 (.004)	
101	304	002 (.921)	501 (.003)	751 (.008)	800 (.003)
		900 (.065)			
002	313	102 (.403)	003 (.102)	502 (.358)	401 (.029)
		800 (.102)	900 (.006)		
102	126	003 (.921)	700 (.016)	401 (.008)	900 (.055)
003	152	103 (.230)	004 (.145)	503 (.474)	751 (.007)
		401 (.039)	800 (.105)		
103	35	004 (.829)	900 (.171)		
004	50	104 (.200)	005 (.080)	504 (.500)	760 (.040)
		401 (.060)	800 (.120)		
104	10	005 (.800)	800 (.100)	900 (.100)	
005	13	105 (.077)	006 (.154)	505 (.538)	800 (.231)
105	1	006 (1.000)			
006	4	106 (.250)	506 (.750)		
106	1	900 (1.000)			
500	1	700 (1.000)			
501	77	011 (.623)	760 (.013)	700 (.182)	900 (.182)
502	112	011 (.652)	760 (.018)	700 (.125)	900 (.205)
503	72	011 (.528)	700 (.208)	900 (.264)	
504	25	011 (.760)	700 (.160)	900 (.080)	
505	7	011 (.857)	900 (.145)		
506	3	011 (1.000)			
011	182	111 (.330)	012 (.115)	751 (.027)	700 (.176)
		401 (.010)	411 (.045)	800 (.280)	900 (.017)
111	61	012 (.459)	700 (.377)	411 (.016)	900 (.148)
012	46	112 (.173)	013 (.174)	760 (.022)	700 (.196)
		401 (.174)	800 (.261)		
112	4	013 (.500)	700 (.250)	900 (.250)	
013	13	113 (.077)	014 (.077)	401 (.383)	800 (.461)
113	1	800 (1.000)			
014	1	800 (1.000)			
751	9	752 (.111)	753 (.333)	754 (.556)	
752	1	800 (1.000)			
753	3	800 (.667)	900 (.333)		
754	5	800 (.200)	900 (.800)		
760	6	800 (.667)	900 (.333)		
700	101	701 (.634)	702 (.267)	703 (.099)	
701	64	800 (.953)	900 (.047)		
702	27	800 (.037)	900 (.963)		
703	10	800 (.600)	900 (.400)		
401	34	402 (.029)	800 (.971)		
402	1	800 (1.000)			
411	25	412 (.040)	800 (.920)	900 (.040)	
412	1	800 (1.000)			
800	294	312 (.118)	313 (.007)	850 (.861)	900 (.014)
312	33	850 (.970)	900 (.030)		
313	2	850 (1.000)			
850	293	320 (.014)	860 (.986)		
900		900 (1.000)			
860		860 (1.000)			

which a notice of appeal is filed are actually settled by the examiner, and (2) cases which are decided by the Board against the appellant may contain claims which can be allowed.

The second column of Table 2 permits computation of the expected number of times an action is taken by each participant in the examination process. Adding together the entries for 101-606, 500-06, 111-114, 401-2, 411-12, 312-313, and 800 yields 3.40, an estimate of the number of times an examiner is required to act on the claims in an application, exclusive of interference and appeals cases. A study of appeals cases (resulting in a separate model) shows that an examiner is required, on the average, to act on each appealed case about 1.8 times. Since about one-fifth of the cases go to appeal, the average number of examiner actions is about 3.75. It is important to note that there is a high degree of arbitrariness in the definition of an examiner action.

Table 2. Selected results from analysis of the chemical operation transition model.

Identification		Prob. of issue	Prob. of occupancy
000	original application	.588	1.000
100	examiner action	.584	.917
001	first amendment	.640	.899
101	examiner action	.612	.613
002	2nd amendment	.655	.621
102	examiner action	.617	.250
003	3rd amendment	.651	.231
103	examiner action	.612	.075
004	4th amendment	.740	.098
104	examiner action	.671	.020
005	5th amendment	.716	.024
105	examiner action	.578	.002
006	6th amendment	.578	.005
106	examiner action	0*	.002
500	final rejection 000	.662	.002
501	final rejection 001	.607	.155
502	final rejection 002	.597	.222
503	final rejection 003	.544	.139
504	final rejection 004	.692	.049
505	final rejection 005	.661	.013
506	final rejection 006	.770	.005
011	1st amend. after final	.770	.368
111	examiner action	.658	.122
012	2nd amend. after final	.855	.098
112	examiner action	.655	.017
013	3rd amend. after final	.981	.026
113	examiner action	.981	.002
014	4th amend. after final	.981	.002
751	interference	.436	.027
752	decided by examiner	.981	.003
753	decided by board	.656	.008
754	decided by board and examiner	.196	.015
760	interference and appeal	.655	.013
700	appeal	.662	.232
701	decided by examiner	.936	.155
702	decided by board	.036	.063
703	decided by examiner and board	.589	.024
401	examiner amendment	.981	.094
402	examiner amendment	.981	.003
411	examiner amendment	.942	.018
412	examiner amendment	.981	.001

*Based on a single case

(continued)

800	allowance	.981	.601
312	amendment	.970	.071
313	amendment not entered	1.000	.005
850	issue	1.000	.591
320	correction	1.000	.006

E. Time Dependent Models

The model presented in previous sections permits one to trace the flow of actions through the network, but at no place does time enter as a parameter. Thus, no conclusions can be drawn concerning length of time required to process an application to final disposition, number of examiners required to staff a collection of arts, or size of the backlogs at a given time. These are all matters of the utmost importance to the Patent Office, so that a model which includes time is essential.

A model with fixed time increments. A single formulation of the model is obtained by tracing the location of the application (or action) during each successive time increment (day, week, half day, etc.). The basic idea is presented in Figure 3.

The probability that an action will be in state S_k at time t_p is the sum of the probabilities of achieving that state at time t_p by all possible routes. For example, the probability that an action, starting at S_1 at time t_1, will be at S_3 at time t_4 is equal to $(.9)(.2)(.2) + (.1)(.7)(.2) + (.1)(.3)(.8) = .074$.

It is clear that Figure 3 can be expressed in matrix form by assigning a symbol to each node in the graph and assigning to each entry of the matrix the probability of transition from one node to another. This representation appears in Table 3. To prevent confusion, the word "state" will be used to denote the S_k and "node" will be used to denote combinations of S_k and t_p.

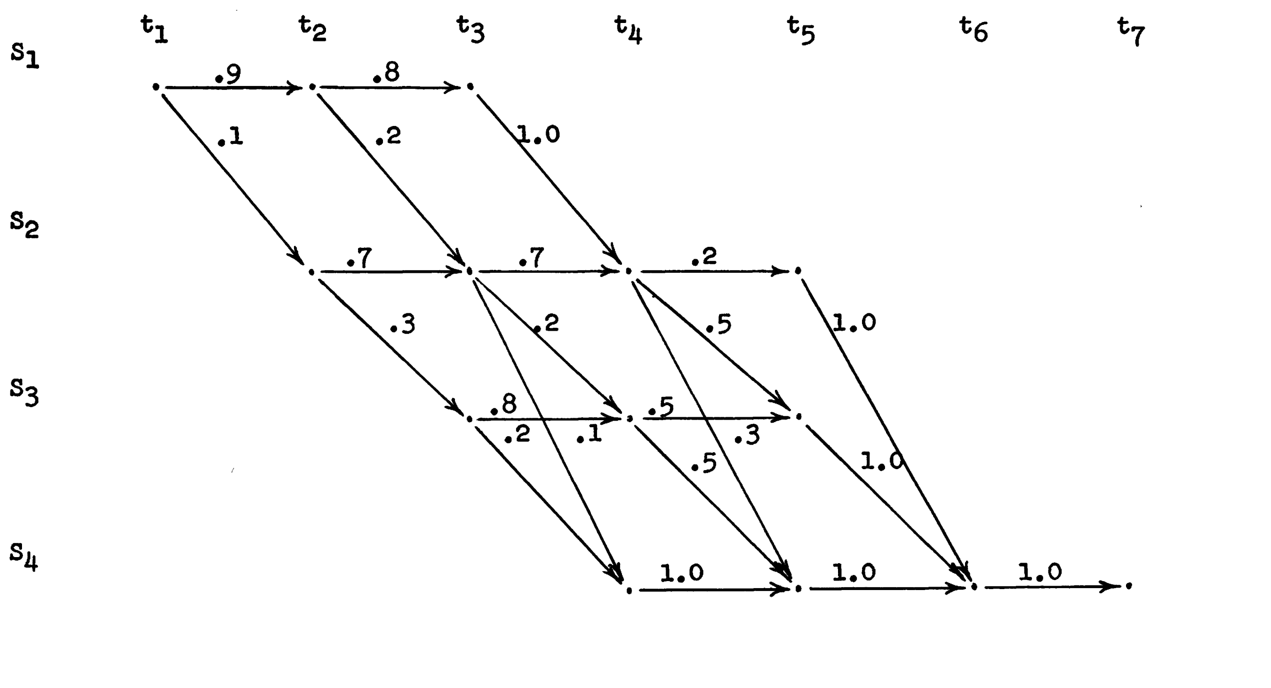


Figure 3. Flow of actions during fixed time increments.

Table 3. Matrix of the transition probabilities of Figure 3.

	S_1t_1	S_1t_2	S_1t_3	S_2t_2	S_2t_3	S_2t_4	S_2t_5	S_3t_3	S_3t_4	S_3t_5	S_4t_4	S_4t_5	S_4t_6	S_4t_7
S_1t_1	0	.9	0	.1	0	0	0	0	0	0	0	0	0	0
S_1t_2	0	0	.8	0	.2	0	0	0	0	0	0	0	0	0
S_1t_3	0	0	0	0	0	1	0	0	0	0	0	0	0	0
S_2t_2	0	0	0	0	.7	0	0	.3	0	0	0	0	0	0
S_2t_3	0	0	0	0	0	.7	0	0	.2	0	0	0	0	0
S_2t_4	0	0	0	0	0	0	.2	0	0	.5	0	.3	0	0
S_2t_5	0	0	0	0	0	0	0	0	0	0	0	0	1	0
S_3t_3	0	0	0	0	0	0	0	0	.8	0	.2	0	0	0
S_3t_4	0	0	0	0	0	0	0	0	0	.5	0	.5	0	0
S_3t_5	0	0	0	0	0	0	0	0	0	0	0	0	1	0
S_4t_4	0	0	0	0	0	0	0	0	0	0	0	1	0	0
S_4t_5	0	0	0	0	0	0	0	0	0	0	0	0	1	0
S_4t_6	0	0	0	0	0	0	0	0	0	0	0	0	0	1
S_4t_7	0	0	0	0	0	0	0	0	0	0	0	0	0	1

An absorbing node, S_4t_7, has been introduced to close the system. It might as well have been introduced as node S_4t_6. The first row of elements u_{kp} of the inverse $(I-Q)^{-1}$ follows, where the first subscript denotes the state and second the time increment:

$u_{11} = 1.0000$ $u_{33} = .0300$
$u_{12} = .9000$ $u_{34} = .0740$
$u_{13} = .7200$ $u_{34} = .4845$
$u_{22} = .1000$ $u_{44} = .0310$
$u_{23} = .2500$ $u_{45} = .3365$
$u_{24} = .8950$ $u_{46} = 1.0000$
$u_{25} = .1790$

The average number of time increments (and hence the average time) in state S_k is obtained by summing on the second subscript. Thus the average time in state S_2 is obtained by

$$\sum_p u_{2p} = 1.424 \text{ time units.}$$

All actions will have reached state S_4 after five time periods. The average time required to reach state S_4 (the absorbing state) can be determined from average times in each node associated with state S_4, i.e., S_4t_4, S_4t_5, and S_4t_6. The average times in node S_4t_4 is, in fact the probability of reaching S_4 in three steps. The probability of reaching S_4 in 4 steps is the average times in S_4t_5 minus the average times in S_4t_4, and so forth. Thus the average time required to reach the absorbing state S_4 is

$$3u_{44} + 4(u_{45} - u_{44}) + 4(u_{46} - u_{45}) =$$
$$3(.031) + 4(.3255) + 5(.6635) =$$
$$4.6325 \text{ time units.}$$

The extension to more complex cases is obvious.

One of the real difficulties with this approach is that the matrix dimensions become quite large. For example, if one uses an average of 15 time units for each state and there are 50 states in the model, a 750 x 750 transition matrix will result. However, with some minor adjustments it can be made triangular. Furthermore, most of the entries are zeros. A method has been developed for manipulating transition matrices of finite Markov chains which depends on the fact that they can be decomposed into linear functions of deterministic matrices (all entries zeros or ones).* We feel that this approach is worth investigating, although there has been insufficient time to try out the method in the preparation of this report.

Time distributions have been developed for the length of elapsed time between all of the pairs of states in the model previously presented. A few are shown in Tables 4 and 5. These tables indicate that one must distinguish between first, second, third, etc., actions and between destinations (i.e., next state). They also reflect the six months statutory limitation on time allowed for amendment. The Patent Office time distributions clearly reflect priorities given to final actions, usually for purposes of appeal. Time distributions between final rejection and appeal are now shown, but they very closely approach the six months statutory period.

A partial list of transition states and probabilities is shown in Table 6.

Table 4. Frequency distributions of elapsed time between receipt of application (or amendment) and examiner action (time in Patent Office).

Days	Original Application	First Amendment Not Final	First Amendment Final	Second Amendment Not final	Second Amendment Final	Third Amendment Not final	Third Amendment Final
0- 29	0	6	11	14	21	7	15
30- 59	11	3	4	15	6	1	11
60- 89	20	19	4	8	12	2	4
90-119	33	12	6	2	12	5	6
120-149	39	26	7	14	9	4	7
150-179	60	41	5	16	7	2	6
180-209	77	44	9	16	10	0	8
210-239	74	55	5	18	9	3	6
240-269	55	41	7	9	8	2	6
270-299	43	24	7	4	6	4	2
300-329	24	13	3	6	6	4	1
330-359	6	7	3	3	6	1	1
360-389	4	1	3	0	1	1	0
390-419	4	3	1	1	1	0	0
420-449	0	3	0	0	0	0	0
450 and over	0	4	0	1	0	0	0

Table 5. Frequency distributions of elapsed time between mailing of rejection and receipt of further amendment (time in hands of applicant).

Days	First Action	Second Action Not final	Second Action Final	Third Action Not final	Third Action Final	Fourth Action Not final	Fourth Action Final
0- 29	22	11	8	23	12	0	27
30- 59	12	8	5	3	4	12	7
60- 89	13	10	4	13	12	7	2
90-119	10	13	7	7	8	0	5
120-149	24	20	7	12	7	4	8
150-179	192	122	8	38	21	6	7
180-209	138	95	5	32	5	6	2
210-239	0	0	0	1	1	0	0
240-269	1	0	0	2	0	0	0
270 and over	1	0	0	2	0	0	0

Machine simulation of the system. The estimated probabilities of Table 1 and the time distributions, of which Tables 4 and 5 are samples, make it possible to simulate the system by electronic computer. For each input to the system (new application) one selects, by drawing random numbers proportional to transition probabilities, a path through the system. The delay at each station is determined by drawing random numbers proportional to

*See S. C. Gupta, "Manipulation of state transition matrices of finite Markov chains," IEEE Transactions on Electronic Computers, February 1963, pp. 19-20.

the time-frequency distributions for some suitable time interval, perhaps 20 or 30 days. The remainder of the program is an accounting scheme to keep track of applications during each increment of time. One must also enter data reflecting the current status of the operation, including applications in the backlog, in the hands of applicants awaiting amendment, in the Board of Appeals, etc. These data are available from administrative data of the Patent Office and from the time distributions, above.

After entry of the required estimated parameters, one simulates the operation of the system for a sufficient time to insure that his estimates do, in fact, reflect the behavior of the system. He then may vary some of the parameters, such as number of examiners, changes in statutory time limits, and so forth, to observe the effect on output and backlogs of these changes. The data which have been gathered and the preliminary analysis of the data indicate that the system can be simulated. If so, the potential of the simulation model for decision-making purposes is obvious.

Table 6. Partial list of transition states and probabilities (in parenthesis) for the transition model with fixed time increments of 30 days.

Present station	Next station (and probability)			
000.1	000.2 (.966)	001.1 (.034)		
000.2	800.1 (.957)	001.1 (.015)	100.1 (.022)	500.1 (.002)
	800.1 (.004)			
000.3	000.4 (.956)	001.1 (.004)	100.1 (.040)	
000.4	000.5 (.927)	001.1 (.006)	100.1 (.067)	
000.5	000.6 (.915)	001.1 (.002)	100.1 (.080)	900 (.003)
000.6	000.7 (.876)	100.1 (.122)	401.1 (.001)	800.1 (.001)
000.7	000.8 (.843)		100.1 (.157)	
000.8	000.9 (.846)	001.1 (.002)	100.1 (.150)	401.1 (.001)
	900.1 (.001)			
000.9	000.10(.887)	100.1 (.112)	800.1 (.001)	
000.10	000.11(.910)	001.1 (.002)	100.1 (.088)	
000.11	000.12(.951)	100.1 (.049)		
000.12	000.13(.988)	100.1 (.012)		
000.13	100.1 (.999)	800.1 (.001)		
001.1	001.2 (.935)	002.1 (.021)	101.1 (.044)	
001.2	001.3 (.965)	002.1 (.013)	101.1 (.022)	
001.3	001.4 (.857)	002.1 (.002)	101.1 (.141)	
001.4	001.5 (.912)	101.1 (.088)		
001.5	001.6 (.704)	002.1 (.100)	101.1 (.192)	751.1 (.004)

DESIGNS OF EXPERIMENTS IN INFORMATION RETRIEVAL

Donald W. King, Westat Research Analysts, Inc.

You have heard some of the inherent difficulties in the patent examining process. One of the major sources of difficulties is the novelty search which the patent examiner is charged to make. The R & D division of the U. S. Patent Office is committed to research, design and development of information storage and retrieval systems to provide the examiner with a workable search tool.

I am here with two purposes in mind. The first is to introduce the field of information storage and retrieval to members of the statistical profession as there has been a distinct lack of the use of statistical techniques in the research in this area. Secondly, I will discuss some experimental design techniques used in the research and development of such systems.

I will, first, briefly describe an information retrieval system. The system involves aggregating a set of books, publications, documents, etc., into a common library or file. An entire document can be put physically into this file or any portion of the document may be extracted and incorporated into a token or artificial file. The token file might include an abstract, classification, descriptions, facets, titles, authors or bibliographies from the document. A searcher may then look over a list of these artifices and request of the file all documents containing one or more artifices of interest to him or which he feels will most likely provide documents of interest to him. The problem then is to describe the documents in a token manner which will provide the examiner with a subset of documents which may be expected to satisfy his search needs. There are many combinations of possible systems. It is necessary, then, to seek out an "optimum" system from the standpoint of various trade-off parameters related to cost, reliability or system quality, and time. Time is distinguished from cost purposely.

The Patent Office is approaching the evaluation of examiner information storage and retrieval in three broad areas:

(1) Determination of the system user's needs.
(2) Determination of the system(s) which will "optimally" fulfill these needs.
(3) Determination of the effect of a new or modified information retrieval system on the entire examining process.

The research in the first area is currently limited to depth interviews and a questionaire survey in potential areas for examiner search aids. The second area involves advanced research into information storage and retrieval from the standpoints of computer software and hardware. Much of this research relates to advancement of the state of the art. In addition, much work is being done to determine the "optimum" systems under the present state of the art. Here, there are many problems to be solved such as defining meaningful file characteristics; establishing the most efficient general approach such as coordinate indices, classification, associative indices, etc.; and determining the best means of preparing the file.

One element of the third area has been discussed. This is the stochastic model describing the flow of applications through the system. A method of evaluating the quality of patent examining has been developed and is in use throughout the office. Some standard quality control practices are being employed in this effort. Another phase of this area involves investigation of the costing aspect of an information storage and retrieval system.* The developmental research for the information retrieved systems in the Patent Office involve a long dependent series of carefully conceived and designed experiments.

Today, I am going to discuss research involving the indexing phase of file preparation. Although research has begun in all of the areas mentioned above, it is further along in this area. The statistical techniques described are well known but are rarely used in research involving information retrieval systems.

Basically, this segment of research is devoted to measuring characteristics important to indexing. These are measures of:

(1) accuracy of indexing
(2) consistency of indexing
(3) time required to index a document.

The most important of these, and probably the most difficult to measure, is accuracy of indexing. Numerous attempts have been made by others to evaluate accuracy of indexing. The real point of departure of this research from other studies is the measurement of the effect of indexing errors in the retrieval of documents. This places the emphasis on the behavior of the system rather than on more philosophic grounds that errors are bad and therefore should be avoided.

*David A. Belsley, "The costing of information retrieval systems in the Patent Office through the application of a generalized costing structure," U. S. Department of Commerce, Patent Office, September 1962 (unpublished report).

(1) Single analyst

(2) Single analyst and a reviewer

(3) Two independent analysts
(set sum)

(4) Two independent analysts
(set intersection)

(5) Two independent analysts and a
reconciliation by a reviewer

(6) Three independent analysts
(set sum)

(7) Three independent analysts
(intersection)

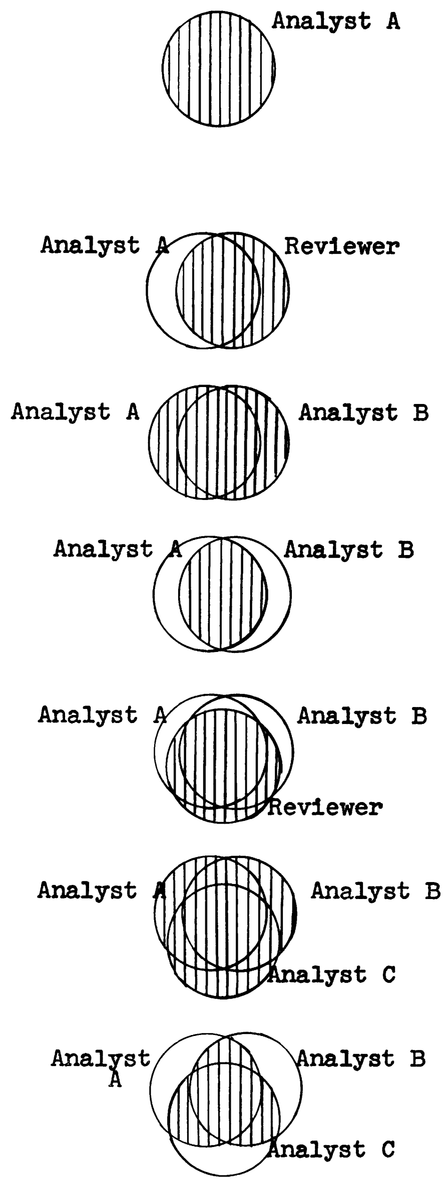


Figure 1. Venn diagrams of codes selected using various analyst modes.

Analyst Experience

Reviewer Experience	Over Three Months in Organometallics Art	Less than Three Months in Organometallics Art
High	$\alpha^{1}_{1(1).23(2)}$ $\alpha^{2}_{2(1).13(2)}$ $\alpha^{3}_{3(1).12(2)}$ $\alpha^{7}_{1(2).23(1)}$ $\alpha^{8}_{2(2).13(1)}$ $\alpha^{9}_{3(2).12(1)}$	$\alpha^{4}_{4(1).56(2)}$ $\alpha^{5}_{5(1).46(2)}$ $\alpha^{6}_{6(1).45(2)}$ $\alpha^{10}_{4(2).56(1)}$ $\alpha^{11}_{5(2).46(1)}$ $\alpha^{12}_{6(2).45(1)}$
Low	$\alpha^{13}_{7(3).89(4)}$ $\alpha^{14}_{8(3).79(4)}$ $\alpha^{15}_{9(3).78(4)}$ $\alpha^{19}_{7(4).89(3)}$ $\alpha^{20}_{8(4).79(3)}$ $\alpha^{21}_{9(4).78(3)}$	$\alpha^{16}_{10(3).11\,12(4)}$ $\alpha^{17}_{11(3).10\,12(4)}$ $\alpha^{18}_{12(3).10\,11(4)}$ $\alpha^{22}_{10(4).11\,12(3)}$ $\alpha^{23}_{11(4).10\,12(3)}$ $\alpha^{24}_{12(4).10\,11(3)}$

$\alpha^{h}_{i(j).ii(j)}$ Experimental arrangement of documents, analysts, and reviewers.

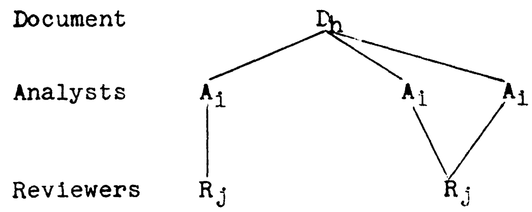
Figure 2. The experimental arrangement of the documents, analysts and reviewers for the organometallics intensive indexing experiment.

Indexing is defined as reducing information in the patent documents to a set of common identifiers and translating this set into unique codes for standard mechanized processing. There are two broad categories of indexing errors. Failure to select a code that should have been selected will be referred to as a Type I error, while selecting a code which should not have been selected is referred to as an error of Type II. Errors can also arise in the translation of information concepts into numerical codes. These, however, all result in errors of Type I or Type II.

We can assume the existence of conditional probabilities in a two-by-two contingency table with the actual selection of a code and whether the code should have been selected as the two bases for classification. A conditional probability that a code is not selected given that it should be corresponds to the Type I error and the conditional probability that a code is selected given that it should not be corresponds to the Type II error.

Some elementary mathematical models were derived to provide a means of estimating the correct retrieval and the false retrieval considering the conditional probabilities of indexing errors, the number of codes used in the search question and the logical relationship of the codes used in the search question.* The proportion of correct retrieval is highly sensitive to the conditional probability of selecting a code given that it should be selected, the number of codes used, and the use of search questions relating the codes conjunctively. Part of the purpose of the research reported here was to verify the simple model experimentally. The close correspondence between hypothetical and experimental data was quite revealing to the information retrieval staff. This was considered to be particularly important since many of the files in the Patent Office involve chemical compounds and coordinate index systems involving conjunctive descriptions of compound fragments.

The most important assumption of the models described was that the conditional probabilities of the codes used in a search question were independent. An experiment was designed to investigate this assumption to estimate the model parameters and other attributes of interest, and to determine the most efficient way of utilizing the indexing personnel. This experiment was conducted on 24 patent documents chosen randomly from the organometallic file containing a total of 3625 documents. The characteristics of interest were the conditional probabilities mentioned previously, measures of consistency of indexing, and estimates of the time required to index a patent document. The characteristics were observed for seven analyst modes. These are described pictorially in Figure 1 by Venn diagrams. The basic experimental arrangement is given in Figure 2. The order of the ith analyst and jth reviewer is given by their appearance in the diagram below:

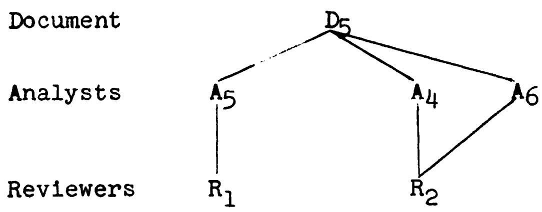


where: D_h, h=1,...,24, denotes documents.

A_i, i=1,...,12, denotes analysts

R_j, j=1,..., 4, denotes reviewers.

For example, the arrangement



appears in Figure 2 as

$$\alpha^{5}_{5(1).46(2)}$$

The primary consideration motivating this design is that an analyst or reviewer cannot repeat his efforts on a given document because of the learning effect.

The design is quite flexible in that all seven analyst modes can be observed, although only five operations are performed on each document. This can be done since the individual analyst's work can be observed prior to review. The single-double-and triple- analyst modes can therefore be generated accordingly. The results of the indexing can be incorporated into the file if the experiment is conducted as a file is being prepared.

Note in the arrangement above that the design provides a means for evaluating analyst experience and reviewer experience. Experience was defined as at least three months in indexing organometallic

*For a more complete discussion of the model, see E. C. Bryant, D. W. King and P. J. Terragno, "Some technical notes on coding errors," WRA PO 7, July, 1963 (informal report to the Office of Research and Development, U. S. Patent Office) documents. All of the analysts experience analyzing chemical compounds in one file or another. The comparison could just as easily have been made with regard to experience indexing, educational background, age, or sex, depending on the hypotheses under investigation. When a file is first being indexed all of the analysts will be inexperienced in the particular art, therefore, this evaluation provides an indication as to what may be lost in the initial phases of indexing that file and what improvement may be expected as the analysts gain experience in the new art.

It is clear that a distinct experimental arrangement is suggested for analysis of each analyst mode. The experimental designs, analysis of variance with expected mean squares, and observed values are given in an informal report to the Office of Research and Development of the U. S. Patent Office.* One difficulty in statistical analysis was that a larger document-to-document variation than was anticipated reduced the sensitivity of the statistical tests. An improved experimental design resulted ** and is used in a similar experiment involving an electrical transistor file.

A summary of the estimates of the conditional probability that a code is selected, given that it should be; the conditional probability that a code is selected, given that it should not be; the total number of codes selected; and total time to index the documents is given in Tables 1 through 4.

An experiment was conducted:

(1) to test the parameters, assumptions and validity of error retrieval models described previously,

(2) to determine the effect of indexing errors on retrievals for three analyst modes, and

(3) to determine if synthetic search questions can be used in evaluating a file

This experiment involved preparing a mechanized token file from 201 organometallic patent documents chosen randomly. Essentially three files were prepared by three analyst modes; i.e., the single-analyst mode, double-analyst (set-sum) mode, and single-analyst-reviewed mode. These three files were searched simultaneously using search questions formulated by examiners using the system in the past. Synthetic questions formulated by analysts were also used. Table 5 gives a summary of the average proportion of correct retrieval for the three analyst modes. These values are plotted by triangles in Figure 3. The model estimates of the proportion of correct retrieval found from the 24 document intensive indexing experiment are plotted linearly on the semi-log scale. The observed average proportion of correct retrieval and model estimates yield similar results for both the double-analyst (set-sum) mode and the single-analyst reviewed mode.

The model assumes that the probability of indexing two or more codes incorrectly is independent for all codes, i.e. $P_3(T_jT_k) = P_3(T_j)\ P_3(T_k)$: This can best be explained by the fact that the terms represent a portion of an entire compound. A further investigation of the errors made by the single analysts reveals that a large number of the errors involve omission of the entire compound rather than merely indexing one fragment of the compound incorrectly.

Joint probabilities of $P_3(T_j,T_k,\ldots,T_r)$ were estimated for one through four codes from the 24 document intensive indexing experiment. Observations of combinations of more than four codes were toorare to be used for this estimation. The scale of the "Average of original and repeated indexings" in Figure 3 gives the average observed proportion of correct retrieval (triangles), plotted least squares estimates of these values (linear plot), and the estimates of the joint probabilities for one through four terms mentioned above (crosses). It is seen that the joint probabilities are very close to those found using a least squares estimate of the observations. This demonstrates that it is quite important to test the assumptions of independence, particularly if there is reason to believe that the codes may be highly related, as they are when the indexing system involve fragments of compounds.

The estimates of the proportion of retrieval using the single-analyst-reviewed mode and the double-analyst (set-sum) mode are apparently representative of the actual observed retrievals. The assumption of independence becomes far less critical in these instances since a second person's review or independent analysis is involved and tends to cancel out the effect of the dependency.

Figure 4 also gives the proportion of correct retrieval over the range of the number of codes per search question for the original indexing and repeated indexing for the single-analyst mode. It is again noted that these plotted lines are least squares lines and are <u>not</u> derived from the model $Y=p_3{}^r$. The very small difference in these independent indexings

*E. C. Bryant, D. W. King, and P. J. Terragno, "Analysis of an indexing and retrieval experiment for the organometallic file of the U.S. Patent Office", WRA PO 10, August, 1963 (informal report to the Office of Research and Development, U. S. Patent Office)

**E. C. Bryant, D. W. King and P. J. Terragno, "Revised design for coding experiment, 307/88.5 file, "WRA PO 9, June 1963 (informal report to the Office of Research and Development, U.S.Patent Office.

Table 1. Estimates of p_3* for the various analyst modes with 95 per cent confidence limits (in parentheses)**

Analyst Mode	Experienced Analysts	Inexperienced Analysts	Combined Analysts
Single-analyst	(.86-.92)	(.78-.86)	.86
Double-analyst			
Set Sum	(.96-.99)	(.89-.93)	.95
Intersection	(.77-.83)	(.66-.73)	.75
Triple-analyst			
Set Sum	(.99-1.00)	(.95-.98)	.98
Intersection	(.88-.97)	(.75-.88)	.88
Single-analyst-reviewed			
Experienced reviewers	(.88-.97)	(.92-.99)	.94
Inexperienced reviewers	(.93-.99)	(.82-.93)	.93
Combined			.94
Double-analyst-reviewed			
Experienced reviewers	(.94-.99)	(.99-1.00)	.98
Inexperienced reviewers	(.92-.98)	(.81-.91)	.91
Combined			.95

*p_3 - is the conditional probability that a code will be selected, given that it should be

** Standard errors, on which the confidence limits are based, contain variation due to differences in documents and analysts.

Table 2. Estimates of the probability that a code will be selected, given that it should not be, p_2, for the various analyst modes.

Analyst Mode	Experienced Analysts	Inexperienced Analysts	Combined Analyst
Single-analyst	(.0002-.0007)	(.0024-.0036)	.0014
Double-analyst			
Set Sum	(.0008-.0013)	(.0011-.0016)	.0033
Intersection	(.0000-.0001)	(.0001-.0003)	.0001
Triple-analyst			
Set Sum	(.0008-.0037)	(.0067-.0113)	.0051
Intersection	(.0000-.0001)	(.0003-.0007)	.0002
Single-analyst-reviewed			
Experienced reviewers	(.0000-.00005)	(.0000-.0004)	.0001
Inexperienced reviewers	(.0000-.0004)	(.0007-.0029)	.0005
Combined			.0002
Double-analyst-reviewed			
Experienced reviewers	(.0003-.0022)	(.0011-.0041)	.0016
Inexperienced reviewers	(.0000-.0004)	(.0013-.0045)	.0008
Combined			.0012

Table 3. Estimates of the total number of codes selected for the various analyst modes.

Analyst Mode	Experienced Analysts	Inexperienced Analysts	Combined Analysts
Single-analyst	(132.4-136.9)	(136.7-141.2)	136.8
Double-analyst			
Set Sum	(146.1-152.4)	(159.8-166.0)	159.0
Intersection	(111.1-117.5)	(106.2-112.6)	111.8
Triple-analyst			
Set Sum	(94.3-227.7)	(120.6-254.0)	174.2
Intersection	(77.9-191.9)	(80.7-194.7)	136.3
Single-analyst-reviewed			
Experienced reviewers	(61.7-248.3)	(24.5-211.1)	136.4
Inexperienced reviewers	(41.7-228.3)	(74.2-260.8)	151.2
Combined			143.8
Double-analyst-reviewed			
Experienced reviewers	(56.1-250.9)	(35.5-230.4)	143.2
Inexperienced reviewers	(41.3-236.1)	(71.4-266.2)	153.8
Combined			148.5

Table 4. Estimates of total time (in minutes) required to index the documents for various analyst modes.

Analyst Mode	Experienced Analysts	Inexperienced Analysts	Combined Analysts
Single-analyst	(40.6-54.1)	(75.0-88.5)	64.6
Double-analyst			
Set Sum	(87.1-100.9)	(156.1-169.9)	128.6
Intersection			
Triple-analyst			
Set Sum	(43.4-238.9)	(147.6-343.1)	193.2
Intersection			
Single-analyst-reviewed			
Experienced reviewers	(0-201.9)	(0-215.7)	96.3
Inexperienced reviewers	(0-190.7)	(62.9-287.7)	126.8
Combined			111.6
Double-analyst-reviewed			
Experienced reviewers	(0-330.9)	(0-365.0)	156.2
Inexperienced reviewers	(0-325.4)	(96.1-479.5)	210.8
Combined			183.5

Table 5 A summary of the average proportion of correct retrieval in the searches of the 201 sample file for three-analyst modes.

No. of codes per search question	No. of searches	Ave. proportion of correct retrieval			
		single-analyst mode	single-analyst-rev. mode	double-analyst mode	single-analyst (joint probabilities*)
1	0	-	-	-	.90
2	3	.80	1.00	1.00	.89
3	12	.76	.82	.95	.85
4	31	.84	.85	.96	.80
5	29	.71	.79	.90	
6	36	.68	.78	.96	
7	54	.66	.79	.88	
8	30	.59	.79	.90	
9	46	.60	.76	.89	
10	35	.50	.59	1.00	
11	44	.52	.54	.79	
12	36	.47	.29	.83	

*The joint probabilities were estimated from the proportion of missed correct codes observed jointly two, three and four codes at a time.

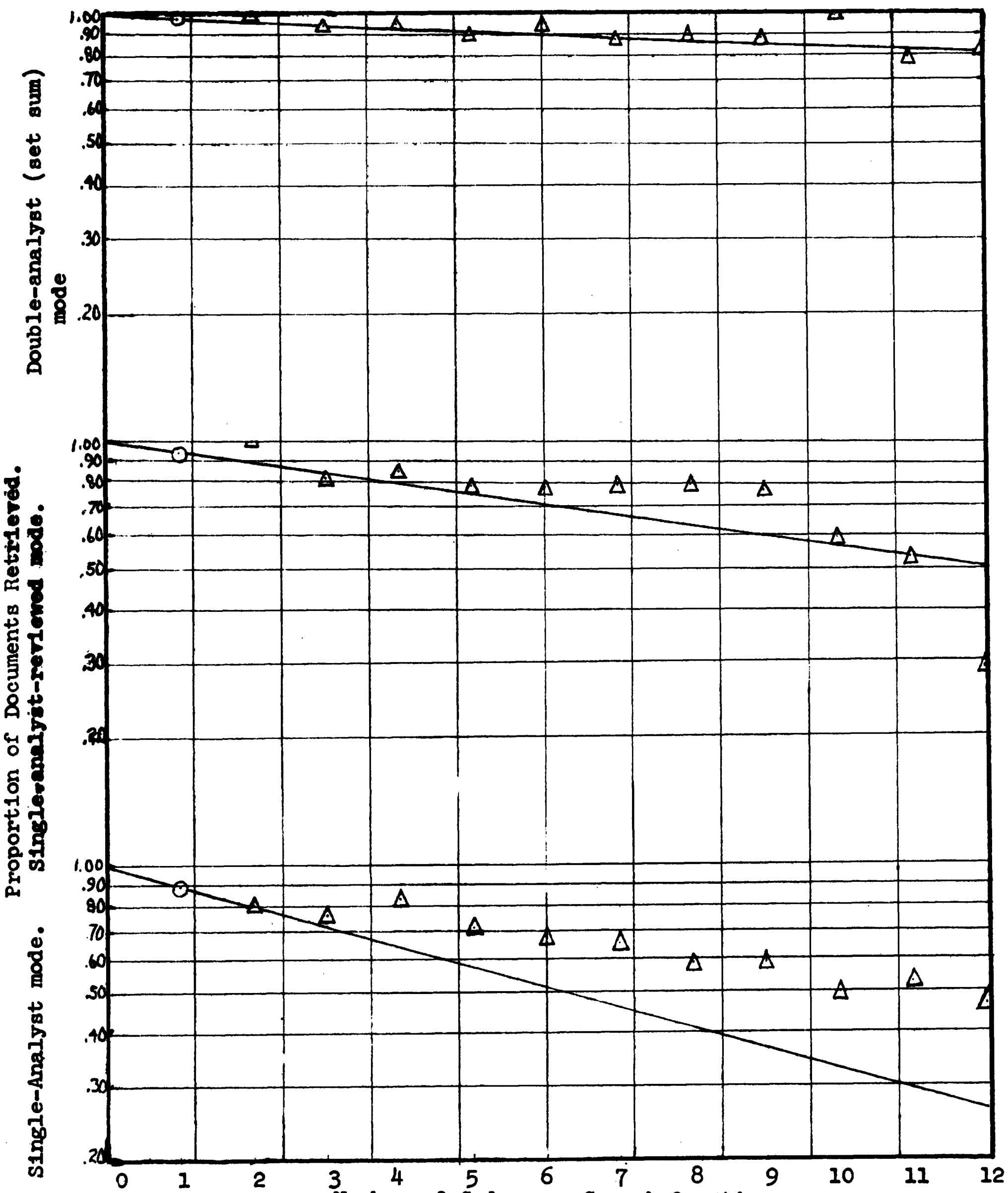


Figure 3 Actual per cent of retrieval compared with expected retrieval from the model $y=p_3^r$ for three analyst modes.

//

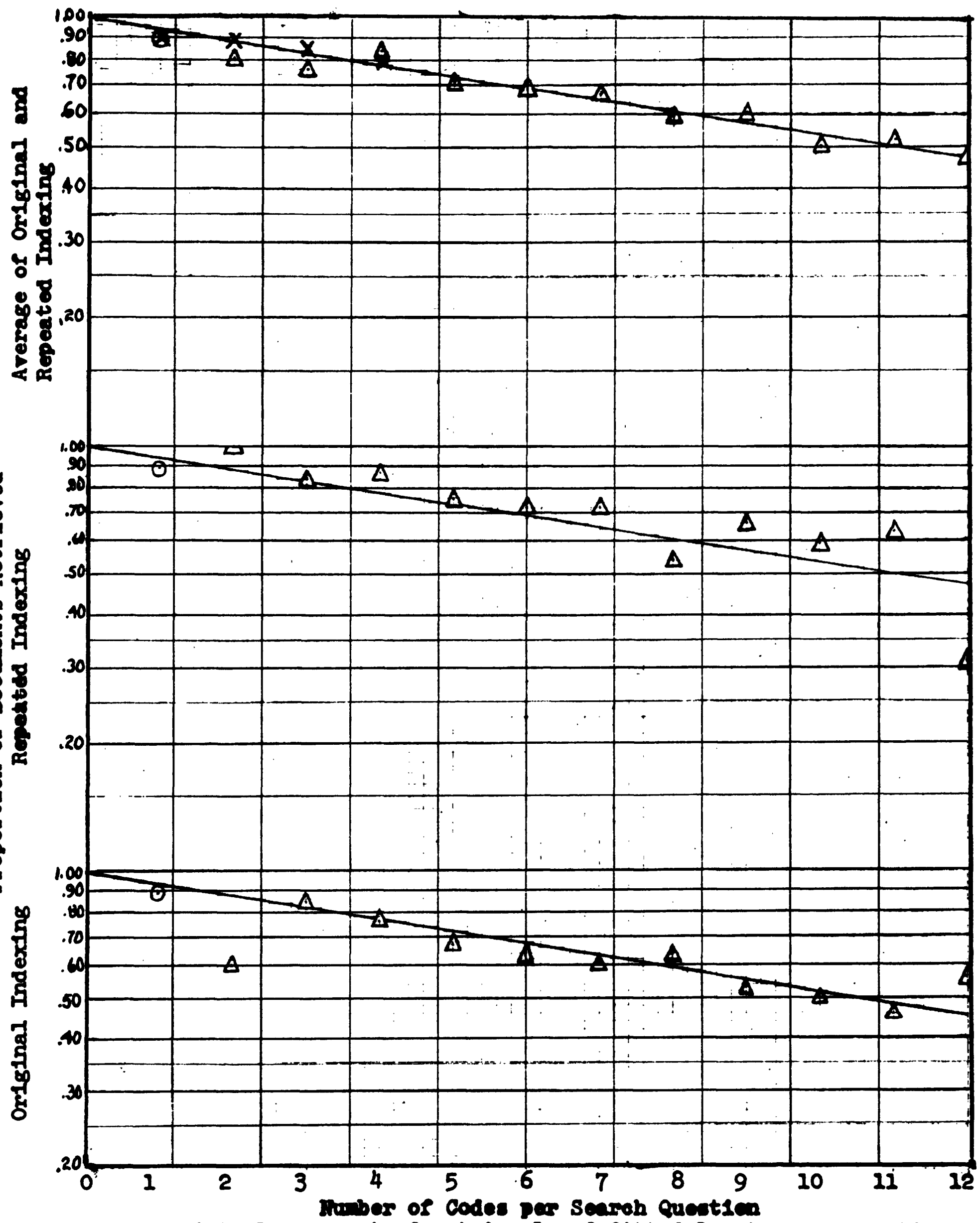


Figure 4 Actual per cent of retrieval and fitted least squares estimate for the original indexing, repeated indexing, and average indexing for the single-analyst mode.

indicates that indexing under experimental conditions (as was done in the repeated indexing) did not significantly affect the experimental findings. The behavior of the data from the intensive indexing experiment also supports the contention that the analysts operated normally even though they were aware that they were being observed.

The average number of correct retrievals, the average number of missed documents and the false drops observed in the searches conducted on the sample file of 201 documents is summarized in Table 6.

The experimentation reported here has evaluated various indexing modes in terms of indexing time, errors in the search file, and errors in retrieval. We have, then, bases for making rational decisions concerning the utilization of indexes so as to accomplish specific objectives. For three indexing modes used in the retrieval experiment, the "tradeoff parameters" are estimated as follows:

	Single-analyst	Double-analyst	Single-analyst-reviewed
Indexing time (Min)	64.3	128.6	111.6
Missed docs. per search	9.6	2.5	6.7
False retr. per search	4.5	9.2	5.6

With a given number of documents to be indexed, a given number of searches to be conducted per year, and some rough indication of the relative importance of missed documents and false retrievals one can determine the analyst mode which fits his particular needs.

Under the current Patent Office examining system the number of missed documents is so important as to lead one to choose the double-analyst (set sum) mode. Furthermore, this permits an easy evaluation of consistency of indexing, a parameter which can easily be controlled by usual methods of industrial quality control.

The importance of accurate coding in preventing missed documents has been demonstrated. Unfortunately it is difficult to assess accuracy of coding because of differences in opinion concerning what should be coded. Consistency of coding is relatively easy to measure, however, and, intuitively, it seems that consistency and accuracy should be closely related. It is clear that high accuracy implies a high degree of consistency between two coders, but that a high degree of consistency will only imply high accuracy if biases are not present.

Within the restriction that the so-called "correct" coding actually may not be correct, one can estimate the conditional probability, p_3, (that a term which should be coded will be coded).

Let $a_{ijk} = 1$ if, in the i^{th} document, term T_j was coded correctly by the k^{th} analyst. Let $a_{ijk} = 0$ otherwise. Let $b_{ij} = 1$ if the j^{th} term should have been coded in the i^{th} document, and let $b_{ij} =$ 0 otherwise. Then,

$$\hat{p}_3^{ik} = \sum_j a_{ijk} \Big/ \sum_j b_{ij} \qquad (1)$$

is an estimate of p_3 averaged over all terms in the i^{th} document and for the k^{th} analyst. Similarly, one can define

$$\hat{p}_3^{i} = \sum_{jk} a_{ijk} \Big/ k \sum_j b_{ij} \qquad (2)$$

where k stands for the number of analysts, as well as an index designator for them. This provides an estimate for the given document. For a given term one can construct the estimate

$$\hat{p}_3^{j} = \sum_{ik} a_{ijk} \Big/ k \sum_i b_{ij} \qquad (3)$$

and so on. An overall estimate is provided by

$$\hat{p}_3 = \sum_{ijk} a_{ijk} \Big/ k \sum_{ij} b_{ij} \qquad (4)$$

We are concerned here primarily with estimates similar to (3) and (4) above.

There are various approaches to estimates of consistency. Consider the following notation for two analysts:

$C_{00}^{ij} = 1$ if neither analyst coded j^{th} term in the i^{th} document.
$= 0$ otherwise

$C_{10}^{ij} = 1$ if the j^{th} term in the i^{th} document was coded by the first analyst, but not by the second.
$= 0$ otherwise

$C_{01}^{ij} = 1$ if the j^{th} term in the i^{th} document was coded by the second analyst, but not by the first
$= 0$ otherwise

$C_{11}^{ij} = 1$ if both analysts coded the j^{th} term in the i^{th} document
$= 0$ otherwise

Table 6 Average number of correct retrievals, missed documents and false drops per search for the 201 sample file for three analyst modes.

No. of codes asked	No. of Searches	Ave. no. of correct retrieval per search	Ave. no. of missed docs. per search			Ave. no. of false drops per Search		
			Single-analyst mode	Double-analyst mode	Single-analyst-rev. mode	Single-analyst mode	Double-analyst mode	Single-analyst-rev. mode.
2	3	1.67	.33	0	0	0	0	0
3	12	2.25	.54	.11	.42	.41	.83	.68
4	31	2.74	.45	.11	.42	.34	.65	.42
5	29	1.96	.57	.20	.41	.24	.62	.31
6	36	1.89	.61	.08	.42	.44	.81	.55
7	54	1.67	.57	.20	.35	.34	.68	.41
8	30	2.10	.87	.21	.43	.32	.70	.36
9	46	1.48	.60	.16	.35	.35	.61	.39
10	35	.63	.31	0	.26	.12	.17	.14
11	44	.93	.44	.19	.43	.08	.25	.11
12	36	.61	.40	.10	.30	.03	.06	.03
Ave. 7.9	32.4	1.54	.53	.14	.37	.25	.51	.31
95% Confidence Limits			.46-.61	.09-.19	.31-.43	.20-.31	.45-.57	.25-.37

This is shown as follows:

		1st Coder No	1st Coder Yes
2nd	No	c_{00}^{ij}	c_{10}^{ij}
Coder	Yes	c_{01}^{ij}	c_{11}^{ij}

Then, some measures of consistency which have been suggested are:*

$$CC^{i} = \sum_j c_{11}^{ij} \Big/ \sum_j \left[c_{11}^{ij} + c_{01}^{ij} + c_{10}^{ij}\right] \quad (5)$$

$$CC^{j} = \sum_j c_{11}^{ij} \Big/ \sum_i \left[c_{11}^{ij} + c_{01}^{ij} + c_{10}^{ij}\right] \quad (6)$$

$$CC = \sum_{ij} c_{11}^{ij} \Big/ \sum_{ij} \left[c_{11}^{ij} + c_{01}^{ij} + c_{10}^{ij}\right] \quad (7)$$

Some work done by the Census Bureau** on response differences is related to this problem. Here the emphasis is on an "index of inconsistency" which, for our purposes, we may define as follows:

$$II^{j} = \frac{n^j g^j}{c_1^j (n^j - c_1^j) + c_2^j (n^j - c_2^j)} \quad (8)$$

where n_j = number of documents over which consistency of the j^{th} term is being observed.

$$g^j = \sum_i (c_{01}^{ij} + c_{10}^{ij}) \quad (9)$$

$$c_1^j = \sum_i (c_{10}^{ij} + c_{11}^{ij}) \quad (10)$$

$$c_2^j = \sum_i (c_{01}^{ij} + c_{11}^{ij}) \quad (11)$$

By summation over j one can obtain an inconsistency index over all terms. Since we are primarily interested in consistency, rather than inconsistency, we use 1 - II to compare with CC, above, and with estimates of p_3.

*See, for example, J. Jacoby and V. Slamecka, "Indexer consistency under minimal conditions", Documentation Incorporated, 7900 Norfolk Avenue, Bethesda, Maryland, November, 1962.

**See Morris H. Hansen, William N. Hurwitz and Leon Pritzker, "The estimation and interpretation of gross differences and the simple response variance", U.S. Department of Commerce, Bureau of the Census, February 15, 1963.

In the sample of 24 documents there were four independent codings of each document; the original coding and three codings in the experiment. Thus there are $\binom{4}{2} = 6$ pairs of codings to be considered. While it is possible to define consistency in terms of the number of agreements among 3 coders (or 4) there is some merit in reducing all such measures to a two-coder basis. This was done by dividing the average (over 6 pairs) of the number of documents in which the given term was coded by two analysts by the average number of times it was coded by either. Thus, for the j^{th} term

$$CC^{j} = \sum_{\alpha}\sum_{j} c_{11\alpha}^{ij} \Big/ \sum_{\alpha}\sum_{j} \left[c_{11\alpha}^{ij} + c_{01\alpha}^{ij} + c_{10\alpha}^{ij}\right]$$

where indicates any of the six combinations of two analysts. A similar averaging process was used for the index of inconsistency.

As can be seen the number of times a term was coded by both members of a pair of analysts serves as the numerator for a coefficient of consistency. As a computational convenience, this variable is averaged over all possible pairs of analysts. An interesting statistical question arises concerning the variance of such an average.

Suppose that two independent coders encode a sample of n documents. We focus our attention on a particular term, T_j, and record the following:

m_{00} = the number of times coded by neither coder
m_{11} = the number of times coded by both coders
m_{10} = the number of times coded by coder 1, but not coder 2
m_{01} = the number of times coded by coder 2, but not coder 1

To obtain the variance of m_{11} we let $y_2 = 1$, if, in a single document, both coders select term T_j and $y_2 = 0$, otherwise. Let p = the probability that term T_j will be coded on a single trial (assumed to be the same for both coders).

Then

$$P(y_2 = 0) = 1-p^2$$

$$P(y_2 = 1) = p^2$$

$$Ey_2 = p^2$$

$$Ey_2^2 = p^2$$

$$\text{Var } y_2 = p^2(1-p^2)$$

Thus, in a sample of n independent document,

$$m_{11} = \sum_{i=1} y_{2i}, \text{ and}$$

$$\text{Var } m_{11} = n\, p^2(1-p^2)$$

Now, consider k independent coders, each of whom encodes the sample of n documents. We wish to obtain the variance of

$$\bar{n}_k = \frac{n_{12} + n_{13} + \cdots + n_{k-1,k}}{\binom{k}{2}} \tag{13}$$

That is, $\bar{n}_k$ is an average of the number of terms encoded in common over all possible pairs of coders. Let

y_k = 0 if no pair of coders has coded T_j

= 1 if exactly one pair of coders has coded T_j

= 2 if exactly two pairs of coders have coded T_j

etc.

Then,

$$p(y_k = 0) = (1-p)^k + k\,p\,(1-p)^{k-1}$$

$$p(y_k = 1) = \binom{k}{2} p^2 (1-p)^{k-2}$$

$$p(y_k = 2) = 0$$

$$p(y_k = 3) = \binom{k}{3} p^3 (1-p)^{k-3}$$

etc.

In general,

$$p(y_k = 0) = (1-p)^k + kp\,(1-p)^{k-1}$$

$$p\left(y_k = \binom{r}{2}\right) = \binom{k}{r} p^r (1-p)^{k-r}, \quad (r = 2, 3, \ldots, k)$$

$$= 0, \text{ otherwise.}$$

Hence,

$$E(y_k) = \sum_{r=2}^{k} \binom{r}{2}\binom{k}{r} p^r (1-p)^{k-r}$$

$$= \binom{k}{2} p^2 \sum_{r=2}^{k} \left\{ \frac{(k-2)!}{(r-2)!\,(k-r)!} p^{r-2} (1-p)^{k-r} \right\}$$

$$= \binom{k}{2} p^2 \tag{14}$$

$$\text{Var } y_k = E(y_k^2) - (Ey_k)^2$$

$$= \sum_{r=2}^{k} \left\{ \binom{r}{2}^2 \binom{k}{r} p^r (1-p)^{k-r} - \binom{k}{2}^2 p^4 \right\} \tag{15}$$

The first term of (15) can be recognized as the expectation of $1/4\, r^2 (r-1)^2$ which reduces to

$$1/4\, [k(k-1)(k-2)(k-3)p^4 + 4k(k-1)(k-2)p^3 + 2k(k-1)p^2]$$

Hence

$$\text{Var } y_k = \left\{\binom{k}{2} p^2 (1-p^2) + 2(k-2)\binom{k}{2} p^3(1-p)\right\} \tag{16}$$

Also,

$$\bar{n}_k = \sum_{i=1}^{n} y_{ki} / \binom{k}{2} \tag{17}$$

where the sum is over all n documents in the sample. Since the documents are independent (by assumption)

$$\text{Var } \bar{n}_k = \frac{n}{\binom{k}{2}} [p^2(1-p^2) + 2(k-2)p^3(1-p)] \tag{18}$$

If instead of $\binom{k}{2}$ pairs of related coders we had $\binom{k}{2}$ independent pairs of coders the variance would be just the first term of the above expression. Hence we lose sensitivity by taking all possible pairs among a group of coders.

It is more important, however, to think in terms of the amount of information supplied per coder. The number of coders required for $\binom{k}{2}$ independent pairs is $2\binom{k}{2}$ and for $\binom{k}{2}$ related pairs is k.

Therefore, the relative efficiency (per coder) of the related-pairs estimate to the independent-pairs estimate is

$$\text{R. E.} = \frac{\dfrac{2\binom{k}{2} n\, [p^2(1-p^2)]}{\binom{k}{2}}}{\dfrac{kn[p^2(1-p^2) + 2(k-2)p^3(1-p)]}{\binom{k}{2}}}$$

$$= \frac{2p^2(1-p^2)\binom{k}{2}}{k[p^2(1-p^2) + 2(k-2)p^3(1-p)]}$$

$$= \frac{(k-1)\ p^2(1-p^2)}{[p^2(1-p^2) + 2(k-2)p^3(1-p)]} \qquad (19)$$

The relative efficiency for k = 6 and k = 12 has been computed for various values of p as follows:

	p								
k	.1	.2	.3	.4	.5	.6	.7	.8	.9
6	2.9	2.2	1.8	1.5	1.4	1.3	1.2	1.1	1.0
12	3.9	2.5	2.0	1.6	1.4	1.3	1.2	1.1	1.1

Since, it is easy, computationally, to handle pairs as though they were independent, it is important to have a correction factor for the adjustment of such variances, to put them on an "independent-pairs" basis. Such an adjustment factor has been computed for various values of p and k and is shown in Table 7. The correction is made by multiplying the dependent-pairs variance by the correction term.

It will be recalled that the coefficient of consistency is, in effect, the set intersection of documents in which a given term is coded divided by the set union. That is,

$$CC = \frac{\bar{n}_k}{n - \bar{n}_k'} \qquad (20)$$

where $\bar{n}_k'$ is the average number of times in which the given term was not coded by both members of a pair, averaged over all possible pairs.

Table 7. Adjustment factors* for variances computed from related pairs of matched codes.

	k					
p	3	4	5	6	7	8
0.1	1.18	1.36	1.55	1.73	1.91	2.09
0.2	1.33	1.67	2.00	2.33	2.67	3.00
0.3	1.46	1.92	2.39	2.85	3.31	3.77
0.4	1.57	2.14	2.71	3.29	3.86	4.43
0.5	1.67	2.33	3.00	3.67	4.33	5.00
0.6	1.75	2.50	3.25	4.00	4.75	5.50
0.7	1.82	2.65	3.47	4.29	5.12	5.94
0.8	1.89	2.78	3.67	4.56	5.44	6.33
0.9	1.95	2.90	3.84	4.79	5.74	6.68

$$* \quad \frac{p^2(1-p^2) + 2(k-2)p^3(1-p)}{p^2(1-p^2)}$$

It is clear that the above means and variances for $\bar{n}_k$ will hold for $n - \bar{n}_k'$ if we replace p by $1 - p = q$, that is, the probability that a given term will not be coded.

We consider now the variance of the ratio given by equation (20). Consider random variables u and v whose true values are U and V, respectively, and consider the variable $W = u/v$ as an estimator of $w = U/V$. Then

$$\text{Var } w^2 = \left[W^2 \frac{\text{Var } u}{U^2} + \frac{\text{Var } v}{V^2} - 2\ \frac{\text{Cov } uv}{UV}\right] \qquad (21)$$

where Cov uv is the covariance between u and v and the other quantities are as defined previously. In our case, we identify u with $\bar{n}_k$ and v with $n - \bar{n}_k'$, and we must find Cov uv.

Consider k coders (analysts) and the following variables:

$y_k = 0$ if no pair of coders has coded

T_k

= 1 if exactly one pair of coders has coded T_k

= 3 if exactly three pairs of coders have coded T_k

etc.

z_k = 0 if no pair of coders has not coded T_k

= 1 if exactly one pair of coders has not coded T_k

= 3 if exactly three pairs of coders have not coded T_k

etc.

In general,

$$p(y_k = 0) = (1-p)^k + kp\ (1-p)^{k-1}$$

$$p\left(y_k = \binom{r}{2}\right) = \binom{k}{r} p^r\ (1-p)^{k-r},\ (\ r = 2, 3, \ldots, k)$$

As before, and

$$p(z_k = 0) = (1-q)^k + k\ q\ (1-q)^{k-1}$$

$$p\left(z_k = \binom{r}{2}\right) = \binom{k}{r}\ q^r\ (1-q)^{k-r} \quad (r = 2, 3, \ldots, k)$$

It can be shown that

$$\mathrm{Cov}\ y_k z_k = \binom{k}{2}\left\{\left[\binom{k-2}{2} - \binom{k}{2}\right]\ p^2 (1-p)^2\right\} \tag{22}$$

This covariance summed over n documents and averaged over all possible pairs of coders would be Cov $\bar{n}_k$, $\bar{n}'_k$. Hence

$$\mathrm{Cov}\ \bar{n}_k\ (n-\bar{n}'_k) = \frac{n\left[\binom{k-2}{2} - \binom{k}{2}\right]\ p^2(1-p)^2}{\binom{k}{2}} \tag{23}$$

where $\binom{k-2}{2}$ is defined to be zero for k=3.

Thus, to obtain an estimate of the variance of the coefficient of consistency, (equation 20), we replace, in equation (21),

W by $\bar{n}_k/(n - \bar{n}'_k)$

U by $\bar{n}_k$

V by $n - \bar{n}'_k$

Var u by eq. (18)

Var v by eq. (18) with p replaced by (1-p)

Cov uv by eq. (23)

and insert an estimate of p from the sample of n documents.

HOW PEOPLE INFLUENCE EXPERIMENTAL RESULTS [1/]

June Roberts Cornog, [2/] National Bureau of Standards

Information is stored so that people can retrieve it when there is need. If people cannot easily learn to manipulate the system by which information is stored when they want to know something, or if the system is undiscriminating enough so that it returns a good deal of unpertinent information when queried, or if the desired information isn't present in what is retrieved, then the whole storage process has been in vain.

If we extend the line of thinking begun in the above paragraph, certain other truisms must also be stated. Any system that people are to use can be validly tested: a) only by having people use it; b) by observing the users under some kind of controlled conditions so factual data about their performance can be gathered instead of just asking their opinions; and, c) by assigning them an experimental task characteristic of those for which the stored information would normally be required. These are obvious truths but their application to any testing situation is much less patent.

If we examine the problems connected with any experimental investigation that fulfills the requirements stated above, one major difficulty at once appears: people under observation do not behave as they normally would! The mere fact that they know they are being watched, or measured, changes their behavior. Considerable thought has been devoted over the years to how best to cope with the "human nature" of the people who must be used in experimental studies, but so far it is doubtful that any research administrator has ever been able to control his Subjects completely.

Changes in behavior due to observation may take many forms but in an experimental study such as that described below, five sources of behavioral variation are very likely to occur: (Refs. 2, 4, 5, 6, 7)

1. People try to help!

   They usually decide what they <u>think</u> the outcome of the study is <u>likely to</u> be and often they consciously or unconsciously weight their efforts in that direction!

2. They set up their own private operating rules.

   Every experimenter confidently lays down the rules of the game, with every expectation that they will be observed by all participants. But rules practically never cover every eventuality and strict inquiry soon reveals that participants are handling the unincluded, or are perhaps just interpreting the rules, in their own individual ways. Systematic error may creep in through loss of uniformity in operation!

3. The Subjects learn as the study progresses.

   People don't stay naive very long. If one of the factors in the experimental task is measurement of learning time, or if the performance of naive Subjects is to be compared with that of experienced people, such individuals can be used only once. The performance or working habits of experienced people often change as they continue to work at jobs they do all the time. If the objective is to measure the effectiveness of a particular method, any changes in work habits or any learning on the job will introduce error.

4. People are never able to be totally consistent in a judgment task.

   The point does not need to be expanded. Every reader will recognize the minor daily variations in his own decisions.

5. People under observation usually show signs of stress.

   They become more careful, give more attention to detail, or, in the case of a manual task such as typewriting, commit an unusual number of mechanical errors.

The U.S. Patent Office research and development program is concerned with the testing of mechanized information storage systems. The information stored is that identifying the inventions in patents already granted. Retrieval of pertinent information may occur when the already patented ideas are searched to determine the novelty of the idea contained in the new application. The data storage systems are usually unique to the "art" they are designed to accommodate so that storage of information about, say electronic transistors, must be quite differently handled than information about chemical insecticides.

After a storage system is designed and at least partly applied, its true applicability to the art it is supposed to store, as well as its usefulness to the patent examiner, must be assessed. This can be done only by having the future users use it - and, of course they are people! Patent examiners are, in addition well above average in intelligence, education and analytical skills so that the kind of "cover story" usually used in experimental situations is useless. In the study described below it was necessary, therefore, to state the purpose straightforwardly and simply to appeal to the good sportsmanship of the participants, by asking them to behave as normally as possible.

The influence of people on the experimental results may be illustrated by a Patent Office research project. One of the chemical "arts" is known as the organometallic group. A project to store the information from these patents in a mechanical system was undertaken.

After the system of encoding and storage was designed and partially carried out, a study of the encoding process was devised in order:

1. To examine the effects of coding on ultimate retrieval of patents (would the encoding used produce all pertinent references and only the pertinent ones upon demand. Discussion of this objective is not included in this paper) and,

2. To derive a method for establishing a criterion for satisfactory coding before full-scale work in this area was begun.

All the available coding people were divided into two as nearly matched groups as possible. All persons were trained simultaneously and their understanding of the work tested in advance. All Subjects worked at their own desks under as normal conditions as possible, each participant was reassured that only mass data would be used - that no one's record would affect his job standing in any way. Only one part of the task was a little out of the ordinary - each coder was asked to keep a record of the time he spent on each case.

For a description of the way the study materials were handled and how the criterion for accuracy in the study was established, I quote the paper delivered by Mr. King a few minutes ago, Ref. (1), Sect. B.

> Briefly, 201 documents were drawn at random from the total file of 3,625 documents. Each of the documents had, during the construction of the file, been coded and reviewed by an experienced analyst, the reviewed encoding being included in the file. Twenty-four of the 201 documents were drawn at random for an intensive coding experiment. Each of the remaining 177 documents was encoded again by random assignment to one of four experienced analysts. Then the two most experienced analysts, with the original encoding, the reviewed encoding and the second encoding before them, selected what, in their combined opinion, was the "correct" encoding. Half of the 24 documents were assigned, at random, to three inexperienced analysts, and half to three experienced analysts. ("Experience" refers to length of time analysts had previously devoted to indexing organometallics documents. All were experienced chemical analysts.) Then the same two senior analysts who selected a correct encoding for the 177 documents made the same determination for these 24, having before them the original encoding, its review, and three more independent encodings. Thus, for 177 documents one can compare actual and experimental coding against a coding defined to be correct. For each of these, measures of consistency between two analysts (the original and the experimental analyst) are available. For each of the 24 documents one can compare the original coding and three experimental codings against a standard and can compute measures of consistency among three experimental codings and an original coding.

The difficulty of a patent was judged by two standards - The chemical compounds included and the sheer thickness of the document. All Subjects were asked not to confer with each other (and asked later to affirm that they had not) when they were told that members of each group of three persons would be duplicating each other's work. Measurement was in terms of:

1. Amount of time spent on each patent;

2. Number of terms encoded;

3. Conditional probabilities as a measure of accuracy;

4. Two indices of indexing consistency.

The work covered a period of four days. Other details of statistical methods may be found in Ref. (1).

Analysis of the data gathered showed some interesting contrasts which illustrate the principles stated earlier. The summary statements were based on both the factual data and on observations made by the administrator as the study progressed. Even though these sources of error were noted, they did not significantly affect the outcome of the study.

1. There was more difference in performance within one of the two experienced coder groups than between the experienced and inexperienced groups, as measured by total number of codes produced, accuracy of those codes and time required in encoding the document. Even though some of the differences among the experienced coders may have been due to variance in the difficulty of the patents worked on, learning among the inexperienced coders occurred so quickly that the pre-test training sessions evidently reduced experience as a test factor. The worst coder in terms of accuracy was one of the experienced people!

2. The time spent on single cases varied between 10 minutes and 3 1/2 hours! The difference in the thoroughness characteristic of the individual coders was undoubtedly strongly felt in these time measures, plus some influence from the variations in their educational backgrounds. The most obvious implication is the previously cited difference in the difficulty of the individual patents encoded, of course.

3. Some Subjects consistently encoded larger numbers of terms than others did but this measure was not related to amount of previous experience. Furthermore, when work performed under the close observation of experimental conditions was tested against the numbers and kinds of codes produced in the course of normal work, additional differences in performance were found although these were not statistically significant.

4. When some reviewers knew which coder's work they were reviewing, they were inclined to become more or less critical according to which coder was involved. Criticism was indicated by a larger number of codes added or deleted.

5. Adequacy of coding was necessarily a matter of judgment but the reviewer tended to go along with the work the coder had done, especially if he didn't know who the coder had been. From the standpoint of omitted terms, two independent analyses or encodings were shown to give better results than one analysis and one review.

Error attributable to the sources listed above was not observable in the data from the groups, but existed in the scores of individuals according to which coder was involved. It is probable that a good deal of the variance among individuals was due to personality variables since no correlation existed between scores and either education or amount of experience.

Consistency among judges was unusually good -- of all the terms encoded inconsistently, only 12% were considered to be ambiguous by the judges. In one other Patent Office information coding and retrieval study, however, there was as much disagreement among the judges as among the coders!

Some further explanation is in order here as to why ambiguous terms are unusually likely to occur in Patent Office work. Inventions always deal with new ideas, with advances in science or technology, with material which has not yet had time to become standardized. Inventors frequently must invent not only the gadget or process but the technical language to describe its function or purpose. If several applicants arrive at nearly the same idea at approximately the same time, they don't necessarily use the same terms to describe it. The patent examiner and coder must look to the concept involved and try to reach a standard language for themselves. Ambiguities are therefore likely to be more numerous than the "outsider" would expect.

The organometallics study yielded one important observation as well as some unusually high consistency measures:

1. It was observed that people who seemed to have greater patience for detail did the best job of encoding patents.

   They were more accurate -- they missed a smaller number of terms and carried through the analysis of chemical compounds more thoroughly. They apparently had a goodly share of the well-known compulsion to "get it right!" Neither the amount of experience nor the educational background of the coder correlated with the quantitative measures taken.

2. Participants showed high levels of accuracy when their work was compared with the criterion codes (3). (See Table I.)

TABLE I

ESTIMATES OF ACCURACY AND CONSISTENCY AND OF LINEAR CORRELATION BETWEEN ACCURACY AND CONSISTENCY BY TERMS FOR SAMPLES OF 24 AND 201 DOCUMENTS.**

	24 documents	201 documents
	Linear correlation coefficients	
a. A measure of accuracy vs. a measure of consistency*	.88	.86
b. A measure of accuracy vs. the Consistency Coefficient*	.90	.95
	Estimates of these characteristics	
c. A measure of accuracy*	.823	.816
d. First measure of consistency*	.862	.726
e. Accuracy squared	.698	.675
f. Consistency Coefficient*	.729	.665

The Standard Errors of these estimates do not exceed .02 in any instance.

*See Ref. (1), Sec. B, for explanation of measures used.

Inspection of data in Table 4, Ref. (1), Sec. B, shows that accuracy and consistency measures of the encoded terms were closely related. The same personality trait probably influenced both measures.

From the data gathered in the organometallics coding experiment, the investigators were able to predict the parameters of the model in the small study and to test the prediction model against the actual, observed retrievals. Apparently the sources of human error described here did not significantly affect the experimental factors.

**Data taken from Tables 3 and 4, and from explanation for Tables 3 and 4, Sec. B, Ref. (1).

Bibliography:

1. Bryant, E.C., King, D.W. and Terragno, P.J., "Designs of Experiments in Information Retrieval," Proceedings of the Social Statistics Section, American Statistical Association, 1963.

2. Edwards, Allen L., "Experiments: Their Planning and Execution," in Handbook of Social Psychology, Vol. 2, pp. 259ff, ed. by Gardner Lindzey, Addison-Wesley Publishing Co., Cambridge, Mass., 1954.

3. Jacoby, J. and Slamecka, V., "Indexer Consistency under Minimal Conditions," Documentation Inc., Bethesda, Md., November, 1962.

4. Jahoda, M., Deutsch, M. and Cook, S.W., Research Methods in Social Relations, Part One, pp. 92-127, Dryden Press, New York, 1951.

5. Jahoda, M., Deutsch, M. and Cook, S.W., Research Methods in Social Relations, Part Two, pp. 463-487, Dryden Press, New York, 1951.

6. Sidman, M., Tactics of Scientific Research, pp. 341-393, Basic Books, Inc., New York, 1960.

7. Underwood, R.J., Psychological Research, pp. 17-48, Appleton-Century-Crofts, New York, 1957.

Footnotes:

1/ Prepared in connection with a special project sponsored by the U.S. Patent Office, on procedures for orienting patent examiners toward non-manual searching methods.

2/ Research Psychologist.

VII

SURVEY SAMPLING AND SAMPLING ERRORS

Chairman, Frederick F. Stephan, Princeton University

CHANGING STRATA AND SELECTION PROBABILITIES*

Leslie Kish, The University of Michigan

Summary

Survey samples are often based on primary units selected from initial strata with probabilities proportional to initial measures M_j. But later samples could better be served with new strata and new measures P_j. The difference between the initial and new strata, and measures, may be due to changes in either the population distribution or in survey objectives. Efficiency dictates retaining in the new sample the maximum permitted number of initial selections. A procedure for changing measures within fixed strata is presented first. Then we develop new procedures for retaining the maximum number of selected units within changed strata. Finally, some problems of selecting units with unequal probabilities are explored.

1. Introduction

Unequal selection probabilities are often assigned to sampling units. Especially in the first stage of survey samples, primary sampling units are often selected with probabilities proportional to their size measures, N_j. Then subsampling, within the selected primaries, of final units with probabilities inversely proportional to the N_j, yields a uniform overall selection rate f:

$$\frac{N_j}{c/f} \times \frac{c}{N_j} = f. \qquad (1)$$

The c and f are constants chosen to yield the desired numbers of primary and final units. Typically the selections are made separately in many strata, with one or two (seldom more) primary selections specified from each stratum.

After the initial selection, the primary units may be used for many surveys. But the initial size measures may differ considerably from new measures that would better suit the needs of current surveys. The difference between initial and desired new size measures may be due to differential changes of size among the sampling units, as revealed by the latest Census. But it can also be due to differences in survey populations. For example, the initial measures may have been based on persons, but now one may want to sample a population of physicians, or college students, or farmers, or values of farm products. If measures of size for the new population are available, and if these differ considerably from those used in the initial selection, then one should prefer to use them.

Furthermore, the best strata for the new population may be sufficiently different from the initial strata to justify changing them. The same changes or differences in population distributions, that lead to changing selection probabilities, also will often motivate changing strata.

A brand new selection would mean changing most of the primary units. But continued use of the initial primary units has several advantages. First, they represent important investments in clerical work, and especially in a selected and trained field force. The field force in each county of a national sample may represent a value of many hundreds of dollars. Secondly, using the same primary units decreases considerably the standard errors of comparisons between surveys, (see data by Kish [3]). This problem is acute for panel samples.

To continue using the entire initial set of primary units would restrict the sample to the initial set of selection probabilities and strata. But one can shift to the new probabilities and strata, and yet retain most of the initial primary selections. To make that shift from the initial to a new set of probabilities and strata, with a minimum number of actual changes of primary units, is the purpose of the methods presented.

Section 2 describes a simple method for changing measures of size within fixed strata, modified in section 5 to introduce further considerations of economy. Section 3 presents methods for retaining initial selections from changed strata, and is the main purpose of this paper; these are derived and further discussed in section 4. Finally section 5 treats some related problems in multiple selections with unequal probabilities.

2. Changing Measures Within Fixed Strata

This section presents a procedure for changing from the initial probabilities p_j to new probabilities P_j for a fixed set of sampling units within a stratum. There are D + I units in the stratum. Of these, D receive a decrease in probability, and I receive an increase, which may also be zero. Subscripts

*This research was done at the Survey Research Center, (Institute for Social Research) with the aid of grant USPHS/GM09889-02 from the National Institutes of Health. The author is grateful to Irene Hess, Vinod K. Sethi, and Roe Goodman for their helpful suggestions.

distinguish the two sets of sampling units, so that

$$P_d < p_d \text{ and } P_i \geq p_i, \qquad (2)$$

where $(d = 1, 2, \ldots, D)$ and
$(i = D + 1, D + 2, \ldots, D + I)$.

The initial probabilities used for the selection of sampling units from a stratum, and the new probabilities to which we want to switch, both sum to unity in the stratum.

$$\sum^{D} p_d + \sum^{I} p_i = 1 = \sum^{D} P_d + \sum^{I} P_i. \qquad (3)$$

The sum of the probability increases must equal the sum of the probability decreases:

$$\sum^{I} (P_i - p_i) = \sum^{D} (p_d - P_d). \qquad (3')$$

The rules for changing probabilities follow.

a. If the initially selected sampling unit shows an increase (or no change in probability), retain it in the sample, as if selected with the new probability P_i.

b. If the initially selected unit shows a decrease, retain it in the sample with the probability P_d/p_d; that is, assign a probability of $1 - P_d/p_d$ for dropping it. Thus the compound probability of original selection and remaining is made $p_d \times P_d/p_d = P_d$.

c. If a decreasing unit is dropped from the sample, select a replacement among the increasing units with probabilities proportional to their increases. The probability of selection of the i-th unit is $(P_i - p_i)/\sum^{I} (P_i - p_i)$. Thus, the total selection probability of an increasing unit consists of p_i initially, but it is properly increased if any of the D decreasing units had been selected:

$$p_i + \sum (p_d - P_d) \times \frac{P_i - p_i}{\sum^{I} (P_i - p_i)} = P_i. \qquad (4)$$

Often only one sampling unit is selected from each stratum. But the rules are equally valid for two or more fixed numbers of selections from the stratum. This method was first presented by Keyfitz [1]. It is illustrated, amplified and modified in section 5.

3. Changing Stratum Boundaries

When the sampling units of a population are shifted across stratum boundaries, from a set of initial strata to a set of new strata, we face new problems of practical urgency, whose solution motivated our search. A key step consists in separating the procedures of this section from those of section 2. First, stratum changes are solved with section 3 procedures to obtain the <u>preliminary probabilities</u> p_j; then these are changed with section 2 procedures to new <u>final probabilities</u> P_j. Of course, $p_j = P_j$ if the final measures equal the original measures; and section 3 without section 2 suffices for situations when strata are changed but measures are not.

We concentrate here on starting with a single initial selection from each initial stratum and ending with a single final selection in each new stratum, and this represents a limitation. But the number and the sizes of the initial strata may differ from those of the new strata, and this represents an important situation for sample surveys.

For brevity and clarity, <u>strata</u> will denote the new strata. Each of these is composed of several <u>sets</u>, which denote the portions of initial strata it contains. Each set contains either zero or one initial selection. We must assume that the units are sorted into the new strata with procedures that guarantee that the initial selection of some units has no effect on their sorting. This can be guaranteed either with definitions that permit no latitude, or by a performer ignorant of the identity of selected units, or both for extra safety. The guarantee is necessary <u>to ensure that selection probabilities before sorting are equivalent to selection probabilities after sorting into the new strata</u>. Except for this guarantee, there is complete freedom in forming the new strata.

We shall denote with M_j the probabilities that the units received in their own initial strata. These measures were probabilities and summed to 1 in the initial strata, but their sum $\sum M_j = M$ in the new stratum is not generally 1.

This sum varies from stratum to stratum, but for brevity we shall denote it as M, without a subscript for the stratum. The measures M_j must be converted to the preliminary probabilities $p_j = M_j/M$, with $\Sigma p_j = 1$. Note that selecting with probabilities proportional to the measures M_j would be equivalent to selecting with probabilities proportional to the preliminary probabilities p_j. Then these could be converted, with section 2 procedures, to the final probabilities P_j. But this would be only a long procedure equivalent to selecting directly with the probabilities P_j. Both procedures disregard our purpose of retaining in the new sample the initial selections.

First Procedure. Suppose now that the stratum consists of several sets, and denote both these sets and their measures as $A + B + C + \ldots = M$. The typical set A has the measure $A = \Sigma_A M_j$, the sum of the measures M_j of the units it contains. First select a set with probabilities proportional to its size, so that set A has the selection probability A/M. If the selected set contains the original selection, accept it as the preliminary selection with $p_j = (A/M) \times (M_j/A) = M_j/M$. If the selected set does not contain the initial selection, select one with probabilities proportional to M_j (or p_j). In either case, compute p_j and convert to P_j, with the section 2 procedures applied to the entire stratum (not to only the selected set).

Understanding this procedure is simple because it depends on only a two-stage selection for the probability $p_j = M_j/M$, and the equivalence (guaranteed) of the possible prior selection with the M_j.

Second Procedure. In the first procedure the sets are selected proportional to their initial probabilities (Σp_j), and the conversion from p_j to P_j is applied to the entire stratum. The second procedure selects with probabilities proportional to the final probabilities (ΣP_j) in the sets, and the conversion from p_j to P_j is confined within the selected sets. The Census Bureau switched from a 333 area sample based on 1950 data to a 357 area sample based on 1960 data, and described its procedure as follows [5]:

> "For generality, consider a revised stratum in the 357 area design made up of parts from more than one stratum in the 333 area design. In principle, one of these parts can be selected with probability proportionate to 1960 size. If, then, the selected part does not include a PSU previously selected in the 333 area design, a selection can be made from PSU's in that part with probability proportionate to 1960 size. On the other hand, if the selected part contains a sample PSU from the 333 area design, selection from the part can be made by the Keyfitz method [1], maximizing the change of retaining the previously-selected PSU."

Third Procedure. This depends on identifying uniquely each of the initial strata with only one of the new strata. Then section 2 procedures are applied for changing measures. All units in the new stratum that did not belong to the specified initial stratum are treated as newly created, with $p_j = 0$; and some new strata may consist entirely of such new units. On the contrary, units of the initial stratum that did not transfer to the specified new stratum are treated as eliminated from it, with $P_j = 0$. Rules are needed to specify the unique identification of the initial strata with the new strata; they must avoid selection bias and preferably should maximize the retention of selected units.

The first two procedures have the advantages of simplicity, that may facilitate their adaptation to designs with further demands. Two of these deserve mention. First, some designs call for two or more selections from each stratum. Second, in some designs the selections from different strata are not independent, because some form of "controlled selection" or "multiple stratification" is used.

However, those two simple procedures have the disadvantage of not utilizing initial selections present in other sets, when the selected set fails to contain one. This disadvantage is moderate if one set predominates within each of the strata. But the disadvantage increases rapidly with the number of sets of the same general magnitude present in the strata.

Fourth Procedure. This procedure is optimal in the sense that it retains all initial selections, subject to conversions with the section 2 procedures. It retains all initial selections, whenever one is present in any set of the stratum. It requires the assumptions of a single selection from each stratum, and independence of selections between strata. These points are developed, together with the justifications for the following selection rules, in section 4.

a. If no initial selection occurs in any set in the entire stratum, select one unit directly with the new probabilities P_j.

b. If a single initial selection occurs in the entire stratum, accept it as the preliminary selection with $p_j = M_j/M$. Then convert p_j to P_j with section 2 procedures applied in the entire stratum.

c. If two or more sets contain initial selections, select one set with probability y. From the selected set accept the selected unit, and convert from the p_j to P_j with section 2 procedures applied to the entire stratum.

   1. When there are only two sets in the stratum, with total measures A and B, select one set with odds <u>inversely</u> proportional to its measure; that is, select set A with probability $y_a = B/(A + B)$, and set B with probability $y_b = A/(A + B)$.

   2. When the stratum contains more than two sets, order all sets into a series of dichotomies, with an objective ordering. Whenever both branches of the dichotomy contain selections, choose one with the rules symbolized with:

$$y_a = \frac{B + (a - b)}{A + B} , \tag{5}$$

where $a = \frac{A}{A'} = \frac{C + D}{C' + D' - C'D'}$

and $b = \frac{B}{B'} = \frac{E + F}{E' + F' - E'F'}$.

To compute the factors A', B', then C', D', and E', F', etc., follow and continue these rules:

If A contains only a single set, A = C and D = 0, then A' = A and a = 1. Similarly for B. If both A and B contain single sets, we have a = b = 1, and the simple case of two sets in the stratum.

If A contains two sets, A = C + D, and C and D contain only a single set each, then C' = C and D' = D and A' = C + D - CD. Similarly for B = E + F.

If C contains two sets C = G + H, then C' = G' + H' - G'H'. If G contains a single set then G' = G; but if G contains two sets, G = T + S, and G' = T' + S' - T'S'. Similarly for H and for D = I + J.

These rules would become cumbersome if carried far. But in practice, rarely will two sets contain selections, when the stratum has many sets.

An objective rule is needed to prevent personal bias in ordering the dichotomies. One reasonable rule is: (a) A set containing more than half of the stratum establishes the first split. (b) Order all others by size; divide them successively into two parts with equal numbers of sets, putting odd numbers into the second half. For example, a stratum composed of a majority set X, and of 9 other sets (denoted with their ranks) would be divided as follows:

$$\{X\}\ \{[(1,2)(3,4)][(5,6)(7,8|9)]\}.$$

4. Justification and Expansion of the Optimum Procedure

To derive the optimal fourth procedure, regard it as the modification of the first procedure, whose justification is obvious. The selection of one of the sets from the stratum with probability proportional to its measure is made to depend on the probability of its containing an initial selection. Based on those probabilities the rules must achieve the equivalent of selecting the set A with probability proportional to A/M.

Let us begin with a stratum containing two sets only, with the measures A + B = M. The presence of only one set would be merely the special case of B = 0. We shall maintain the odds of choosing between the two sets at A/B, because this ratio of the measures is the basis for the selection within the sets. We use the important fact that these measures A and B also represent the probability for each set that it contains the initial selection from the initial stratum. Hence, if we can assume independence between the two probabilities A and B, we have:

$(1 - A)(1 - B)$ = probability that neither set contains a selection (6)

$A(1 - B)$ = probability that set A, but not set B contains a selection

$B(1 - A)$ = probability that set B, but not set A contains a selection

AB = probability that both sets contain a selection.

Our strategy shall be as follows: When neither set contains a selection we choose between the two sets with the desired odds A/B. When selection is present in only one set, we choose it with certainty as desired. But this introduces an imbalance in the odds. When selection is present in both sets, choose set A with odds $y_a/(1 - y_a)$, computed so as to redress the imbalance introduced before.

In other words, the desired odds A/B are achieved over all possibilities by achieving them separately for the case of no selection, and jointly over the case of one or two selections. We need to solve for y_a in the expression that equates the desired odds A/B to the compound probabilities of the occurance of a case and the selecting of the set:

$$\frac{A(1 - B) \times 1 + (1 - A)B \times 0 + AB \times y_a}{A(1 - B) \times 0 + (1 - A)B \times 1 + AB \times (1 - y_a)} = \frac{A}{B} \qquad (7)$$

and $(1 - B) + By_a = (1 - A) + A(1 - y_a)$.

Thus when both sets contain selections, set A should be chosen with probability

$$y_a = \frac{B}{A + B}, \quad \text{or odds} \quad \frac{y_a}{1 - y_a} = \frac{B}{A} \qquad (8)$$

Direct extension of the above method proved too complex even for three sets. But one can arrange a larger number of sets into a series of dichotomies, and then apply in sequence procedures similar to those for two sets. For example, a stratum containing four sets can be divided into group A containing sets C and D, and into group B containing sets E and F. We can choose group A with the odds $A/B = (C + D)/(E + F)$, then select C with odds C/D from $A = C + D$, and finally a unit within C with the probability M_j/C. The overall probability of selecting the unit would be $A/(A + B) \times C/(C + D) \times M_j/C = M_j/M$.

We want an equivalent procedure in which the odds A/B, C/D, and E/F are preserved. A new problem arises in the probability that a group of two sets contains either one or two selections. The probabilities for the two groups are not simply $A = C + D$ and $B = E + F$; they are $A' = C + D - CD$ and $B' = E + F - EF$ respectively. But we can deal more conveniently with the odds A/B, proportional to the sums of measures, than with A'/B'. Hence we must adjust accordingly the probabilities y_a and $(1 - y_a)$ for choosing between the two groups, when both contain selections. We set these so as to attain, jointly for double selections and single selections, the desired odds A/B:

$$\frac{A'(1 - B') + A'B'y_a}{(1 - A')B' + A'B'(1 - y_a)}$$

$$= \frac{Aa'(1 - Bb') + Aa'Bb'y_a}{(1 - Aa')Bb' + Aa'Bb'(1 - y_a)}$$

$$= \frac{A}{B}, \qquad (9)$$

where $A' = Aa'$ and $B' = Bb'$. Then $a'(1 - Bb') + Ba'b'y_a = (1 - Aa')b' + Aa'b'(1 - y_a)$ and $y_a(Aa'b' + Ba'b') = (b' - a') + Ba'b'$.

Thus when both groups A and B contain selection group A should be chosen with probability

$$y_a = [B + (1/a' - 1/b')]/(A + B), \text{ or}$$

$$y_a = \frac{B + (a - b)}{A + B}, \text{ where } a = \frac{A}{A'} \text{ and } b = \frac{B}{B'}. \qquad (10)$$

Thus $(a - b)$ represents an adjustment of the probability of choosing group A over group B, when both contain selections, and when each is composed of two or more sets. Of course, group B would be chosen with $(1 - y_a) = [A + (b - a)]/(A + B)$. Note that a value of $(a - b) > A$ would result in the impossible choice of $y_a > 1$; similarly $(b - a) > B$ would mean $y_a < 0$. But this cannot occur if A and B are both not greater than 3. Since this requirement will be met easily in practice, we desist from pressing further solutions for this problem.

The presence of only three sets in the stratum means that group A contains only one set $A = A'$ and $a = 1$. The presence of only two sets means $a = b = 1$, and (10) becomes (8). When there are only four sets, and $A = C + D$ and $B = E + F$, then

$$A' = C + D - CD \quad \text{and} \quad B' = E + F - EF \qquad (11)$$

Here C and D, and E and F all represent the probabilities for each of the four sets, that it contains one or more selections.

When there are more than four sets in the stratum, the formation of dichotomies can be continued as needed. Any and all of the four groups, C, D, E, and F, can be considered as having been composed of two sets. In general then, instead of (11) we write

$$A' = C' + D' - C'D' \text{ and } B' = E' + F' - E'F' \quad (12)$$

where each of C', D', E', and F' must denote the probability that the group contains one or more selections. For example, if C contains a single set, $C' = C$; if it contains exactly two sets, $C = G + H$, then $C' = G + H - GH$; if both G and H are groups of sets then $C' = G' + H' - G'H'$, where G' and H' are similarly defined.

These procedures for single selections per stratum would become too complex if applied to multiple selections from strata. More direct and general solutions seem to be possible if some strong symmetries are imposed on the strata and on the selections. But generalizations in this direction are not helpful if they impose conditions that are typically absent in practical work.

The simplest example would be the initial selection with simple random sampling of n units from the entire population of N units; that is, all $\binom{N}{n}$ combinations equally probable. These can be sorted into arbitrary strata with procedures that guarantee no effect from the initial selections on the sorting. The n selections can be accepted as equivalent to random selections within the strata. The stratum samples then can be increased or decreased at random to obtain the numbers of random selections needed from each stratum. The symmetries of simple random selections are sufficient, and no more complicated proofs are needed.

If the initial selection was comprised of simple random selections within the initial strata, then they are also random selections within the corresponding sets of the new strata. Treating the sets as substrata within the new strata, some treatment can be evolved to yield a sample approximating a stratified random sample within each new stratum. But for selecting units with unequal probabilities this is likely to become too complicated to be practical.

The problem of selecting two (or more) units with unequal probabilities from a stratum can be simplified by dividing the stratum into two (or more) random substrata of equal size. This can be done with a random sorting of units, as in the first procedure of section 6. Selecting one unit from each substratum reduces the problem to that of a single selection per stratum. The initial strata and the new strata can both be treated this way. We should add that after the single selections in the new strata are secured, other selections can be added either with replacement, or by using one of the procedures of section 6 for selecting without replacement.

Single selections per strata characterize much practical work, and to these the optimal procedures are directly applicable.

The requirement of independence between strata is violated when controlled selection is used. I fear that it would be too difficult to achieve formal treatment of the joint probabilities for the several sets within strata; but this seems most unlikely as a source of genuine bias. However, imposing a controlled selection on the new strata seems difficult to achieve.

5. Modifications and Illustration of Changing Measures Within Fixed Strata

a. A simpler procedure than that of section 2 can handle a problem confined chiefly to large growth in a small proportion of the primary units. For example, suppose we have a sample of blocks selected with the initial probabilities p_j, proportional to the initial sizes N_j. Suppose also that for a new sample one is willing to retain the original probability p_j for all blocks, except for those that have at least doubled in size. That is, if the new size $N_j' < 2N_j$, we accept the initial p_j; but for the <u>growth</u> blocks, with $N_j' \geq 2N_j$, we want new probabilities P_j proportional to N_j'.

Generally, place into new <u>growth strata</u> the portion $(N_j' - N_j)$, denoting the size increase of primary units designated as <u>growth units</u>. Then select from these growth units a sample with probabilities proportional to the values $(N_j' - N_j)$. In most situations, the constant of proportionality will be the same (c/f) as for the initial selection, to preserve the uniform overall sampling fraction f; but a different fraction can be introduced for the growth stratum.

Thus the growth units are selected with the sum of two probabilities in the growth strata and in the initial strata:

$$\frac{N_j' - N_j}{c/f} + \frac{N_j}{c/f} \qquad (13)$$

If selected, these primaries should be subsampled with the rates c/N_j' so that the overall selection probability becomes:

$$\left(\frac{N_j' - N_j}{c/f} + \frac{N_j}{c/f}\right) \times \frac{c}{N_j'} = f. \qquad (13')$$

Note that, as desired, the planned subsample is c̲ when a unit becomes selected. This can occur both in the initial strata and in the growth strata. The procedure amounts to splitting the growth units into two parts, consisting of an initial size N_j plus a growth size $(N_j' - N_j)$, and subjecting them to separate selections. The ordinary units, that do not qualify as growth units, remain subject to the initial selection rates $N_j/(c/f) \times c/N_j = f$.

b. Into the procedures of section 2 one can introduce further considerations of efficiency, knowing that generally it is neither necessary nor possible to have precise measures of size. It is desirable to have probabilities approximately proportional to size measures. And it is necessary and sufficient for applying the procedures of section 2 that the sum of changes in probabilities be zero in the stratum, (so that $\Sigma P_j = \Sigma p_j = 1$ can be satisfied). Within that requirement we can adjust the selection probabilities to satisfy some criteria of change large enough to be recognized as important, and to deliberately neglect smaller changes.

For many sampling units the changes in probabilities are small enough that they can be deliberately neglected. If these units are numerous, and if neglecting their changes adds little to the overall variance, then one can reduce the number of changes significantly with little sacrifice in the variance. To these units one can reassign the initial probabilities; they constitute a set of units for which $p_i = P_i$ is arbitrarily assigned. This "flexible" procedure reduces the number of units one must switch; it also eliminates the task of revising office records for selected units with no change in probabilities. But this "flexible" procedure requires more computations for balancing the probabilities within the strata than the rigid "strict" plan does. The overall advantage of the flexible plan is important when it confines expensive switching to a small proportion of units. This may require tolerating rather wide limits of differences in actual measures for the units to which $P_i = p_i$ are arbitrarily assigned.

We used the flexible procedure in changing from the 1940 to 1950 Census measures for the 54 counties, representing as many strata in the national sample of the Survey Research Center. In choosing criteria we balance the costs of changing counties against the increase in survey variances due to small distortions in the selection probabilities. These were considered relative to other sources of variations in sample size. We decided on the following procedure: (a) Define an important increase as 10 per cent or more $(P_j/p_j \geq 1.10)$. Compute the sum of the increases in the stratum. (b) Add enough of the largest decreases (smallest values of P_j/p_j) to balance <u>exactly</u> the sum of the increases; that is, $\Sigma (P_i - p_i) - \Sigma (P_d - p_d) = 0$. (c) Consider all other counties as not having changed probabilities, with $P_i = p_i$. The details are in [2].

Two other modifications deserve brief mention. Faced with rather small probabilities of change $(1 - P_d/p_d)$ in each of 16 strata, we did not draw independently within each stratum. Instead we controlled the number of changes by cumulating the probabilities of change from one stratum to another, then applied an interval of one, after a random start. Thus the actual number of changes, which was three, was controlled within a fraction of the expected number of changes.

Second, we need not merely accept the changes from one Census period to another; we can also project them forward into the middle of the period of the use of the Census frame. This may be worth doing for the fastest and steadiest growing counties. For example, the California counties that have shown unusual growth in one decade may also be expected to show similar growth in one decade may also be expected to show similar growth in the next decade. The projection may improve the design of a sample of counties for use during that decade.

6. Some Related Problems

When selection probabilities and strata are changed, the situation may also require changes in the boundaries of sampling units. Some may be split, and others combined. The initial selection probabilities p_j of a unit may be divided among several portions according to specified rules; similarly several of the p_j may be combined into one new unit. The creation of entirely new units is noted with $p_j = 0$, and the elimination of initial units with $P_j = 0$. But we must avoid a detailed treatment of this problem.

Selecting more than one unit with probabilities proportional to unequal measures is difficult to do "without replacement". (It is even more difficult to find simple formulas for computing the lower variances that selection without replacement yields.) This problem has been the subject of several investigations, most recently by Fellegi [6] and Rao [4]. He discusses the first procedure below, but the other four procedures are new, I think. For brevity two selections per stratum are discussed, but the methods can be extended to more selections.

1. Sort the units of the stratum at random into two equal substrata and select one unit from each, with probability proportional to P_j. This can also be accomplished by putting the units in random order and selecting two with a systematic interval.

2. Select from the stratum with equal probability and without replacement 8 units and put them in the same random order in which they were selected. Obtain the sum of their measures, Σ. Now select a random number from 1 to $\Sigma/2$; then add $\Sigma/2$ to it. These two numbers are the two selection numbers from the stratum. The number 8 is suggested merely as a desirable compromise: we want to save labor, but we also want to avoid situations in which double selection is possible, because a unit is larger than $\Sigma/2$.

3. When selecting units with equal probability, reselection of any unit is avoided by substituting some unselected unit. This is simple because of the complete symmetry of equal probabilities, which is generally lacking in units with unequal probabilities. But we can use the presence of partial symmetries, when at least two units appear with the same size, for any size, in the same stratum. If one unit is selected twice, select at random one of the others with the same size. If exactly equal sizes do not exist, generally one can establish strict rules for "best matches" of sizes. When two sizes P_j and P_k are matched, the uniform subsampling rate $\underline{f}$ may be obtained by subselecting in both with $(P_j + P_k)/2$, whenever both units appear in the sample.

4. First select $K = 6$ units using the desired measures and with replacement, which is generally easy. Then sort these at random into three pairs, but without allowing the same unit twice into a pair. Now select one of the pairs at random. The probability for a unit with $M_j = 0.10$ of appearing twice in two draws is 1/100; but its appearing more than three times in $K = 6$ draws is reduced to about 1/800. For $M_j = 0.05$ the reduction is from 1/400 to about 1/12,000. (If K sampling units are selected with probabilities proportional to measures P_j, subselection of k units with equal probabilities yields a sample of k units with probabilities proportional to the P_j.)

5. An initial selection with equal probabilities would be a special case of the procedures in section 2, with the $p_j = 1/N$ for a single selection, and $p_j = 2/N$ for 2 selections; then convert these to probabilities proportional to the P_j. This procedure reduces but does not eliminate the probability of duplicate selection of the same unit.

References

[1] Keyfitz, Nathan, "Sampling with Probabilities Proportional to Size", Journal of the American Statistical Association, 46(March, 1951), pp. 105-109.

[2] Kish, Leslie and Hess, Irene, "Some Sampling Techniques for Continuing Survey Operations", Proceedings of the Social Statistics Section, American Statistical Association, Washington (1959).

[3] Kish, Leslie, "Variances for Indexes from Complex Samples", Proceedings of the Social Statistics Section, American Statistical Association, (1961), pp. 190-199.

[4] Rao, J. N. K., "On Three Procedures of Unequal Probability Sampling Without Replacement", Journal of the American Statistical Association, 58(March, 1963), pp. 202-215.

[5] U. S. Bureau of the Census, "The Current Population Survey - A Report on Methodology", Technical Paper No. 7, (1963). U. S. Government Printing Office, Washington, p.67.

[6] Fellegi, Ivan P., "Sampling with Varying Probabilities Without Replacement", Journal of the American Statistical Association, 58(March, 1963), pp. 183-201.

A CONTROLLED SELECTION WHICH PERMITS UNBIASED ESTIMATES OF SAMPLING VARIANCES

Roe Goodman, Bureau of the Census

Controlled selection has been utilized in the sampling for a number of statistical surveys during the dozen or so years since this method of sampling was introduced. The original application was in the selection of a nation-wide sample of primary sampling units by the Survey Research Center of the University of Michigan. (1) That sample of PSU's, or an up-to-date version of it, was used extensively by the Survey Research Center over a period of perhaps ten years. Other uses of controlled selection have been made by the Census Bureau in the Current Population Survey (5), by the Bureau of Labor Statistics in the city sample for the Consumer Price Index (4), and by the University of Michigan's Bureau of Hospital Administration in the sampling of hospitals and hospital patients (2). A current application of the method is that being made by the Bureau of the Census in the selection of PSU's for agricultural surveys to be conducted as a part of the program of the 1964 Census of Agriculture.

Present and past uses of the method of controlled selection have been undertaken despite the fact that there has generally been no information in the particular instance regarding the gains (or losses) which may have resulted from the use of this sampling procedure as compared to some alternative sampling method. Evaluation of the gains (if any) which are achieved with the use of controlled selection has generally been difficult. Nevertheless the method seems to have been used because it was believed that the survey results would be at least as reliable as those which could have been obtained under an alternative sampling scheme.

As far as estimates of sampling variances are concerned, with controlled selection recourse is ordinarily made to the same types of approximations as are utilized for other sampling designs for which the sample data alone do not yield unbiased estimates of variances. One of the methods which is widely used is the method of collapsed strata. The method of collapsed strata was first devised for the case of stratified random sampling when the stratification was carried to the point that only one unit had been selected within each stratum. Similarly, with controlled selection sampling variances may be computed considering the sample as though consisting of sets of two or more units selected at random within collapsed strata, the estimated variances then being computed based on the variability among sample units within a set, with the use of suitable weighting factors. This method seems generally to serve well enough for purposes of approximating standard errors of survey estimates both for the ultimate case of stratified sampling (one unit per stratum) and for the case of controlled selection as well.

At some point however it becomes necessary to go farther than merely to approximate sampling variances and sampling standard errors and to determine rather definitely the relationship between the sampling variances found under controlled selection and those which obtain if some alternative sampling procedure is used. A logical standard for comparison in such instances is that of the sampling variance under stratified random sampling. The need for these comparisons of the variances is especially real due to the fact that variance estimates under the method of collapsed strata are on the average over-estimates of the true variances. Moreover, the theory shows that the greater the gain which has been achieved by the last stage of stratification or by the refinements of controlled selection the greater will be the degree of over-estimation of the variances with the method of collapsed strata. From the estimates of sampling variances and sampling standard errors obtained in practice then we are left with no clue whatsoever concerning the gains which may be achieved by the use of the controlled selection.

It has frequently been recognized that complete data for a population, such as Census data, permit the drawing of repeated samples and the preparation of estimates for each sample and that estimates of sampling variances can then be computed directly simply by computing the variance among the different sample estimates. This approach has frequently been used experimentally and, in fact, use of this method accounts for what little is known about the gains from the use of controlled selection to date. Considerable further use may yet be made of precisely this approach to the problem. However, it is undeniable that during the past decade very little work has been done along this line - in part no doubt for reasons to be explained below.

The present study is an attempt to progress toward the preparation of more useful estimates of sampling variances with controlled selection than have been available in the past. The results to be presented here and those being derived in our present studies (but not so far available) do depend upon information relating to a number of PSU's in addition to those in the original sample. However, the approach is more one of estimating variances from sample data and does not require the use of repeated samples. The present approach has considerable promise in that often times the drawing of repeated samples is not practicable. In the first place the sampling process may be extremely intricate involving several steps and the use of controls at each step. For this reason the drawing of additional samples becomes very laborious. In addition, more and more there is the desire to do the kind of thing Dr. Kish has just been talking about, namely, to build a new sample upon an old sample, retaining as many unite of the previous sample as possible. As discussed by Dr. Kish the process involves changing the composition of strata and even re-defining some of the primary sampling units in the population. Moreover, controlled selection may have been used in the original sampling and it is now to be used again in the revision of the sample in order to bring it up to date. The procedure about to be described therefore seems a logical way to go about studying the gains achieved from the use of controlled selection.

For the benefit of those who may not be entirely clear on the exact distinction between some of these sampling procedures let us begin with a simple illustration of the use of controlled selection. See Table 1. In this example it is assumed that it is desired to select a sample of Standard Metropolitan Statistical Areas for the United States, or some major region, for use in multi-purpose sample surveys of households. In the illustration there are given three strata consisting of PSU's in Virginia, Maryland, Delaware, New Jersey and the eastern part of Pennsylvania but excluding the major cities of Washington, Baltimore, Philadelphia and New York City (the part in New Jersey). It is assumed that these large cities would be selected with certainty and hence that there would be no sampling of them at the first stage. It is assumed that there are other strata containing the SMSA's in other parts of the Region covered by the surveys, and that these other strata would be sampled also.

In the illustration the sampling probabilities were computed so as to be proportionate to the total population of each SMSA. In order that the sum of the probabilities for each stratum should be exactly 1.000 the chances of selection for occasional PSU's are divided a part being placed in one stratum and the remainder being placed in another stratum. The SMSA Newport News-Hampton is split in this way; it has a total probability of 0.150 of which 0.011 is placed in the first stratum and 0.139 in the second. The main consideration in the original stratification was the size of the central cities within the SMSA. In the columns to the right can be seen the possible samples which may be selected under one controlled selection scheme. Each possible sample consists of three places, reading across. The probability given beside the city in the last column indicates the probability of selection of the particular sample. Thus a sample consisting of Norfolk, Reading and Wilkes-Barre has a probability of 0.160 and finally one of Newport News, Scranton and Lancaster has a probability of selection of 0.011. The samples containing Norfolk have probabilities adding to 0.387, those containing Richmond have probabilities adding to 0.273 - and so on - and the sum of the probabilities for all possible samples is of course 1.000. From examination of the PSU's it can be seen that SMSA's located in Virginia have probabilities adding to 0.990 (0.387 plus 0.273 plus 0.150 plus 0.106 plus 0.074). Therefore every sample contains one and only one PSU in Virginia with the exception of the one Allentown, Wilmington and Lancaster which has a probability of 0.010. Looking at the coastal cities and seaports it can be seen that the sum of the probabilities for Norfolk, Newport News, Wilmington and Atlantic City is 0.890. Again no two of these cities appear in the same sample except for the sample consisting of Norfolk, Scranton and Atlantic City with a probability of 0.11. Samples having total probabilities of 0.121 then contain none of these port cities. On the whole a good control of the selection of port cities has been achieved. The fact that one possible sample contains both Norfolk and Atlantic City illustrates another common characteristic of controlled selection, namely, that with this method the goals sought are approached but usually not fully achieved. (Often if one goal is fully achieved another goal has to be sacrified somewhat).

Now for comparison with stratified random sampling it may be noted that, if the selections were to be made independently within the three strata, Norfolk and Wilmington would have a probability of (0.387) (0.245) or 0.095 of appearing in the same sample whereas with the controlled selection the probability of this joint occurrence is zero. In the case of Norfolk and Atlantic City the probability of their appearing in the same sample with stratified random sampling is (0.387) (0.108) or 0.042 compared with the 0.011 mentioned above. Numerous other comparisons can easily be made in the example at hand which would show clearly the effects of using controlled selection as an alternative to the well-known stratified random sampling.

Let us turn now to the method of estimating sampling variances which has been devised at the Bureau of the Census for purposes of the present analysis. See Table 2. In a generalized solution for sampling with varying probabilities, given by Horvitz and Thompson (3) ten years ago it was shown that to obtain unbiased estimates of sampling variances from sample data it is necessary that each pair of units in the population should have had a chance of appearing in the same sample. The solution in the present instance then is to supplement a sample chosen by the method of controlled selection by adding one or more units in such a way that every pair in the population has a chance of being in the amended sample. The idea of supplementing a sample, say a systematic sample, in such a way as to fulfill this condition is not new although it is doubtful if results based on this approach have been published. The exact method used now is first to select a simple random sample of ℓ strata out of the L strata and then within each stratum so selected to select one additional PSU. Prior to the selection of the additional PSU the one originally selected is "replaced", thus permitting the selection of the same one a second time. Under these conditions the formulas as given in the second table apply and it is now theoretically at least possible to obtain unbiased estimates of the sampling variances using data for the sample as amended.

A few remarks concerning the formulas and their purposes may be useful at this point. First, please note that in the estimated sample total, $\hat{X}$, only the data for PSU's in the original sample are used. $\hat{X}$ is defined in this way because the purpose is to estimate the variance for precisely the sample as originally selected. Data for the additional PSU's are to be used only for purposes of estimating the variance but not for estimating the population total, X. The reasons for this decision stem from the fact that the estimated variance is itself subject to a great amount of variability and the kind of analysis being made must therefore be considered as a laboratory project rather than a procedure which would be used in the conducting of actual sample surveys. Since the formulas are to be applied in cases in which Census data are available for every PSU in the population the possibility exists of supplementing the original sample with a large number of additional units in order to obtain the desired degree of precision of the estimated variances.

From the variance formula itself the importance of the difference, $x'_{h(1)} - x'_{h(2)}$, may be noted. Basically, what is done is to obtain this difference between the estimated stratum total in each stratum for which a second PSU is selected and then to use this difference, both to obtain sums of squares (the first term) and sums of cross-products (the second term). In the second term the sign of the difference is important of course since it is multiplied by the estimated stratum total, based on the original sample for every <u>other</u> stratum, and then summed. A feature of the formula is that the first term reflects the variance of an estimated total for stratified random sampling and the second term the gain (or loss) from the use of controlled selection. In order to have a gain then the sum of the crossproducts, the second term of the formula, must be negative.

It may be noted that it is easily possible to set ℓ equal to L, that is, to select an additional unit within each stratum. In any use of the formula when ℓ is not extremely small it is to be expected that in some strata the second PSU selected will be the same as the original selection. In such cases both the squared term and the sum of crossproducts naturally become zero for the particular stratum. The proof that the estimated variance is unbiased is a simple one as shown in C. From the second step to the third step the formula becomes simplified due to the fact that the expected values of so many of the product terms are equal to zero. The selection of the additional PSU's independently from stratum to stratum accomplishes this result even though the PSU's in the original sample are not selected independently within the different strata.

The results obtained to date with the use of this formula have been found to be of little value due to the extreme variability of the estimated variances. It has been found that estimated variances computed from samples of no more than 36 PSU's and supplemented by an additional set of 36 PSU's, still do not yield meaningful results. Until a sample of adequate size is used many estimated variances turn out to be negative and it is clear that no satisfactory measures of gains or losses can be derived unless much larger samples are used.

At the Census Bureau experience has now been gained with the use of this formula as applied to data for past Censuses and a computer program has been tested and utilized in the experimental work done to date. We will now proceed to utilize the sample of some 400 PSU's, supplemented by as many as 400 PSU's, and perhaps even supplemented all over again by an additional 400, in an attempt to obtain reliable estimates of sampling variances with the particular controlled selection being used in the new agricultural sample. Meanwhile, there is room also for the possible development of other estimates of sampling variances, estimates which need not necessarily be unbiased provided the estimates are consistent and the bias is not unduly large. It appears then that satisfactory solutions to this problem will be possible, especially with the aid of the computers.

Table 1.--CONTROLLED SELECTION -- ILLUSTRATION
Population Consisting of SMSA's, m = 3
(Assumed to be part of larger population and sample)

Standard Metropolitan Statistical Area	P_{hj}	Approx. pop. of central cities (000's)	Possible samples under one controlled selection scheme and corresponding probabilities (read across)		
			Stratum I	Stratum II	Stratum III
Stratum I			.387 Norfolk-	.184 Reading	.024 Wilkes-Barre
Norfolk-Portsmouth, Va.	.387	420	Portsmouth		.160 York
Richmond, Va.	.273	220			
Allentown-Bethlehem-				.127 Scranton	.116 Harrisburg
Easton, Pa.	.329	185			.011 Atlantic City
Newport News-Hampton,Va.	.011	200			
	1.000			.076 Trenton	.076 Wilkes-Barre
Stratum II					
Newport News-Hampton, Va.	.139	200	.273 Richmond	.152 Wilmington	.132 Wilkes-Barre
Wilmington, Del.	.245	95			.020 Lancaster
Trenton, N. J.	.178	115			
Scranton, Pa.	.157	110		.102 Trenton	.006 Lancaster
Reading, Pa.	.184	100		.019 Scranton	.115 Harrisburg
Roanoke, Va.	.097	95			
	1.000		.329 Allentown-	.139 Newport News-	.139 Lancaster
			Bethlehem-	Hampton	
Stratum III			Easton		
Roanoke, Va.	.009	95		.097 Roanoke	.097 Atlantic City
Lynchburg, Va.	.074	55			
Atlantic City, N. J.	.108	60		.093 Wilmington	.074 Lynchburg
Harrisburg, Pa.	.231	80			.010 Lancaster
Lancaster, Pa.	.186	60			.009 Roanoke
Wilkes-Barre, Pa.	.232	65			
York, Pa.	.160	55			
	1.000		.011 Newport News-	.011 Scranton	.011 Lancaster
			Hampton		

Table 2.--CONTROLLED SELECTION EXTENDED

Estimators and Expected Value of Estimated Variance

A. Estimated total - based on original sample only.

$$\hat{X} = \sum_{h=1}^{L} \sum_{j=1}^{1} \frac{x_{hj(1)}}{P_{hj(1)}}$$

Subscript (1) indicates units originally selected.
L = number of strata.
P_{hj} = probability of selection for j-th unit in h-th stratum.

B. Estimated variance.

$$s_{\hat{X}}^2 = \frac{L}{2\ell} \sum_{h=1}^{\ell} (x'_{h(1)} - x'_{h(2)})^2 + \frac{L}{\ell} \sum_{h=1}^{\ell} \sum_{\substack{g=1 \\ g \neq h}}^{L-1} (x'_{h(1)} - x'_{h(2)})(x'_{g(1)})$$

where

$$x'_{h(1)} = \frac{x_{hj(1)}}{P_{hj(1)}}, \text{ etc.}$$

Subscript (2) indicates unit selected subsequently, with replacement.
ℓ = number of randomly chosen strata in which second units are selected independently.

C. Expected value of $s_{\hat{X}}^2$.

$$E(s_{\hat{X}}^2) = E\left[\frac{L}{2\ell} \sum_{h=1}^{\ell} (x'_{h(1)} - x'_{h(2)})^2 + \frac{L}{\ell} \sum_{h=1}^{\ell} \sum_{\substack{g=1 \\ g \neq h}}^{L-1} (x'_{h(1)} - x'_{h(2)})(x'_{g(1)})\right]$$

$$= \frac{L}{2\ell} E \sum_{h=1}^{\ell} \left[(x'_{h(1)} - X_h) - (x'_{h(2)} - X_h)\right]^2$$

$$+ \frac{L}{\ell} E \sum_{h=1}^{\ell} \sum_{\substack{g=1 \\ g \neq h}}^{L-1} \left[(x'_{h(1)} - X_h) - (x'_{h(2)} - X_h)\right] \left[(x'_{g(1)} - X_g) + X_g\right]$$

$$= \frac{1}{2} \sum_{h=1}^{L} E \left[(x'_{h(1)} - X_h)^2 + (x'_{h(2)} - X_h)^2\right]$$

$$+ \sum_{h=1}^{L} \sum_{\substack{g=1 \\ g \neq h}}^{L-1} E \left[(x'_{h(1)} - X_h)(x'_{g(1)} - X_g)\right]$$

$$= \sum_{h=1}^{L} E(x'_{h(1)} - X_h)^2 + \sum_{h=1}^{L} \sum_{\substack{g=1 \\ g \neq h}}^{L-1} E \left[(x_{h(1)} - X_h)(x'_{g(1)} - X_g)\right]$$

$$= E\left[\sum_{h=1}^{L} (x'_h - X_h)\right]^2 = E\left[\sum_{h=1}^{L} x'_{h(1)} - E \sum_{h=1}^{L} x'_{h(1)}\right]^2$$

$$= \sigma_{\hat{X}}^2, \text{ by definition.}$$

REFERENCES

1. Goodman, Roe and Kish, Leslie, "Controlled Selection - A Technique in Probability Sampling", Journal of the American Statistical Association, 45: 350-372 (1950).

2. Hess, I., Riedel, D. C., and Fitzpatrick, T. B., "Probability Sampling of Hospitals and Patients". Ann Arbor: The University of Michigan, Bureau of Hospital Administration, Research Series No. 1, 1961.

3. Horvitz, D. G. and Thompson, D. J., "A Generalization of Sampling without Replacement from a Finite Universe", Journal of the American Statistical Association, 47: 663-685 (1952).

4. Wilkerson, Marvin, "The Revised City Sample for the Consumer Price Index", U.S. Department of Labor, Bureau of Labor Statistics, Monthly Labor Review, Oct., 1960.

5. U.S. Department of Commerce, Bureau of the Census, "The Current Population Survey A Report on Methodology" Technical Paper No. 7.

A COMPARISON OF THREE RULES FOR ASSOCIATING FARMS AND FARMLAND WITH SAMPLE AREA SEGMENTS IN AGRICULTURAL SURVEYS.*

W. A. Hendricks, D. T. Searls and D. G. Horvitz
Research Triangle Institute

1. Introduction

The application of area sampling techniques to sample farm surveys requires some rule for associating farms and farmland with the selected area segments. The rule that was adopted in the United States about a quarter of a century ago, when area sampling first began to be applied to farm surveys in this country, and which has been used most generally up to the present time, is the so-called "headquarters rule." By this rule a point on every farm which can be defined rigorously, and which can be identified by interviewers in the field, is employed as a reference point. If this point for a particular farm falls within the boundaries of the area segment, the farm is regarded as being "in" the segment. But if the reference point is outside the segment, the farm is considered outside the segment, even though some or most of all land in the farm may fall inside the segment.

The Census Bureau and other agencies involved in farm surveys have developed elaborate definitions of the reference point which is to serve as the "headquarters" of the farm in any particular instance, such as the following:

a. If the operator of the farm lives on the farm, his house is the headquarters.
b. If the operator does not live on the farm but there is one and only one house on the farm, that house is the headquarters.
c. If there is more than one house on the farm and the operator does not live on the farm, the house of greatest value is the headquarters.
d. If there are no houses on the farm but other buildings are present, the building of greatest value is the headquarters.
e. If there are no buildings on the farm, the "main entrance" to the farm is the headquarters.
f. If no point can be identified as the main entrance, the corner of the farm farthest west and farthest north (in that order) is the headquarters.

Such sequences are established to permit the use of reference points that can be identified by interviewers with the least difficulty and possibility of error. However, errors in associating farms with area segments by a "headquarters" rule are usually all to prevalent. The problem is aggravated by difficulties and errors that often arise in deciding which parcels of land constitute the "farm."

To avoid these troublesome problems, the agency of the United States Department of Agriculture now known as the Statistical Reporting Service has in recent years made extensive use of a so-called "Closed Segment" rule. By this rule interviewers must account only for items on land that lies entirely within the segment boundaries. To obtain estimates of numbers of farms, all persons, or a sample of them, living within the segment boundaries must be screened to determine how many are farm operators. This is accomplished by ascertaining the nature and extent of their agricultural operations, if any, regardless of where they are performed. These persons are also interviewed to obtain data that can only be obtained for the farm as a whole, such as sales and purchases of various commodities. Wherever possible, estimates of such items are made to conform to related information obtained from land within the segment boundaries. For example, universe estimates of cattle inventories are obtained from the numbers found within the sample segments on the date of the enumeration. But estimates of sales are obtained by applying the ratio of farm sales to farm inventories, for entire farms as reported by farm operators in the segments, to the estimate of total inventories derived only from numbers within the segment boundaries.

This "closed segment" rule has several advantages over a "headquarters" rule, but it also has some disadvantages. On the plus side we note (1) the rule is less troublesome for interviewers to apply in the field, (2) as the land to be accounted for in each sample segment is depicted on the interviewers' aerial photos, both interviewers and supervisors can recognize gross errors in reported data that might otherwise pass undetected, (3) reporting errors can be reduced because respondents are asked to report separately for specific tracts of land pointed out to them on the aerial photos, and in many cases where their holdings inside the segment boundaries represent only a portion of all their holdings, they are not required to disclose information about the portions outside the segment, and finally (4) between-segment sampling variation can be reduced because the boundaries of each segment place a limit on the total land to be accounted for in connection with that segment.

On the minus side we observe that farm operators who live in cities and towns can only be counted if some sample segments are allocated to urban areas. The task of identifying farm operators in such areas is often formidable and considerable undercounting may result. Furthermore, there is also the risk that some urban residents classified as farm operators may have tenants on their farms who would be classified

* This study was conducted for the Bureau of the Census, United States Department of Commerce, under Contract No. CcO-7574.

as the operators if they fell into the open-country portion of the sample. This is perhaps the greatest defect in the closed segment procedure. As pointed out previously, the fact that the method is not well adapted to surveys seeking data which can be reported conveniently only for farms as a whole must also be considered on the debit side.

Another approach, which for want of a better name has been called the "weighted segment" procedure, seems to offer a solution that retains many of the advantages of both the "headquarters" and "closed segment" rules and is also free of the most serious objections levelled at those two methods. So far as the authors of this paper are aware, it has not been applied to farm surveys in the United States as yet, although an agency of the Department of Agriculture has sponsored a rural land-ownership survey in which this method was employed. The Statistical Reporting Service made use of a similar principle in a survey for estimating the availability of farm grain storage facilities a few years ago.

As applied to farm surveys, the "weighted segment" approach regards every farm with some of its land inside a sample segment as associated with that segment. Data are recorded for every such farm as a whole, but are multiplied by the fraction of the farmland in the farm that falls within the segment before incorporating them into segment totals.

This procedure has a number of advantages over the two previous rules. First, it is a rule that can be applied by interviewers in the field with less difficulty and possibility of error than a headquarters rule. Also the need for canvassing urban areas is eliminated and the sample can be confined to the open country. All data are recorded for entire farms; hence no special treatment is required for items that can not be reported conveniently for portions of farms as in the closed segment approach. At the same time the weighting of the data for each farm by the fraction of its land falling within the segment can reduce the between-segment variability of segment totals to a level comparable to that obtained with closed segment data. In fact, this variability can be expected to be lower because of the averaging effect achieved by prorating entire-farm data to land within the segment, rather than recording data only for the land within the segment.

Among the disadvantages, as compared with the closed segment procedure, we observe that interviewers are still faced with the problem of deciding which parcels of land must be defined as a "farm" and this is sometimes difficult. However, errors introduced by including too much land, or too little, in defining a farm tend to be partially neutralized by the weighting procedure. The fact that all land reported as being in a farm may not be covered by the interviewers' aerial photo eliminates some of the visual verification that can be performed by supervisors on closed segment data. The weighting that must be applied to individual farm data is a computational step that is not required with either the headquarters or closed segment rules, but with automatic data processing procedures that are now in rather general use this is not a serious matter. With sample segments of a given size, data must be recorded for about twice as many farms by this rule as compared with a headquarters rule. With long questionnaires this can increase the time required to be spent in each sample segment, although it would by no means double the time required with a headquarters rule. The proper application of a headquarters rule requires a complete canvass of each sample segment to ascertain the places eligible for enumeration; considerable time is often consumed in screening out ineligible places.

2. Objectives of the Present Study

The purpose of the present study was to investigate the sampling variances encountered with each of the preceding rules for associating farms with sample area segments. Although some fragmentary information on the headquarters and closed segment rules has accumulated over the past few years, no systematic comparison of the two approaches in this respect has ever been made. So far as the weighted segment rule is concerned, objective data on variances are nonexistent.

The data employed in this study were obtained in the 1954 and 1959 Census of Agriculture Evaluation Programs. Data from 384 identical segments in 175 primary sampling units (PSU) were available for both of those years in a form that made such an analysis possible. In addition to detailed information about the characteristics of each farm covered in the Evaluation Program, data were recorded separately for the portions of those farms that fell inside the segment boundaries. To investigate the behavior of the headquarters rule, only two alternative reference points on each farm were considered as headquarters: (1) if the operator lived on the farm, his residence was the headquarters, and (2) if the operator did not live on the farm, the point on the farm farthest west and farthest north (in that order) was the headquarters. Sketches of the farm in relation to the segment boundaries were available to make the appropriate determination in each case.

All relevant information was placed on punch cards to facilitate the computations, which were performed on an IBM 1401 computer. The items studied in the analysis were:

1. Numbers of farms (omitted in closed segment analysis)
2. Acres of farmland
3. Acres of cropland harvested
4. Acres of corn harvested
5. Acres of wheat harvested
6. Acres of cotton harvested
7. Acres of soybeans harvested
8. Acres of oats harvested
9. Acres of hay cut
10. Acres of tobacco harvested

Between segment variability within primary units was estimated for the above items separately for each of three regional strata of the United States and for each of the three rules employed to associate sample farms with sample area segments. Coefficients of correlation between 1954 and 1959 data were also computed for each rule of association. The variances and covariances within primary units were computed under the assumption that a large sample of farms in the United States would be a one-stage design with sample segments allocated proportionally to all PSU's in the universe rather than a two-stage design of the kind used in the Evaluation Program. To indicate how much improvement could be attained by basing 1959 estimates of agricultural items upon 1954 data through the use of difference, ratio, and regression estimators, the variability of such estimates was computed on a per-segment basis for comparison with the per-segment variances of the 1959 segment totals. To investigate possible gains in precision by excluding large farms from the area sample, all variances and covariances were computed with data for farms of 2,000 acres or more included and excluded.

The three regional strata and the number of sample segments in each are as follows:

Region I-North (154 segs.)	Region II-South (116 segs.)	Region III-West (114 segs.)
Connecticut	Alabama	Arizona
Illinois	Arkansas	California
Indiana	Delaware	Colorado
Iowa	Florida	Idaho
Maine	Georgia	Kansas
Massachusetts	Kentucky	Montana
Michigan	Louisiana	Nebraska
Minnesota	Maryland	Nevada
Missouri	Mississippi	New Mexico
New Hampshire	N. Carolina	N. Dakota
New Jersey	S. Carolina	Oklahoma
New York	Tennessee	Oregon
Ohio	Virginia	S. Dakota
Pennsylvania		Texas
Rhode Island		Utah
Vermont		Washington
West Virginia		Wyoming
Wisconsin		

These regions do not conform exactly to the regions for which official agricultural statistics are customarily summarized. The regional strata employed here were adopted mainly to achieve greater equality in the number of sample segments per stratum.

3. Computational Methods

The 175 primary units in the three regions had been selected in 1954 with unequal probabilities and the sampling rates within those PSU's had been adjusted to arrive at a self weighted sample. For the present analysis some PSU's containing only one sample segment each needed to be combined with other PSU's to permit the computation of within-PSU variability. After those combinations were made, the 384 segments were contained in 124 new PSU's which were regarded as being selected with probabilities comparable to the original PSU. The within-PSU sampling rates were regarded as proportional to the reciprocals of those probabilities. The average 1959 variance between segments within PSU's for a given region was therefore computed from the formula

$$s_y^2 = \frac{1}{n} \sum_i \frac{n_i}{n_i - 1} \sum_j (y_{ij} - \bar{y}_i)^2 \qquad (1)$$

in which

- y_{ij} = a 1959 segment total for the j-th segment in the i-th PSU.
- $\bar{y}_i$ = the per-segment average of the y_{ij} for all segments in the i-th PSU.
- n_i = the number of sample segments in the i-th PSU.
- n = the total number of sample segments in the region.

Average covariances between 1954 and 1959 data were computed in a similar fashion:

$$s_{xy} = \frac{1}{n} \sum_i \frac{n_i}{n_i - 1} \sum_j (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i) \qquad (2)$$

in which x_{ij} and y_{ij} are comparable segment totals in 1954 and 1959.

Clearly, if an estimated universe total of an agricultural item is to be made for a current year, such as 1959, without reference to related data for previous years, the relvariance (RV) of that estimate will be equal to the relvariance of the per-segment average for that item in the sample. If such an estimate is represented by Y_1, we have

$$RV(Y_1) = \frac{s_y^2}{n\bar{y}^2} \qquad (3)$$

in which s_y^2 is the between segment variance of the sample segment totals, $\bar{y}$ is the per-segment average of all segment totals, and n is the number of segments in the sample.

If a difference estimate, which makes use of data for the universe and the sample in a previous year such as 1954 is computed, such an estimate takes the form

$$Y_2 = X + (Y_1 - X_1) \qquad (4)$$

in which X is the universe total in the base year, Y_1 is the current year estimate obtained by applying the reciprocal of the sampling rate to the current year sample total, and X_1 is the corresponding estimate of X derived from base year data in the sample. The relvariance of Y_2 is given by

$$RV(Y_2) = \frac{s_x^2 + s_y^2 - 2s_{xy}}{n\bar{y}^2} \qquad (5)$$

in which the various terms are self-explanatory.

Instead of employing a difference estimator, one might invoke a ratio estimate of the form

$$Y_3 = \frac{Y_1}{X_1} X \qquad (6)$$

The relvariance of Y_3, written in terms of the relvariances of X_1 and Y_1, together with the relative covariance (RCV) of X_1 and Y_1 is approximately:

$$RV(Y_3) = RV(X_1) + RV(Y_1) - 2RCV(X_1Y_1) \qquad (7)$$

A third alternative would be to employ a regression estimator of the form

$$Y_4 = Y_1 + b(X - X_1) \qquad (8)$$

in which b is the estimated regression coefficient of y on x and the other symbols have the same meaning as before. The relvariance of Y_4 is approximately

$$RV(Y_4) = (1 - r_{xy}^2)\ RV(Y_1) \qquad (9)$$

in which r_{xy} is the average within PSU coefficient of correlation between 1954 and 1959 totals for the same segments.

4. Numerical Results •

Simple Expansion Estimates

The relvariance of a simple expansion estimate, as indicated previously, is identical with the relvariance of the per-segment average of that item. The estimated relvariances for all farms for each of the selected items under consideration are shown in Table 1 on a per-segment basis (n = 1) by region for each of the three association rules, and also with farms of 2,000 acres or more excluded. No such large farms were in the Region I sample.

The table indicates that the exclusion of large farms from the sample (Regions II and III) did not reduce the relative variances to any great degree. Aside from farmland itself, some reduction was effected in the relative variances of a few items such as corn, oats, and hay acreages in Region II.

The Closed Segment and Weighted Segment Rules tend to yield the lowest variances, with the Weighted Segment showing a slight edge over the Closed Segment. Table 2 shows that the use of the Weighted Segment Rule about doubles the number of farms from which information is obtained as compared with the Headquarters approach. This does not double the interview cost because interviewers must account for all land in a segment to identify farms with headquarters in the segment.

Table 2. Farms with Land in Segment by Location of Headquarters, 1959 EPA Rural Area Sample

Region[1/]	Number of farms in sample		
	With Hq. in Segment	With Hq. Not in Segment	Total
North	1,192	1,251	2,443
South	1,072	1,033	2,105
West	291	240	531
U. S.	2,555	2,524	5,079

1/ The regions are as defined for the 1959 Census, not as for this study.

1954-59 Correlation Coefficients

Average correlation coefficients between segments within PSU are shown in Table 3 for the items covered in Table 1. Correlations are highest in Region I and lowest in Region III. In Region II they are a bit larger when large farms are retained in the sample, but in Region III the reverse seems to be true in several instances. One extremely large place of about 128,000 acres in Region III was omitted from the computations because it tended to dominate the results unduly. Generally speaking, correlations are highest with the Closed and Weighted Segment Rules of association. Of these two, the difference is again in favor of the Weighted Segment Rule.

The correlations are of sufficient magnitude, almost in general, to yield a considerable increase in statistical efficiency with estimation procedures which include prior Census or base year information as compared to simple expansion estimation. This is verified in Tables 4, 5 and 6. The average correlations between segments within PSU's for the years immediately following a Census year should be even larger than those obtained for the 5-year interval in this study. It should be pointed out, however, that base year data assembled for the sample segments during a Census should be in a form comparable to the data that will be collected according to the specified association rule in intercensal sample surveys.

The negative correlations for cotton acreage in Region II with the Headquarters Rule were due to two segments with fairly large acreages in 1954 but small cotton acreages, according to this rule in 1959. This was due to a shift of the headquarters of one or two farms out of these segments in 1959, rather than a larger reduction in cotton acreage between 1954 and 1959. The correlations for cotton acreage in this region for the Closed and Weighted Segment Rules remained substantially postive.

Difference Estimates

The relative variances of 1959 difference estimates, based on 1954 data from the same seg-

Table 1. Estimated Average Within Primary Sampling Unit Relative Variances (n=1) for <u>Simple Expansion Estimates</u> obtained with Three Rules of Associating Farm Land with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
			Region I			
Farms	0.388	*	0.381	0.388	*	0.381
Farmland	.544	.420	.420	.544	.420	.420
Cropland Harvested	.720	.561	.567	.720	.561	.567
Corn Acreage	.913	.858	.746	.913	.858	.746
Wheat Acreage	2.150	2.625	1.998	2.150	2.625	1.998
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	2.418	1.630	1.343	2.418	1.630	1.343
Oats Acreage	.974	1.038	1.123	.974	1.038	1.123
Hay cut Acreage	.870	.844	.781	.870	.844	.781
Tobacco Acreage	**	**	**	**	**	**
			Region II			
Farms	.598	*	.713	.596	*	.710
Farmland	1.775	1.194	1.194	1.335	.992	.992
Cropland Harvested	3.914	3.174	2.585	4.019	3.185	2.654
Corn Acreage	3.241	2.776	1.950	2.092	1.749	1.418
Wheat Acreage	11.142	9.581	8.238	11.046	9.495	8.148
Cotton Acreage	6.188	2.228	2.017	6.135	2.218	2.003
Soybean Acreage	20.219	18.021	17.975	20.043	17.863	17.818
Oats Acreage	12.685	13.320	9.105	8.194	5.794	5.323
Hay cut Acreage	7.296	6.585	5.405	6.567	5.620	5.304
Tobacco Acreage	7.837	5.902	6.804	7.769	5.850	6.745
			Region III			
Farms	.553	*	.440	.546	*	.436
Farmland	5.612	.878	.878	.721	.688	.688
Cropland Harvested	1.086	.604	.577	1.098	.555	.563
Corn Acreage	1.142	.750	.643	1.092	.740	.641
Wheat Acreage	1.326	1.186	1.208	1.347	1.134	1.129
Cotton Acreage	12.289	15.264	11.742	12.182	14.455	12.039
Soybean Acreage	29.449	31.518	26.694	29.196	31.246	26.450
Oats Acreage	2.536	2.026	2.130	2.473	2.018	2.136
Hay cut Acreage	1.133	.961	.619	1.042	.944	.639
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

Table 3. Estimated Average Within Primary Sampling Unit Correlations between 1954 and 1959 Area Segment Totals obtained with Three Rules of Associating Farmland with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
			Region I			
Farms	0.716	*	0.798	0.716	*	0.798
Farmland	.744	.942	.942	.744	.942	.942
Cropland Harvested	.828	.960	.966	.828	.960	.966
Corn Acreage	.810	.839	.922	.810	.839	.922
Wheat Acreage	.623	.857	.905	.623	.857	.905
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	.686	.705	.793	.686	.705	.793
Oats Acreage	.750	.722	.895	.750	.722	.895
Hay cut Acreage	.732	.751	.752	.732	.751	.752
Tobacco Acreage	**	**	**	**	**	**
			Region II			
Farms	.799	*	.686	.799	*	.686
Farmland	.887	.932	.932	.832	.918	.918
Cropland Harvested	.851	.947	.936	.828	.930	.928
Corn Acreage	.828	.851	.834	.657	.704	.758
Wheat Acreage	.608	.731	.656	.610	.730	.657
Cotton Acreage	-.070	.685	.734	-.070	.405	.594
Soybean Acreage	.574	.820	.702	.574	.824	.699
Oats Acreage	.697	.769	.728	.182	.374	.360
Hay cut Acreage	.660	.671	.690	.360	.379	.538
Tobacco Acreage	.962	.957	.960	.962	.957	.960
			Region III			
Farms	.691	*	.714	.704	*	.718
Farmland	.855	.319	.319	.429	.378	.378
Cropland Harvested	.597	.817	.730	.684	.772	.740
Corn Acreage	.430	.559	.661	.491	.582	.663
Wheat Acreage	.446	.892	.682	.537	.827	.665
Cotton Acreage	.663	.743	.741	.663	.652	.712
Soybean Acreage	.250	.578	.344	.250	.578	.344
Oats Acreage	-.013	.572	.576	.571	.654	.738
Hay cut Acreage	.388	.388	.479	.500	.367	.453
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

ments with the same rules of association, are shown in Table 4. As compared with the variances of the simple expansions shown in Table 1, the results are as would be expected from the magnitudes of the correlation coefficients involved. Most items show some improvement in Regions I and II. Improvement was least noticeable in Region III.

The superiority of the Closed and Weighted Segment approaches is quite noticeable and as expected from the higher correlations. In a number of instances, the indicated gain in statistical efficiency would more than offset the increased cost, if any, of the Weighted Segment approach.

Ratio Estimates

The relative variances of 1959 ratio estimates, comparable to the difference estimates of the preceding section, are given in Table 5. As anticipated there are no striking differences between the results in Tables 4 and 5. In some cases the difference estimates are better but in others the ratio estimates have less variability. There seems to be a small edge in favor of ratio estimates.

Regression Estimates

Regression estimates should have less variability than difference or ratio estimates because sampling fluctuations in base data have only a negligible effect, whereas in difference and ratio estimates such fluctuations exert considerable effect. For the present computations the effects of sampling fluctuations in 1954 base data on the regression estimates were ignored completely. The results are shown in Table 6.

As expected, the relative variances are generally smaller than for the other kinds of estimates. The Closed and Weighted Segment Rules of association again show considerable superiority, with the latter being a bit better. Eliminating the large farms from the sample did not change the relative variances appreciably except in a few instances. This was also true for the difference and ratio estimates. The items affected the most by removing the large farms were oats acreage in Regions II and III and farmland in Region III.

5. Summary and Conclusions

On the basis of the variances observed in this study, the Closed and Weighted Segment Rules of association are decidely preferable to the Headquarters Rule. This holds true for each of the four methods of estimation considered. Relative variances generally are a bit lower for the Weighted Segment Rule than for the Closed Segment Rule.

The elimination of farms of 2,000 acres or more from the sample reduced the relative variances somewhat, but the reduction was not particularly striking. However, the number of large farms in the sample was small. When large farms are eliminated from an area sample and treated separately, the relative variance of the estimate for the two strata combined may be reduced appreciably.

When current estimates are computed by difference, ratio or regression procedures with matching data from a previous Census year, an appreciable reduction can be achieved in the relative variances of the agricultural items studied. The reductions in Region III (where the correlations between the 1954 and 1959 data were lower) would probably be less than in Regions I and II. Ratio and difference estimates would tend to have similar precision, with perhaps a slight edge for the ratio estimates. As expected, regression estimates would be the most precise.

The gain in precision to be achieved with the Closed and Weighted Segment Rules as compared to the Headquarters approach is even greater with estimation procedures which make use of base year data, since the correlations are generally higher for these approaches.

Obviously, other considerations in addition to the size of the sampling error must be taken into account when recommending one procedure over another. However, when the farm as a whole is regarded as the unit of observation, the Weighted Segment Rule appears to be preferable to other rules that might be considered for associating farms with sample segments. Not only are sampling errors smaller with this rule; the rule is less likely to be misinterpreted or misapplied by interviewers in the field.

For items where the farm as a whole does not necessarily have to serve as the unit of observation, the Closed Segment Rule has much to commend it. Perhaps the most serious objection to this rule is that some supplemental procedure must be used in conjunction with it to arrive at a count of number of farms and to obtain data on items that apply to the farm as a whole. The rule does have the advantage that interviewers have less difficulty determining the land to be covered in the enumeration. If all of the tracts enumerated are delineated on aerial photos, interviewers and supervisors can perform approximate visual verifications on much of the data reported by respondents. Both of these properties should have the effect of reducing nonsampling errors.

Table 4. Estimated Average Within Primary Sampling Unit Relative Variances (n=1) for Difference Estimates obtained with Three Rules of Associating Farm Land with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
			Region I			
Farms	0.286	*	0.178	0.286	*	0.178
Farmland	.350	.049	.049	.350	.049	.049
Cropland Harvested	.274	.045	.039	.274	.045	.039
Corn Acreage	.335	.257	.112	.335	.257	.112
Wheat Acreage	2.564	1.214	.584	2.564	1.214	.584
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	1.775	.909	.640	1.775	.909	.640
Oats Acreage	.663	.713	.244	.663	.713	.244
Hay cut Acreage	.445	.446	.367	.445	.446	.367
Tobacco Acreage	**	**	**	**	**	**
			Region II			
Farms	.589	*	.481	.584	*	.476
Farmland	.453	.222	.222	.540	.272	.272
Cropland Harvested	1.226	.328	.320	1.420	.437	.374
Corn Acreage	1.213	.857	.734	1.614	1.088	.848
Wheat Acreage	7.030	4.473	4.854	6.945	4.446	4.765
Cotton Acreage	14.314	4.494	2.195	12.353	5.975	2.703
Soybean Acreage	16.882	11.100	12.383	16.736	11.094	12.393
Oats Acreage	15.347	13.032	12.215	29.185	30.813	21.817
Hay cut Acreage	4.197	3.620	2.957	6.690	5.454	4.280
Tobacco Acreage	1.742	1.816	1.322	1.727	1.800	1.310
			Region III			
Farms	.391	*	.309	.374	*	.296
Farmland	1.514	2.551	2.551	.974	.716	.716
Cropland Harvested	1.309	.262	.367	1.051	.305	.322
Corn Acreage	1.120	.574	.419	1.010	.548	.423
Wheat Acreage	4.029	.486	.780	2.546	.591	.754
Cotton Acreage	48.395	13.032	14.932	47.972	21.355	17.584
Soybean Acreage	29.054	21.442	24.622	28.804	21.258	24.397
Oats Acreage	8.296	2.174	1.799	3.093	2.045	1.189
Hay cut Acreage	2.114	1.544	.924	1.587	1.573	.974
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

Table 5. Estimated Average Within Primary Sampling Unit Relative Variances (n=1) for Ratio Estimates obtained with Three Rules of Associating Farm Land with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
			Region I			
Farms	0.209	*	0.140	0.209	*	0.140
Farmland	.309	.048	.048	.309	.048	.048
Cropland Harvested	.278	.045	.038	.278	.045	.038
Corn Acreage	.374	.282	.123	.374	.282	.123
Wheat Acreage	2.198	1.320	.540	2.198	1.320	.540
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	2.035	1.206	.825	2.035	1.206	.825
Oats Acreage	.507	.533	.236	.507	.533	.236
Hay cut Acreage	.454	.393	.348	.454	.393	.348
Tobacco Acreage	**	**	**	**	**	**
			Region II			
Farms	.257	*	.378	.255	*	.376
Farmland	.394	.162	.162	.458	.176	.176
Cropland Harvested	1.118	.400	.381	1.301	.524	.446
Corn Acreage	1.022	.767	.600	1.230	.922	.639
Wheat Acreage	7.573	4.953	5.463	7.442	4.899	5.326
Cotton Acreage	9.817	1.514	1.028	9.732	2.828	1.538
Soybean Acreage	13.564	6.915	9.618	13.447	6.603	9.486
Oats Acreage	8.388	6.379	4.987	16.951	11.466	8.512
Hay cut Acreage	4.168	3.647	2.859	6.295	5.263	3.945
Tobacco Acreage	.602	.517	.540	.596	.512	.535
			Region III			
Farms	.302	*	.234	.287	*	.227
Farmland	1.742	2.052	2.052	.787	.705	.705
Cropland Harvested	.924	.226	.320	.756	.259	.290
Corn Acreage	1.041	.556	.392	.931	.530	.397
Wheat Acreage	2.448	.314	.708	1.722	.421	.684
Cotton Acreage	24.565	7.005	6.383	8.830	9.022	6.531
Soybean Acreage	58.914	42.573	48.035	58.388	42.174	47.613
Oats Acreage	4.615	1.523	1.470	2.017	1.361	.979
Hay cut Acreage	1.324	1.112	.742	1.014	1.120	.786
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

Table 6. Estimated Average Within Primary Sampling Unit Relative Variances (n=1) for Regression Estimates obtained with Three Rules of Associating Farm Land with Sample Area Segments, by Region, with and without Farms of 2000 or More Acres.

Item	All Farms Association Rule			Excluding Large Farms Association Rule		
	Head-quarters	Closed Segment	Weighted Segment	Head-quarters	Closed Segment	Weighted Segment
				Region I		
Farms	0.189	*	0.138	0.189	*	0.138
Farmland	.243	.048	.048	.243	.048	.048
Cropland Harvested	.226	.044	.038	.226	.044	.038
Corn Acreage	.314	.254	.112	.314	.254	.112
Wheat Acreage	1.315	.690	.361	1.315	.690	.361
Cotton Acreage	**	**	**	**	**	**
Soybean Acreage	1.281	.819	.499	1.281	.819	.499
Oats Acreage	.426	.497	.224	.426	.497	.224
Hay cut Acreage	.403	.368	.339	.403	.368	.339
Tobacco Acreage	**	**	**	**	**	**
				Region II		
Farms	.216	**	.378	.215	*	.376
Farmland	.378	.157	.157	.410	.157	.157
Cropland Harvested	1.081	.326	.318	1.264	.431	.367
Corn Acreage	1.020	.483	.594	1.189	.832	.603
Wheat Acreage	7.025	4.465	4.689	6.936	4.440	4.626
Cotton Acreage	6.158	1.181	.932	6.105	1.854	1.296
Soybean Acreage	13.563	5.918	9.129	13.446	5.743	9.104
Oats Acreage	6.515	5.441	4.286	7.922	4.986	4.632
Hay cut Acreage	4.121	3.619	2.831	5.718	4.814	3.772
Tobacco Acreage	.592	.500	.532	.586	.495	.528
				Region III		
Farms	.289	*	.215	.275	*	.212
Farmland	1.511	.789	.789	.589	.590	.590
Cropland Harvested	.699	.201	.270	.585	.224	.255
Corn Acreage	.931	.516	.362	.829	.490	.360
Wheat Acreage	1.062	.242	.646	.959	.358	.630
Cotton Acreage	6.891	6.832	5.300	6.831	8.303	5.933
Soybean Acreage	27.614	20.973	23.534	27.377	20.791	23.319
Oats Acreage	2.535	1.363	1.424	1.666	1.155	.971
Hay cut Acreage	.963	.817	.477	.781	.817	.508
Tobacco Acreage	**	**	**	**	**	**

*Data on number of farms not available.

**Not computed since average acreage per segment was zero or close to zero.

VIII

POSTCENSAL STUDIES OF SCIENTIFIC AND SPECIALIZED PERSONNEL

Chairman, Robert W. Cain, National Science Foundation

Page

THE POSTCENSAL STUDIES PROGRAM: BACKGROUND AND CONTENT

Norman Seltzer, National Science Foundation

The need for data on the Nation's resources of all types of manpower has become more urgent as an awareness grows that such information is vital in planning and evaluating many economic, educational, and scientific policies and programs in all sectors of the economy. In particular, our concern regarding our resources of high-level manpower required to carry out the multitude of tasks in science and technology has been reflected in the National Science Foundation's programs of scientific manpower studies. These programs have been in effect since the establishment of the National Science Foundation more than a decade ago, and although the number and variety of studies and surveys supported by the Foundation in recent years has been quite extensive, the realization that we have yet much to accomplish is with us at all times.

The genesis of the Postcensal Studies Program goes back to 1957 when the Foundation together with the President's Committee on Scientists and Engineers appointed a special advisory panel to review requirements for scientific manpower data. In its report[1/] issued in 1958, this panel found, not surprisingly, that on the whole, data on the number, demand, supply, utilization, and other economic and social characteristics of scientific and technical personnel were not adequate for formulating policies and undertaking programs related to the welfare and security of the Nation. Among the projects recommended as highly urgent was "a special survey of a large sample of persons recorded in the 1960 Census enumeration as college graduates or as persons currently or last employed in scientific and technical positions, whether college graduates or not, to determine relationships between training and subsequent occupations." It was reasoned that the 1960 Census would provide a rare opportunity (not available again for possibly another 10 years) to obtain valuable data directly and efficiently from the individuals classified in scientific and technical occupations regarding their demographic, economic, and social characteristics. In addition, by surveying all types of college graduates, regardless of occupation, comparable data would be provided on other highly trained personnel in other professions (such as law and medicine), in the humanities, in business, and in managerial and administrative positions in business and government as well as those currently not in the labor force.

With the knowledge that Census Bureau data processing would have available by 1962 a tape with the occupational sample information, the National Science Foundation in 1960 began to consider the feasibility of undertaking such a major project, and requested that the National Opinion Research Center, affiliated with the University of Chicago, prepare a planning statement on a series of postcensal studies of scientific and professional workers and college graduates. A detailed planning statement was prepared which provided the initial basis for the proposed studies. During the first part of 1961, the Foundation, aware of the interests and missions of other Federal agencies, initiated a series of meetings with these agencies to acquaint them with the proposed studies and provide them with the opportunity to participate in or co-sponsor the nation-wide survey under consideration. By late fall, 1961, four other Federal agencies--the U. S. Office of Education, National Institues of Health, Bureau of Labor Statistics, and the Veterans Administration--had made definite commitments to participate in the survey and provide the necessary support.

Beginning in the Spring of 1961, the National Opinion Research Center with the aid and consultation of the National Science Foundation and the Bureau of the Census began to develop a basic mail questionnaire which would be used as the primary survey instrument for the Postcensal Studies Program. Since another paper has developed in detail the mechanics and operation of the survey itself, the remainder of this paper will discuss the content of the questionnaire and some of the possible uses of the data which we were seeking.

The availability of a large sample based on the entire population was very attractive bait for requesting information on an extremely wide variety of subjects which could well have covered numerous economic, demographic, sociological, and psychological areas. The temptation to run wild, so to speak, was held in check by the very obvious realizations that the burden imposed on the respondents might be such as to obviate any possible success in obtaining a meaningful rate of response. In addition, of course, was the fact that the Foundation's program of manpower studies enables us to obtain data from a variety of sources, much of which would both complement and supplement data derived through the Postcensal Studies Program. These manpower studies include among others, the National Register of Scientific and Technical Personnel, employment surveys of scientific and technical personnel in various economic sectors, follow-up studies of college graduates, and pilot efforts dealing with the labor market behavior and mobility of persons in selected occupations.

The content of the Postcensal Studies Program can be reviewed in some detail through an examination of the survey questionnaire itself. Even with self-imposed limitations, the final version of the basic questionnaire ran to eight "fully-packed" pages with an additional sheet enclosed containing a list of pre-coded fields of specialization to be used in answering questions dealing with employment and training. For

1/ A Program for National Information on Scientific and Technical Personnel, NSF 58-28, National Science Foundation, 1958.

a small sample of persons in occupations in which the National Institutes of Health had particular interest, another one page supplemental questionnaire was added dealing in detail with questions of financial support received by the respondent for both training and research purposes.

To begin with, among the main foci of interest were questions pertaining to the employment, occupation, and job activities of persons classified in selected professional, scientific, and technical occupations during the 1960 Census. In the 1960 Census schedule, the amount of information available pertaining to a person's work activities is quite limited. In fact, the only direct questions deal with what a person does (in terms of an occupational classification) and what type of employer he works for. Even information collected by both Government and non-government organizations for studies dealing with job analysis or vocational guidance have provided largely some general outlines when dealing with professional and technical personnel. The occupational label used in classifying personnel such as "engineer", "chemist", or "college professor" actually covers persons in a wide range of specializations.

Therefore, we were interested in determining first, for the most current period possible, how many engineers, for example, were working in civil compared to nuclear engineering?; how many chemists considered themselves to be in organic chemistry compared to physical chemistry?; and, what fields college teachers considered their primary area of specialization? Beyond this, we were hopeful of obtaining some insight into the extent to which interdisciplinary work in science and technology has resulted in engineers working in an area of the physical or life sciences, physicists concerning themselves primarily with some aspect of the medical sciences, or mathematicians calling astronomy their field of work specialization.

Another equally important area of job information is the activities or duties that are actually performed; that is, what do people classified in professional and technical occupations of interest to us "really do" in their jobs. Most of our information in this area, up to the present time, comes from other surveys which give us only an indication of the functions in which an individual is primarily engaged; for example, the National Register and the employment surveys mentioned earlier. Although we may have some indication that a certain number of physicists may be involved in "research", what the varied job requirements or duties of these personnel are, has not really been known. For some, this may mean that aside from engaging in applied research, the job may entail consulting customers on technical matters, coordinating a team of other professional personnel, and writing technical reports; for others, there may be administrative duties, and making estimates of markets for new products. For persons in other occupations, such as engineers, mathematicians, economists, and college teachers, there are of course a similar wide range of activities which make up the different types of jobs in which such personnel are engaged. In addition to obtaining an overview of the varied activities making up the jobs of professional and technical workers, respondents in the study were also requested to indicate which two activities were primary in the sense of most time being spent on them.

The organization of work in professional and technical occupations and the interpersonal relationships in the work environment is another area in which little information has been developed. Although in the past, the professional, in particular, was either self-employed or worked largely alone even when employed in an organization, the development of the professions and the complexities of scientific and technical work in an increasingly industrialized setting have resulted in considerable changes in the organizational environment. In order to obtain some understanding of this environment, a series of questions were directed at the respondent regarding the size of organizational unit in which employed; the number of employees being supervised, if any; whether he works as part of a team, either with personnel from his own field of specialization or from others; whether he has an immediate supervisor, and if so, if the supervisor's field of specialization is similar to his.

If our knowledge concerning the current employment and job activities of professional and technical personnel has been rather limited, this has been even more so about the process over time by which such highly trained persons are allocated to various jobs and employers, the career paths which may characterize different professions, and the movement of professional and technical personnel between various employers, occupations, and work specializations. It was determined that some insight into this complex area would be helpful in dealing with an assortment of problems including the supply and demand of scientific and technical personnel. To this end, questions on employment and job activities were related to three points in time--current employment (e.g., mid-1962 when the survey schedules were sent out), April 1960 (the date of the decennial census when the persons in these occupations were originally enumerated), and the first full-time job held at age 24. It was obvious, of course, that such information could not provide complete work histories, but it would give us a broad overview of mobility patterns. For these time periods, it will be possible to analyze many factors in relation to changers and nonchangers among the various occupational groups; for example, there are those who have always had the same occupation with the same employer; those who have changed employers one or more times but remained in the same occupations; those who have remained with the same type of employer and occupations but whose area of work specialization has shifted during their careers, etc.

What insights can be provided by such data? The period 1960-1962 has been marked by, among other things, an increase in vast Federal Government expenditures for research and development, a build-up of activities in both government and industry for the space program, an increase in existing as well as new programs for medical and health research, an increased emphasis on the development of new products in many science-oriented industries, and an expansion of college and university facilities to accomodate the influx of new students and provide for expanding research programs. Against this background, the recent mobility data will provide an evaluation of the movement between employers, jobs, activities, and fields of specializations. For example, are more scientists moving from academic employers to industrial jobs than vice versa? Are a greater proportion of engineers concerned with administrative or supervisory duties than heretofore? Are certain industries attracting a higher proportion of the mobile personnel? Does there appear to be a shifting or upgrading of persons in non-professional jobs (the technician occupations) to professional occupations?

By going back to the age 24 starting point for job histories, it may be possible to establish typical and variant career histories for specific occupations and occupational groups, for respondents with specific levels and types of training, and for those with certain demographic characteristics. Several additional general questions on employment which were included will provide some further insights into the overall work history patterns: respondents were asked to indicate all the different types of employers worked for; the number of different employers for the current field of work specialization as well as the total number of years worked in the present field of specialization; and finally, some data on the different fields of work specialization in which the respondent was engaged during his career other than those already indicated for the specific points in time requested.

Turning now to our third main area of inquiry--training--an intensive effort has been made to obtain a considerable amount of detail on various facets of both formal education and informal types of training. By and large, persons in the occupations covered in this survey have a fairly high level of training, especially when compared to the general population. Not only is some information in this area available from a variety of other studies, but we are also aware that the requirements for employment in these professional, scientific, and technical occupations require this background--even more so in the past several decades.

To begin with, since information on training was obtained as of 1962, we were able to update the Census occupational information on number of years of formal training completed. However, our primary interest lay beyond this data, in that we wanted to determine some of the specifics of higher education obtained in relation to subsequent employment. Data was therefore requested on major fields of specialization for undergraduate and graduate study at every institution attended as well as the different types of degrees granted, where appropriate. As a subsidiary question, we requested information on sources of financial support received by respondents for undergraduate and graduate or professional training. This will provide some historical insights on the varied sources of support obtained by persons who received their training in different fields of study. In addition, because of our general knowledge that a substantial amount of training takes place outside of the formal educational system, several questions were included about informal types of training received, such as company training programs, military training applicable to civilian occupations, home study correspondence courses, special workshops and seminars, etc. Some of the more apparent uses of this information includes: a detailed description of the formal education and training of persons in various professional, scientific, and technical occupations; an analysis of current occupation and field of work specialization as well as overall job histories in relation to major fields of study at both undergraduate and graduate levels; the extent to which persons with less than a college degree are employed in professional occupations, and what types of informal training as well as experience may have contributed to their attaining such positions; and, an analysis of the personal and other background characteristics of the respondents to determine whether any insight can be obtained regarding differences in levels of training and subject matter studied.

Lastly, as previously indicated, some information was sought on background and personal characteristics both to supplement data available from other sources and as factors to relate to data obtained in the areas dealing with employment and training.

The information obtained in several of these areas includes: (1) Attitudes toward work in terms of the respondent's current occupation--respondents were asked to indicate the relative importance of and degree of satisfaction with selected characteristics of occupations. Information provided here may aid in identifying clusters of values which characterize specific occupations or groups of occupations. In addition, we may obtain clues regarding continuity of employment and future turnover among persons in various occupations, and in relation to such factors as age, geographic location, and training background. (2) Marital status and fertility--this includes both the marital status of the respondent as well as the number and ages of the respondent's children as possible factors in job mobility. Also, we are provided with a measure of the reproductive rates of an important segment of the population. (3) Professional characteristics--this area covers membership in professional associations and data on publication of articles or books and presentation of papers at professional meetings. Obviously, these two characteristics are only a few of the many which could be explored regarding status or professionalism among the occupations being studied.

The second major group covered in the Postcensal Survey encompasses a sample of all college graduates broken into a number of subgroups. The two main subgroups included: (1) those in the labor force in 1960 employed in all other occupations not covered in what might be called our target occupation sample; and (2) those in the labor reserve in 1960 (employed at some time between 1950 and 1960 in occupations covered in our professional and technical occupations sample). For all these respondents, information was also obtained on their work and career histories, training background, and various personal characteristics. Not only will the data provided by the persons in this group result in a comparative analysis of the Nation's college-educated population, but in terms of our own particular interests, we will be able to determine in large measure the extent to which persons trained in scientific and technical fields were, in 1960, working in occupations seemingly unrelated to this training; the same for persons who started their careers in professional, scientific, and technical occupations and were employed elsewhere in 1960; and finally, what potential exists among those in the professional and technical labor reserve in 1960 for possible reemployment in professional, scientific, and technical fields.

In closing, I should like to reiterate our hope that the program of postcensal studies outlined here will provide many insights helpful in contending with manpower problems affecting all sectors of the economy and the national welfare. The data and information culled from these studies and added to information from other past, current, and future studies will hopefully bring us closer to the day when the formulation of policies affecting our highly trained manpower will be undertaken with much greater assurance and confidence than heretofore.

THE POSTCENSAL STUDY - DATA COLLECTION, PROCESSING AND TABULATING

Stanley Greene and David L. Kaplan, Bureau of the Census *

The Postcensal Study of Professional and Technical Manpower represents a major survey undertaking of the Bureau of the Census. There were various tasks involved covering a wide range of technical activities.

The major tasks associated with this project presently completed by the Bureau of the Census are as follows:

1. Design and printing of questionnaires and other forms.

2. A pretest covering 600 cases.

3. Sample selection of some 70,000 persons covering 45 specified professional and technical occupations and college-graduate groups from the 1960 Census of Population records.

4. Matching of selected sample cases to the 1960 Population Census schedules to obtain name and address for mailing purposes.

5. Mailing operation consisting of an original mail-out, follow-up as required by two reminder letters and, finally, a reminder letter under the National Science Foundation letterhead.

6. Independent subsampling of the two classes of nonresponses--
(1) those returned by the post office as nondeliverable, and (2) those apparently delivered but not answered. The two groups were subsampled for further follow-up by, respectively, (1) addressing new questionnaires to the "postal rejects" in care of their employers (requiring a search and match of the 1960 Census of Population returns for "names of employers" and a directory search for the corresponding address) and (2) having the "nonanswer" cases telephoned by Census Bureau enumerators in the areas covered by the Current Population Survey.

7. Manual editing and coding of the returned questionnaires.

8. Card punching the information (requiring six punch cards per case).

The following phases of the project remain to be implemented although much of the planning work has been completed:

1. Transfer of punch card data to computer tape.

2. Preparation of the computer tape record for each case and weighting of same.

3. Tallying the required tabulations.

Details of the various phases of the planning, implementation, and results are discussed in this paper.

Universe

Several major classes of people comprised the universe included in the survey. The largest class consisted of persons who were reported as being in the experienced civilian labor force in specified professional occupations in the 1960 Census.1/ This included those who were employed in the specified occupations and those who were unemployed, but whose last job was in one of the selected occupations.

The original planning called for 33 professional occupations. Three of these were dropped 2/ before the survey was taken, whereas librarians were limited to those employed in public libraries, and elementary or secondary schools and sampled as separate groups. Thus there were 31 distinct professional categories in the survey. These are listed on table 1.

A second major class included in the survey comprised those persons in the "Experienced civilian labor force" in seven technical occupation groups. The occupations included were designers, draftsmen, surveyors, medical and dental technicians, electrical and electronic technicians, other engineering and physical sciences technicians, and technicians not elsewhere classified.

In addition to the two major classes of occupations listed above, the survey included a sample of persons who had completed four or more years of college. This last major class was subdivided into the following seven groups. The first three groups were in the labor reserve 3/ in 1960. The three labor reserve groups covered:

1. Female, ages 20 to 54 years, with experience in one of the selected professional or technical occupations.

2. Other persons with experience in one of the selected professional or technical occupations.

3. All persons in labor reserve with experience in occupations not selected for the survey.

* The authors wish to acknowledge the assistance of Mr. John Priebe in preparing this paper.

The persons in the "experienced civilian labor force" who were in occupations other than those selected for the survey were subdivided into the following three groups:

1. Managers, officials, and proprietors (not elsewhere classified) who were working in the following industries:

    Agriculture, forestry and fisheries
    Mining
    Construction
    Manufacturing
    Transportation, communications, and other public utilities
    Entertainment and recreation services
    Professional and related services
    Public administration

2. Balance - Females, ages 20 to 54 years

3. All others

The remaining group consists of the remaining noninstitutional population, 20 years old and over not in the Armed Forces.

The complete list of 45 classes and the detailed components are outlined in table 1.

Design and printing of questionnaires and other forms

The original questionnaire was designed by the National Opinion Research Center. This questionnaire was reviewed for feasibility by the Bureau of the Census. These two organizations in consultation with the sponsoring agencies, developed the questionnaire that was used in the Census Bureau pretest.

The questionnaire used in the pretest consisted of eight pages divided into four sections. The first section dealt with current employment, asking questions on their present employment status, and, if working, on the respondent's occupation, industry, earnings, job activities, work attitudes, and the holding and nature of a second job.

The second section asked questions on the employment status as of April 1, 1960 (the date of the Decennial Census) and the respondent's first full-time job after reaching age 25 (an age where most persons had completed their formal education).

Section III inquired about the educational and training level of the respondent. It asked questions on the colleges attended, field of study, type of degree granted and year work was ended. This section also asked about the source of finances for their post-high school training and other types of training they may have received, such as company training programs, home study correspondence courses, and military training applicable to civilian occupations.

The last section requested background information such as age, sex, type of residence when growing up, marital status and number of dependents. An analysis of the results of the pretest questionnaire was the basis for redesigning the questionnaire. Most of the changes were in the format, but some changes were made in the items with several additions being made to the section IV on background information.

Three variations of the questionnaire were designed and used in the survey. The basic questionnaire was used for the selected professional occupations, and the three "experienced civilian labor force" classes. A variation of the basic questionnaire was used for the technicians. The major changes in this questionnaire were in the list of job activities, and the technicians were not asked work attitudes. A second variation of the questionnaire was used for the labor reserve and the last class of those not in the labor force nor the labor reserve. The major difference in this questionnaire was in the method of asking for past work experience.

A supplementary questionnaire was sent to a portion of the biologists and psychologists on sources of research support they may have received during their graduate studies.

Pretests

A feasibility pretest of this survey, covering 275 cases, was conducted in the Chicago area by the National Opinion Research Center. Another pretest was conducted by the Bureau of the Census beginning in the fall of 1961. Persons in professional and technical occupations used in this survey were selected from a special evaluation project file which provided the names and addresses of respondents. Approximately 600 cases were selected for the pretest. An original mailing was followed by two reminder mailings sent to the nonrespondents. The response rates of this pretest are given below.

	Number	Responses	
		Number	Percent
Total..........	591	419	70.9
Original mailing.....	591	254	43.0
First follow-up......	445	116	26.1
Second follow-up.....	229	49	21.4

A subsample of the nonresponse cases, amounting to 51 cases, was drawn for further follow-up activity. This work consisted of a personal phone call reminder to the nonrespondent and produced 23 additional returns. Therefore the final number of completed questionnaires received in the pretest was 442 or 74.8 percent. (A figure quite similar to our results in the main study.)

These completed questionnaires were then analyzed and tabulated focusing on the problem of nonresponse by item and inconsistency between items. The result of this analysis was the final determinant in preparing the format and wording of the questionnaires.

Sample selection

The Bureau, in consultation with the sponsoring agencies, selected the sample for the survey. First, within the limits of financing and statistical reliability, the number of sample cases required for each occupation and other group in the universe was determined. (See col. 1 of table 2.) Estimates were made--since the universe counts were not yet available at the time--of the number of cases of each of these groups that would appear on the 1960 Census 25-percent sample tape file. These two figures provided the basis for determining a differential sampling ratio for each group to supply the required number of sample cases (col. 3). Since the basic universe was not known but had to be estimated, a very liberal sampling ratio was adopted to assure that a sufficient number of sampling cases would be selected from the Census 25-percent sample file. Using these sampling ratios, the first selection (and count of the total in each category) was made by the computer on a sample "every K case" basis. The computer identified and selected by the predetermined sampling ratio each category of the sample universe (shown in col. 4).

Revisions in the groups to be surveyed were also made. For example, pharmacists were deleted from the study and became the basis of a special project.

Such revisions in the groups were cause for increasing the number of sample cases required for certain of the remaining groups (col. 2). The revised number of sample cases required for the study was then compared to the first sample selection based upon the liberal sampling fraction. A division of these two figures for each group provided a subsampling fraction (col. 5). The computer then applied the subsampling fraction to the first sample selection and selected the final sample (col.6). This was accomplished in the following manner. A random start between zero and the final sampling fraction was selected for each category. To this random start the sampling fraction (to five decimal places) was added for each case in the first sample selection. When this sum exceeded or equaled "one" the case thus identified was selected and the sum reduced by one. If the sum for the case did not equal or exceed "one" the case was not selected and the next addition was made.

The computer thus identified the sample cases and also selected for high-speed printouts, pertinent data for the sample case, providing a basis for searching original Census records for purposes of matching and name and address determination for mailing the questionnaires.

A subsample of 1,500 biologists and 1,000 psychologists was selected to receive the supplementary questionnaire on research support. These cases were selected by using a random start and every "n"th case thereafter. "n" was computed by dividing the number of cases selected to receive the supplementary questionnaire by the total number of cases in the survey with the specified occupational code.

Matching and mailing operations

When the sample was selected from the 1960 Census tapes, certain identification items were selected for each case and printed out on a listing. Some of the identification items used were the codes for State, county, enumeration district (ED), occupation, industry, age, and highest school grade completed. Each case was also assigned a control number. With this information the Census schedule books were searched to ascertain the name and address of the individual.

At the same time the names and addresses were being located, punch cards were being prepared for control purposes. These cards noted the control number, State, and a code indicating the type of questionnaire required. The name and address, as ascertained from the match of Census records, was also typed on the card. This typed address was reproduced by a Xerox process and used for the address labels. The card itself was used for check-in control (those not showing a notation of receipt of schedule being sent additional mailings as required).

Although there were 45 independent samples comprising the survey, they broke down into three major components for purposes of schedule design and into four separate groups for purposes of the mailing operation.

The mailing operation consisted of an original mailing and three follow-up mailings. Each mailing--the original and follow-up--was color coded by varying the color of the schedule. This was done primarily for control of the mailout sequence. The mailouts were divided into four groups as determined by the respondents' status in the 1960 Census. The first group represented selected professional workers in the labor force (excluding 2,500 biological scientists and psychologists).

All biological scientists and psychologists were sorted from the professional group described above. A sample of about 1,500 biological scientists and 1,000 psychologists was then merged into one group. The portion of the biological scientists and psychologists not selected in the sample was returned to their original file.

Another group consists of those persons with technical occupations. The last group is composed of the labor reserve.

The mailing pieces to each of these groups consisted of (1) the respective questionnaire, (the biological scientists and psychologists also received a supplementary questionnaire), (2) an introductory letter, (3) a "Fields of Specialization List," (4) a return envelope.

Receipts

The endeavors described in the mailing operation elicited 51,505 completed questionnaires from the original panel of 71,300. The rate of receipt amounted to 72.2 percent. This figure compares favorably with our pretest experience where the return rate amounted to 70.9 percent.

Variations in the categories may be noted in table 1. (This table shows rates of receipt by each of the 45 classes.) For the professional group, the highest receipt rate was achieved, amounting to 72.6 percent, whereas the technical worker group--somewhat under the average return rate--amounted to 63.9 percent. Among the professional workers, it may be noted that the highest return rate is 82 percent (foresters and conservationists).

Field follow-up procedures

About 12,500 of the original cases did not respond to any of the four original mailings and constituted the "nonanswer" file of nonrespondents. This group was sampled at approximately a 1 in 4 rate for personal follow-up. Thus about 3,000 cases required follow-up, all of which, by design, fell into Primary Sampling Units of the Bureau's Current Population Survey and thus an existing field staff was available to implement the procedure. The procedure called for all sample cases to be selected in the Bureau's central office and identified by their PSU number and other relevant information (name, address, phone number, appropriate schedule). This information was packaged along with required forms and instructions and sent to the Bureau's Regional Offices. The Regional Offices in turn transmitted the materials to the proper interviewers. The interviewers contacted each nonrespondent by telephone, asking them to complete a questionnaire. Those cases indicating cooperation were mailed one by the interviewer, along with a Regional Office return envelope. Those cases indicating a refusal to complete a questionnaire were asked eight basic questions on the phone.

When the interviewer completed this phase of the work, she sent a record of the results of her assignment to the Regional Office. The Regional Office matched the completed questionnaires received to the record of results. The unmatched forms for those who were mailed questionnaires were returned to the interviewers, who again called the person and proceeded to ask the basic questions.

In regard to the "postal reject" file (that group never delivered by the post office), amounting to 7,100 cases, a sample of 1,000 random cases was drawn. A further attempt to locate these cases was made through their last known employer. Since the 1960 Census results provided the name of the employer, we had a basis for operation.

The steps required to implement this follow-up required a matching and searching of the original census record. After the case was located, the company name entered on the schedule was transcribed to a special listing. The address of the establishment was then obtained by checking through city directories and other reference material. The questionnaires were then mailed to the respondent in care of his employer using the normal mailing procedures with provision made for the follow-up mailings. These activities resulted in a return rate of about 30 percent.

Coding and editing of schedules

The processing work was accomplished by dividing the work into two major portions, namely "General Coding" and "Occupation and Industry Coding." The schedules were designed to minimize coding by annotating the entry boxes where possible with predetermined punching codes. Where this was not possible, as in the cases of "institution attended," "type of degree granted," "name of sponsoring institution," "subject of training," and "State and county of residence," codes had to be predetermined and, as in the case of "subject of training," a three-digit code was formulated and a special publication prepared noting the subject field content of each broad three-digit field. Also, during the "General Coding" phase, extensive editing rules were applied to the items to account for some blanks, obvious inconsistencies, consideration of fractions, improper placement of entries, dual entries, finding midpoints of ranges (if given), conversion of income entries to codable items, conversion of improper time basis to acceptable basis. Further editing of this nature will also be implemented in the computer.

The "Occupation and Industry Coding" phase of the work was done in accordance with the 1960 Census of Population classification scheme, with some minor modifications. All clerical work was verified completely on a dependent basis.

Preparing the record and weighting

Prior to tallying the tabulations in the Postcensal Study, certain programing activities are required to prepare the computer tape record.

Each questionnaire required six 80-column punch cards to accommodate the data. This information must first be transferred from punch cards to computer tape and the six cards for each case must be consolidated into a single record for a person (eliminating the duplication of identification items required on each punch card).

Each of the 45 occupations receives a differential weight. The methodology involved in this weighting calls for a consideration of the three following classes of responses:

1. Initial responses

2. Responses from a field follow-up program

3. Responses from a file of "postal rejects"

The latter two classes have to be weighted to the totals from which they are drawn.4/ The determination of these weights will be done clerically and incorporated in the punch card. After these intermediate weights are on the record and are applied to the latter two classes, this file will be merged with the initial responses (class 1). The final weights to be applied to each occupation group would be the proportions these merged totals bear to their respective grand total as determined by the 1960 Census results.

FOOTNOTES

1/ For information on the classification of occupations in the 1960 Census, see U.S. Bureau of the Census, 1960 Census of Population, Alphabetical Index of Occupations and Industries, Revised Edition, Washington, D.C., 1960, and its companion volume U.S. Bureau of the Census, 1960 Census of Population, Classified Index of Occupations and Industries, Washington, D.C., 1960. For information on the definition of concepts used by the Bureau of the Census, see the text in the following reports: U.S. Census of Population: 1960, Detailed Characteristics, United States Summary, Final Report PC(1)-1D, Washington, D.C., 1963, and U.S. Census of Population: 1960, Occupational Characteristics, Final Report PC(2)-7A, Washington, D.C., 1963. The second report will be released in October of 1963.

2/ Professional nurses, pharmacists, and physicians and surgeons who were employed by any level of government, but not working in hospitals.

3/ In the 1960 Census the term labor reserve was used for those persons who had worked sometime during the period of 1950 to 1960, but were not in the labor force at the time of the census.

4/ The methodology outlined herein is subject to review of the reliability of the follow-up data by Bureau sampling experts.

Table 1.--DETAILED COMPONENTS OF THE UNIVERSE AND RECEIPTS IN THE POSTCENSAL STUDY OF PROFESSIONAL AND TECHNICAL MANPOWER

Occupations and other groups sampled	Number of cases in survey	Cases returned: Number	Cases returned: Percent
I. Occupations in the survey and their Census codes	71,300	51,505 1/	72.2 1/
A. Selected professional occupations	56,137	40,768	72.6
021 Chemists	2,500	1,839	73.6
College presidents, deans, and professors and instructors, nonscientific subjects	1,260	905	71.8
030 College presidents and deans			
054 Professors and instructors, nonscientific subjects			
Professors and instructors, natural science	2,501	1,856	74.2
031 Professors and instructors, agricultural sciences			
032 Professors and instructors, biological sciences			
034 Professors and instructors, chemistry			
041 Professors and instructors, geology and geophysics			
042 Professors and instructors, mathematics			
043 Professors and instructors, medical sciences			
045 Professors and instructors, physics			
052 Professors and instructors, natural sciences, not elsewhere classified			
Professors and instructors, social science	1,494	1,155	77.3
035 Professors and instructors, economics			
050 Professors and instructors, psychology			
051 Professors and instructors, statistics			
053 Professors and instructors, social sciences, not elsewhere classified			
040 Professors and instructors, engineering	2,000	1,529	76.5
060 Professors and instructors, subject not specified	1,249	873	69.9
080 Engineers, aeronautical	1,999	1,383	69.2
081 Engineers, chemical	1,270	974	76.7
082 Engineers, civil	1,948	1,354	69.5
083 Engineers, electrical	3,499	2,533	72.4
084 Engineers, industrial	2,000	1,457	72.9
085 Engineers, mechanical	1,999	1,399	70.0
090 Engineers, metallurgical and metallurgists	1,000	726	72.6
091 Engineers, mining	1,000	708	70.8
092 Engineers, sales	1,000	682	68.2
093 Engineers, not elsewhere classified	2,782	1,971	70.8
103 Foresters and conservationists with 4 or more years of college	1,000	820	82.0

1/ Figures include 966 cases received after the tally by occupation, thus detail will not add to total.

Table 1.--DETAILED COMPONENTS OF THE UNIVERSE AND RECEIPTS IN THE POSTCENSAL STUDY OF PROFESSIONAL AND TECHNICAL MANPOWER--Con.

Occupations and other groups sampled	Number of cases in survey	Cases returned	
		Number	Percent
111 Librarians - elementary and secondary schools) ..with 4 or more years of college	1,751	1,335	76.2
111 Librarians - public libraries)			
130 Agricultural scientists	1,991	1,494	75.0
131 Biological scientists	3,502	2,528	72.2
134 Geologists and geophysicists	2,000	1,351	67.6
135 Mathematicians	1,909	1,321	69.2
140 Physicists	2,295	1,714	74.7
145 Miscellaneous natural scientists	1,022	787	77.0
172 Economists	1,136	805	70.9
173 Psychologists	2,150	1,570	73.0
174 Statisticians and actuaries	1,000	716	71.6
175 Miscellaneous social scientists	878	613	69.8
182 Teachers, elementary schools (Public schools only)	2,999	2,164	72.2
183 Teachers, secondary schools	3,003	2,206	73.5
B. Selected technical occupations	7,999	5,108	63.9
072 Designers	1,000	673	67.3
074 Draftsmen	1,000	701	70.1
181 Surveyors	1,000	587	58.7
185 Technicians, medical and dental	1,000	619	61.9
190 Technicians, electrical and electronic	999	636	63.6
191 Technicians, other engineering and physical sciences	2,000	1,274	63.7
192 Technicians, not elsewhere classified	1,000	618	61.8
II. Persons with an educational attainment of four or more years of college			
A. In experienced civilian labor force and not in the selected professional or technical occupations	2,948	1,903	64.6
1. Managers, officials, and proprietors (not elsewhere classified) who were working in the following industries	943	1,903	64.6
Agriculture, forestry and fisheries			
Mining			
Construction			
Manufacturing			
Transportation, communications, and other public utilities			
Professional and related services			
Public administration			
2. Balance - Females, ages 20 to 54 years)			
3. All others)	2,005		
B. Labor reserve	3,313	2,160	65.2
1. Females, ages 20 to 54 years, with experience in one of the selected professional or technical occupations	2,000	1,681	74.2
2. Other persons with experience in one of the selected professional or technical occupations	267		
3. All persons in the labor reserve with experience in occupations not selected for the survey	1,046	479	45.8
C. Persons 20 years old or older not in the labor force, labor reserve nor institutions	903	600	66.4

Table 2.--SAMPLE SELECTION FOR THE POSTCENSAL STUDY OF PROFESSIONAL AND TECHNICAL MANPOWER

Occupation or classification	(1) Original number of sample cases required	(2) Final number of sample cases required	(3) Original liberal sampling fraction	(4) Original sample count	(5) Sub-sampling ratio	(6) Final sample selected
Total in survey	73,000	76,869	-	152,510	-	71,300
Total professional occupations	55,000	59,869	-	90,774	-	56,137
Total college presidents, deans and professors	7,000	8,500	-	11,230	-	8,504
College presidents, deans, and professors and instructors, nonscientific subjects	1,000	1,250	1/4	2,465	0.50710	1,260
Professors and instructors, natural science	2,000	2,500	1/4	2,548	0.98117	2,501
Professors and instructors, social science	1,000	1,500	1/4	2,167	0.69221	1,494
Professors and instructors, engineering	2,000	2,000	1/1	2,359	0.84782	2,000
Professors and instructors, subject not specified	1,000	1,250	1/8	1,691	0.73921	1,249
Total engineers	18,000	20,282	-	32,654	-	18,497
Engineers, aeronautical	1,500	2,000	1/4	3,284	0.60902	1,999
Engineers, chemical	2,000	2,000	1/8	1,270	1.0	1,270
Engineers, civil	2,500	2,500	1/20	1,948	1.0	1,948
Engineers, electrical	2,500	3,500	1/10	4,618	0.75791	3,499
Engineers, industrial	2,000	2,000	1/8	3,095	0.64621	2,000
Engineers, mechanical	2,500	2,500	1/20	1,999	1.0	1,999
Engineers, metallurgical and metallurgist	1,000	1,000	1/2	2,305	0.43384	1,000
Engineers, mining	1,000	1,000	1/2	1,526	0.65531	1,000
Engineers, sales	1,000	1,000	1/2	7,170	0.13948	1,000
Engineers, not elsewhere classified	2,000	2,782	1/4	5,438	0.51159	2,782
Foresters and conservationists (4 years of college)	1,000	1,000	1/1	2,936	0.34060	1,000
Librarians	2,000	2,000	1/4	5,250	4 yrs.college	1,751
Total natural scientists	16,000	16,800	-	19,237	-	15,219
Agricultural scientists	2,000	2,000	1/1	1,991	1.0	1,991
Biological scientists	4,000	4,000	1/1	3,502	1.0	3,502
Chemists	2,000	2,500	1/8	2,617	0.95530	2,500
Geologists and geophysicists	2,000	2,000	1/1	4,695	0.42599	2,000
Mathematicians	2,000	2,000	1/1	1,909	1.0	1,909
Physicists	2,000	2,300	1/1	3,501	0.65696	2,295
Miscellaneous natural scientists	2,000	2,000	1/1	1,022	1.0	1,022

Table 2.--SAMPLE SELECTION FOR THE POSTCENSAL STUDY OF PROFESSIONAL AND TECHNICAL MANPOWER--Con.

Occupation or classification	(1) Original number of sample cases required	(2) Final number of sample cases required	(3) Original liberal sampling fraction	(4) Original sample count	(5) Sub-sampling ratio	(6) Final sample selected
Total social scientists	5,000	5,287	-	10,080	-	5,164
Economists	1,000	1,137	1/1	4,814	0.23619	1,136
Psychologists	2,000	2,150	1/1	3,014	0.71334	2,150
Statisticians and actuaries	1,000	1,000	1/4	1,373	0.72834	1,000
Miscellaneous social scientists	1,000	1,000	1/1	879	1.0	878
Teachers, elementary public schools	3,000	3,000	1/50	4,197	0.71480	2,999
Teachers, secondary schools	3,000	3,000	1/25	5,190	0.57804	3,003
Total technicians	7,000	8,000	-	32,934	-	7,999
Designers	1,000	1,000	1/10	1,672	0.59809	1,000
Draftsmen	1,000	1,000	1/50	1,061	0.94251	1,000
Surveyors	1,000	1,000	1/5	2,291	0.43650	1,000
Technicians, medical and dental	1,000	1,000	1/20	1,734	0.57671	1,000
Technicians, electrical and electronic	1,000	1,000	1/1	23,176	0.04315	999
Technicians, other engineering and physical science	1,000	2,000	1/10	4,684	0.42699	2,000
Technicians, not elsewhere classified	1,000	1,000	1/5	3,340	0.29941	1,000
Persons in "Other" groups with 4 years of college	11,000	9,000	-	23,778	-	7,164
Experienced civilian labor force not in target occupations	3,000	3,000	-	16,522	-	2,948
Selected managers	1,000	1,000	1/100	943	1.0	943
Balance females ages 20 to 54	1,000	1,000	1/20	8,207	0.10309)	2,005
All others	1,000	1,000	1/100	7,373	0.13211)	
Labor reserve	5,000	4,000	-	6,353	-	3,313
Females ages 20 to 54 in target occupations	4,000	2,000	1/25	2,950	0.67797	2,000
All others in target occupations	1,000	1,000	1/100	267	1.0	267
Not in target occupations	-	1,000	-	3,136	0.33333	1,046
Persons 20 years old or over not in the labor force, labor reserve nor institutions	2,000	2,000	1/200	903	1.0	903

PHYSICISTS AND MATHEMATICIANS IN THE POST CENSAL SURVEY OF SCIENTIFIC AND TECHNICAL PERSONNEL*

Seymour Warkov, National Opinion Research Center

The rapid growth of the scientific and technical occupations during the past decade has been well documented. However, the occupational explosion involving engineers, technicians and physical, natural and social scientists, has not been explored in sufficient detail to permit a comprehensive picture to emerge of the duties and job content, training history, patterns of mobility and social characteristics of persons in these scientific and technical occupations. To meet the need for this kind of information, the National Science Foundation has commissioned the series of post-enumeration studies of scientific and technical workers that is the topic of this symposium. A major portion of these studies, entitled "The Post Censal Survey of Technical and Scientific Manpower," was committed to the comparative analysis of some forty-five scientific and technical occupations classified in the 1960 Census of Population among "Professional, Technical and Kindred" workers under the three-digit occupational code.

This paper presents some of the preliminary findings for two of the forty-four titles selected for study: Mathematicians and Physicists. As you know, the 1960 Census classified one-fourth of the population by occupation and a number of other key characteristics. This listing provided the frame for drawing samples of one thousand to five thousand persons per occupation. Some 1,300 Mathematicians and 1,700 Physicists were sampled: the data presented in this paper are based on a take rate of 69 per cent for Mathematicians and 75 per cent for Physicists.

A special deck of punched cards was prepared by the Bureau of Census[1] for our use at NORC. Accordingly, I wish to stress that these findings are tentative and subject to revision at the time machine tabulations are provided by the Bureau of the Census. Furthermore, the data may be heavily skewed in the direction of Physicists employed in non-academic positions since Physicists also could be classified under the title "Professors and Instructors of Physics." When the data are weighted according to the sampling ratios employed in selection of respondents in each of the occupational titles, it will then be possible to merge the two sets of data and provide a more balanced portrait of Physicists. With these reservations in mind, let me begin.

Today we examine the composition of the two groups in terms of their ages, sex and educational attainment and then compare the two occupations on a number of items related to their 1962 occupation and 1962 employment in terms of these crucial demographic and social variables.

Age, Sex and Education

"Mathematicians" in the 1960 Census differ from "Physicists" as follows:

First, Mathematicians are in a more youthful occupational category with 59 per cent less than 35 years of age as compared with 41 per cent of the Physicists;

Second, while both are essentially male occupations, Mathematicians are in the more heterogeneous group; one out of four is a woman in contrast with one out of twenty Physicists; and

Third, both occupations recruit heavily from the ranks of college graduates. Fully 87 per cent of the Mathematicians have at least a Bachelor's degree as do 89 per cent of the Physicists, but the educational summit of higher education was reached more frequently by the latter: 28 per cent have a doctorate as compared with 11 per cent of the Mathematicians.

The distribution of men and women among the age-education segments of the sample of Mathematicians is presented in Table I.b. Younger Mathematicians with advanced degrees almost invariably are men; indeed, all Ph.D.'s under age 45 in this sample are males--while women almost equal their male counterparts among the older Mathematicians advancing no further by 1962 than the Bachelor's degree (only 50 per cent of the B.A.'s 45 years and older are male). Female Mathematicians hardly ever take advanced degrees. Furthermore, there is a hint of a life cycle effect: women appear in substantial numbers in the ranks of Mathematicians after age 44 when, presumably, child rearing tasks are completed.

Putting these findings together, one would conclude that the greater educational attainments of the Physicists as measured by the percentage holding the doctorate is accounted for by occupational differences in sex composition. Table I.c shows that this is not the case, however. Even when comparisons are made for men only, more Physicists hold the doctorate. As expected, the percentage of workers holding the Ph.D. increases among both groups in each successive age grade; but it <u>is</u> surprising that within each age grade, relatively more persons identified as Physicists in the 1960 Census should hold the doctorate. While explanations of this difference readily come to mind, suffice it to note that vital

*This paper was made possible by funds provided under National Science Foundation contract NSF-C288. The statements made and views expressed are solely the responsibility of the author.

[1]I wish to thank Mr. Stanley Greene, Population Division, Bureau of the Census, for rendering this form of assistance. I also wish to acknowledge the able support provided by Sanford Abrams in preparing the tabular materials used in this paper.

differences do exist among the two occupations and need to be considered in subsequent analyses.

Work Status, 1962

Virtually all persons identified as Physicists and Mathematicians in the 1960 Experienced Civilian Labor Force (ECLF) were still employed two years later. Among Physicists only three per cent were no longer in the ECLF at the time they returned their completed questionnaires to the Bureau of the Census and seven per cent of the Mathematicians were likewise removed from the labor force. While both are relatively youthful occupations, these low rates of withdrawal testify once again to the efficacy of formal education at and beyond the college level in securing and maintaining a position in the labor market. (Parenthetically, of the 92 Mathematicians no longer in the ECLF, 73 were women.)

Occupation, 1962

We now consider a major topic on occupational analysis, namely, how many persons identified as incumbents of an occupation at one point in time are identically employed and classified two years later? Changes in the occupational designation of workers may result from (1) a change in job and job content; (2) respondent error involving a change in label although the same work functions are performed; and (3) transcription error. Whatever the reason may be, let us for the moment document the amount of occupational change experienced by the two groups as indicated by the occupational classification employed and applied by Census personnel to responses to the question: "What kind of work were you doing (last week)?" The reference point at time 1 is April, 1960; at time 2 it is Summer, 1962. Table III shows that some 70 per cent of the workers classified as Physicists in the 1960 Census of Population were again classified as Physicists in the 1962 Post Censal Survey; only 56 per cent of the Mathematicians retained the same occupational label some 27 months later. I think we all agree that the amount of occupational change exhibited in this two-year follow-up survey is startling. Certainly, it raises questions about the system of classification currently employed to identify persons in these two occupations and perhaps in other scientific, technical and engineering occupations as well.

Tables III.a, b, and c suggest that changes in occupational affiliation found in these two samples are not simply a function of procedural vagaries in classification because age, sex and educational attainment all seem to be systematically involved. Table III.a shows that among male Physicists, occupational stability declines after age 45; for male Mathematicians, there are no age-associated differences in movement; and among female Mathematicians, occupational stability slightly increases with age. The second part of the table demonstrates the holding power of an occupation for men in both samples as they ascend the educational hierarchy, with Ph.D.'s the most likely to stay put occupationally.

The situation is further complicated when age and education are jointly considered. Among Physicists, age and education generate an effect such that young Ph.D.'s are least likely to switch occupations (90 per cent are still in Physics), old Bachelors' are much more likely to have moved out (only 55 per cent of those 45 years of age or more are still there) while only 34 per cent of the older men without a Bachelor's degree are still in Physics. In Mathematics, the picture is muddied because of the effect of sex on occupational stability. Note that among the Ph.D.'s--that exclusive domain of men insofar as these two occupations are concerned--occupational stability declined with age (the percentage distributions are almost identical in the two groups) while occupational stability increased with age among Mathematicians who are no further along than the Bachelor's degree.

Table III.c presents rates of occupational stability among Mathematicians when all three variables are taken into account. Among the men, age and education work together to provide a pattern similar to that found among the Physicists. Thus: 89 per cent of the young Ph.D.'s remain in the occupation; only 29 per cent of the older men lacking academic degrees do likewise. There is no coherent pattern among the women. Young women in the 1962 ECLF who lack degrees are least likely to remain in Mathematics (28 per cent) while older women holding the Bachelor's degree in three out of four cases are still there. In sum: relative youth and advanced degrees both promote occupational stability among men in Mathematics and Physics; for women in Mathematics, additional information on the life cycle and family formation should help us discern a pattern.

Employment Mobility: 1960-1962

The preceding analysis of occupational changes as measured by the three-digit Census occupational code raises as many questions as it answers. One of the first concerns the relationship between employment and occupation. Do persons who change employment also change occupations? Is the opposite true? Or do the two types of change occur independently? Let us begin by looking at the patterns of employment of workers classified as Mathematicians and Physicists in 1960. The relevant questionnaire item asked: "Were you working for the same company, business or organization in April, 1960 as you were in your major employment last week?"

Some 81 per cent of the Physicists and 80 per cent of the Mathematicians were with the same employer two years later. Sex-age rates of employment stability are given in Table IV.a. Chances of changing employers during the two-year interval decline with age, and within each age group male Mathematicians are more likely to change than male Physicists, while female Mathematicians are the least likely to switch employers during the two-year interval. Among males, the occupational difference persists even when age grade is taken into account; and among Mathematicians, sex differences also persist when age is held constant up to age 45.

Differences among Physicists in rates of employment mobility in each age grade depend on educational attainment as well. Table IV.b shows that workers who hold the doctorate are more likely to move than those holding the Bachelor's: thus, 68 per cent of the young Ph.D.'s are still with the same employer in constrast with 91 per cent of the older Physicists taking the Bachelor's and 99 per cent of the older men without degrees. Among the Mathematicians, there is a pattern related to age but not to education. Once again, the sex composition of Mathematicians may be obscuring the relationship.

The effects of age and education on employment and occupation mobility among Physicists should be noted. Occupational stability and employment mobility seem to go hand in hand for the young Ph.D. (they stay with Physics but change employers) while the opposite is true of the older Physicist lacking postgraduate credentials (the latter stay with their employer but change Census-designated occupation).

Types of Mobility

The distinction between occupation and employment is one which we wish to consider in greater detail. Because of time limitations, the analysis is limited to Physicists. We have shown that the same socio-demographic variables have different effects on rates of occupational mobility and employment mobility among Physicists when each is treated separately. Our next step is to see what happens when they are put together.

A combination of the two types of mobility behavior yields the following classification of Physicists:

--Stable Physicists maintain both occupation and employment affiliation during the two-year interval;

--Itinerant Physicists move on to other, presumably greener, pastures but retain their occupational affiliation;

--Organization Professionals maintain their employment ties but move out of Physics during the two-year study period into other professional occupational categories; and

--Mobile Professionals are workers making a double switch.

Of course, this classification of the different modes of adapting to the world of work is somewhat arbitrary for in the long run everybody will have moved out of their occupation and employment. Nevertheless, after a twenty-seven month interval about six in ten Physicists remain stable, one in ten is an Itinerant, two in ten are no longer Physicists in the same organization and one in ten has changed both employer and occupation.

Reading across Table V.a we see that the probability of becoming an Itinerant is more than twice as great for the man with the doctorate as for the Bachelor's, but the chances of remaining in the same employing organization in a different occupation are more than twice as great for the 1960 Physicist with the Bachelor's degree than for the Ph.D. Furthermore, an advanced degree makes a difference for remaining a Stable Physicist but all are equally likely to become displaced occupationally and organizationally.

The importance of age grade and formal education for career patterns among types of Physicists is indicated in Table V.b. The upper panel identifies the Itinerant Physicists within each age-education segment of the 1960 sample. The young Ph.D. is most likely to have moved within the twenty-seven month period to another organization while remaining a Physicist (23 per cent did) in contrast with three per cent of the older Physicists who were Bachelor's recipients only and none among those lacking degrees. The probability of becoming an Organization Professional--one no longer identified as a Physicist --and a Mobile Ex-Physicist also become sharply differentiated when age and educational attainments are jointly considered. For example, only three per cent of the young Ph.D.'s remain in the organization but change occupations; 40 per cent of the older Bachelors' and 66 per cent of those without a four-year degree who were called Physicists in 1960 have now moved into other occupations within the employing organization.

While age and education are shown to be importantly linked to the prospects for a change in employment and/or a change in occupation, there are other factors as well. Consider the field of training for which the highest degree was obtained: Do Physicists who trained in the field of Physics show the same propensity for mobility as do 1960 Physicists whose highest degree was obtained in other fields? Since college and graduate training entail substantial investments of time, one would anticipate the emergence of a sense of commitment to the area of work for which training was secured. Table VI shows that Itinerant Physicists were most likely to have secured their highest academic degree in the field of Physics while Organization Professionals, of whom only one in three received their highest academic degree in Physics, were least likely to have the field of formal training correspond to their 1960 occupation.

Organizational Roles

Central to the entire problem of change of occupation given the current classifications is the nature of the work role within the employing organization. As a first, quick attempt to understand the substantial change in occupations describing the kind of work 1960 Physicists were doing in their 1962 employment, we examined responses to a question asking for a description of the work role in terms of the following:

Are you....(check one)

An administrator (concerned mainly with policy making, planning, overall supervision)?

A supervisor (concerned mainly with technical matters)?

A coordinator (concerned mainly with liaison)?

Other?

Administrator and Coordinator imply managerial roles at the periphery of the professions' core work. If this is the case, then workers changing occupations during the two-year interval may have been promoted to administrative positions. They are, perhaps, involved in managerial tasks that are no longer appropriately termed Physicist. Table VII supports this notion: reading down the table, we see that the Itinerant Physicist and the Stable Physicist are half as likely to designate themselves in managerial terms as are the other two types of workers. Thus, a change in occupation frequently appears to entail involvement in managerial tasks. On the other hand, technical supervision is more frequently related to employment stability: the Stable Physicist is most likely to be in technical supervision; the man who has made a double switch, least likely.

Work Activities in 1962 Employment

A closer approximation of the occupational roles of Physicists is afforded by a thirty-item inventory of work activities. Respondents were asked to check off all activities which "may be part of your major current position." The question that followed asked: "Of all these you checked above, which TWO did you spend the most time doing?" The analysis today is based on the responses to the latter question. Table VIII analyzes three of the "two-most-time-consuming" activities in the 1962 employment of 1960 Physicists by age, grade, and educational attainment.

The first panel shows that eight out of 10 young Ph.D.'s are engaged in "basic" research but only three out of 10 older Bachelors' consider basic research as one of their two-most-time-consuming activities. Young Bachelor's and Master's degree Physicists have relatively high levels of participation in "applied" research. In contrast, young Ph.D.'s are less likely to describe their work as applied but their older counterparts get increasingly involved in this type of work. The likelihood of applied research declines with age at the Master's level and there is a slight age-related decline among Bachelor's. Because our data are cross-sectional, not longitudinal, we cannot tell whether a pattern of work entailing applied and basic research represents a genuine transformation in occupational roles as one grows older or whether generational differences alone account for the higher incidence of basic research among young, highly trained Physicists. To describe the change in content of work would, of course, require that cohorts be followed over a period of time. Nevertheless the findings suggest that as young researchers seeking "basic" answers "burn out," they move into other roles. In addition to the presumably less arduous task of conducting applied research, Physicists of all academic stripes more frequently "administer and supervise research and development after age 35." These findings support the proposition that there is a career sequence in the occupational histories of Physicists, each career stage absorbing these professionals in its own distinctive round of activities.

Salary Rates

The final topic which is considered today concerns another key factor in mobility behavior: financial rewards. We do not know the 1960 salary rate of Physicists but this information was secured for the 1962 major employment. Before we look at the salary rates for "current major employment" among the four types of Physicists, I wish to show how age grade and educational attainment together with a change in occupation during the two-year interval affect salary rates for 1962 employment. Inspection of Table IX.a reveals that age grade and highest degree attained are almost equally predictive of the proportion of 1960 Physicists having salary rates of $12,000 or more in their 1962 major employment. An added advantage accrues to the 1960 Physicist who changes occupation during the two-year interval provided that he is less than 45 years of age. Beyond this point in the life cycle, a change in occupation appears to be irrelevant. Thus: one out of ten Physicists under 35 years of age with Bachelor's degrees earns $12,000 or more while nine out of ten Ph.D.'s in the older age brackets work for comparable salaries.

In Table IX.b data are given for salary rates by age and highest academic degree among the four types of Physicists. It contains a complicated story. Reading down each column, it is noteworthy that:

First, each group of Physicists (classified by type of mobility behavior) exhibits an orderly progression in salary by education and age. Invariably, 1960 Physicists 35 years of age and older who secured the doctorate are most likely to be earning $12,000 or more while the young man without a Ph.D. is least likely to be at this level.

And second, we see that education is the primary variable accounting for salary differences, age always exercising a secondary influence. As a result the young Ph.D. more frequently earns this sum than the older man who stopped short of the vaunted academic mark. Parenthetically, we have already shown that the young Ph.D. hardly ever becomes an Organization Professional or makes the double switch; hence two cells in the table are almost vacant.

Reading across the table, we see that within each age-education segment of the sample, differences obtain among the four types of Physicists in their ability to command salaries of \$12,000 or more. Among old Ph.D.'s, <u>everybody</u> who changed occupation <u>and</u> employer is working at the higher salary rate, testifying perhaps to the efficacy of money in allocating manpower in short supply. Close to nine out of ten Stable Physicists and Organization Ex-Physicists do as well followed by eight out of ten in the remaining group. Among the young Ph.D.'s the direction in differences is similar with Stable Physicists having only a slight advantage over the Itinerant, suggesting in turn that career factors other than money are of greater significance in differentiating Itinerants and Stable Physicists at earlier stages in scientific careers. Among the older Physicists below the level of the doctorate, men who make the double switch again are the most likely to be earning higher salaries (66 per cent do) but there the similarity to their Ph.D. counterparts ends: the Stable Physicists among the latter rank well below the men who make the double change in the proportion who earn higher salaries, and the older non-Ph.D. who is occupationally mobile but organizationally stable is <u>least</u> likely to reach the higher salary rate. Among the younger men without the doctorate, the pattern is similar to the one described above: the Stable Physicist is least likely to be earning \$12,000, the Itinerant most likely to make it.

In sum: Education and age group together tell a substantial part of the story on salary rates among Physicists while occupational change was found to provide an added increment up to age 45. When occupational change was combined with employment mobility to yield the four patterns of behavior during the twenty-seven month interval, it was found that at each age and education level the reward system operated differently among the four types of Physicists. Salary differences attributable to type of mobility were most substantial among the older non-Ph.D.'s; factors extrinsic to core professional work appear to be more salient at this career stage than is the case among young Ph.D.'s whose mobility behavior seems to have little influence on salary rates.

<u>Summary</u>

The purpose of this paper was to present some of the initial findings for two of the forty-four occupations included in the Post Censal Survey of Scientific and Technical Manpower. Our work to date suggests that a useful start can be made in explaining both inter- and intra-occupational variation by examining the age, sex and educational attainment components of occupations. The more detailed analysis of Physicists indicated that much can be learned about occupational life through an internal analysis of the data--a "case study" of the occupation, as it were--but the student of the sociology of occupations leans toward the comparative perspective. In the near future tabulations will be available for all forty-five occupations providing materials out of which we hope to fashion a broad, systematic analysis of scientific, engineering and technical occupations.

In the course of our inquiry into "the relationship between training and subsequent occupation,"[2] occupations are to be differentiated in terms of work roles, training patterns, mobility behavior and the like. To take one example, the distinctive age-education patterns of involvement by Physicists in basic and applied research, and in administering and supervising research and development probably do not obtain across the board. Empirically determined occupational differentiation, then, opens the way for constructing new typologies of occupations. The materials which were reported today illustrate the types of comparisons that we plan to extend across the entire spectrum of occupations in the sample.

[2] National Science Foundation, <u>A Program for National Information on Scientific and Technical Personnel</u>, NSF 58-28 (August, 1958), Page 6.

TABLE I.a

SEX, AGE, AND HIGHEST DEGREE ATTAINED BY PHYSICISTS AND MATHEMATICIANS

Occupation	Sex		N	NA	Total
	Men	Women			
Physics . .	95	5	1,704	6	1,710
Mathematics.	74	26	1,304	7	1,311

Occupation	Age			N	NA	Total
	Under 35	35-44	45 and over			
Physics . .	41	41	18	1,703	7	1,710
Mathematics.	59	29	12	1,302	9	1,311

Occupation	Highest Degree Attained				N	Other & NA	Total
	No Degree	Bachelors	Masters	Ph.D.			
Physics . .	11	37	24	28	1,689	21	1,710
Mathematics.	13	49	27	11	1,295	16	1,311

TABLE I.b

AGE AND HIGHEST DEGREE ATTAINED BY SEX

Age	Physics - % Male			
	No Degree	Bachelors	Masters	Ph.D.
Under 35	96 (49)	94 (321)	95 (180)	98 (146)
35-44	94 (66)	99 (233)	96 (163)	99 (227)
45 and over . . .	78 (73)	94 (72)	95 (57)	100 (97)

Total = ... 1,684
NA + Other Degree 26
Total Physics 1,710

Age	Mathematics - % Male			
Under 35	58 (95)	69 (429)	88 (185)	100 (54)
35-44	64 (39)	70 (152)	85 (123)	98 (58)
45 and over . . .	59 (37)	50 (52)	74 (38)	92 (25)

Total = ... 1,287
NA + Other Degree 24
Total Mathematics 1,311

TABLE I.c

SEX BY AGE BY DEGREE

Sex	Age	Per Cent				N
		No Degree	Bachelors	Masters	Ph.D.	
		Physics				
Men	Under 35	7	45	26	22	663
	35-44	1	34	23	33	672
	45 and over . . .	21	25	20	35	276
Women	Under 35	6	58	27	9	33
	35-44	24	18	41	18	17
	45 and over . . .	70	17	13	0	23
		Total = ...				1,684
		NA + Other Degree				26
		Total Physics				1,710
		Mathematics				
Men	Under 35	10	52	29	9	568
	35-44	9	37	35	19	293
	45 and over . . .	22	26	29	23	99
Women	Under 35	21	68	11	0	195
	35-44	18	57	24	1	79
	45 and over . . .	28	49	19	4	53
		Total = ...				1,287
		NA + Other Degree				24
		Total Mathematics				1,311

TABLE II.a

WORK STATUS

Occupation	In Labor Force	Out of Labor Force	Total		NA	N
			Per cent	N		
Physics . .	97	3	100	1,707	3	1,710
Mathematics	93	7	100	1,310	1	1,311

TABLE II.b

SEX BY WORK STATUS

(Per cent Out of Labor Force)

Occupation	Men	Women	N	NA	Total
Physics . . .	2 (1,624)	22 (77)	1,701	9	1,710
Mathematics .	2 (970)	22 (333)	1,303	8	1,311

TABLE II.c

AGE BY WORK STATUS

(Per cent Out of Labor Force)

Occupation	Under 35	35-44	45 and over	N	NA	Total
Physics . .	4 (702)	1 (695)	6 (303)	1,700	10	1,710
Mathematics	9 (771)	4 (375)	4 (155)	1,301	10	1,311

TABLE II.d

DEGREE BY WORK STATUS

(Per cent Out of Labor Force)

Occupation	No Degree	Bachelors	Masters	Ph.D.	N	Other and NA	Total
Physics . . .	7 (189)	3 (626)	4 (400)	2 (471)	1,686	24	1,710
Mathematics .	12 (172)	9 (635)	3 (348)	0 (139)	1,294	17	1,311

TABLE III

OCCUPATION - 1962

Occupation - 1960	Same as 1960	Engineering	Other Profession	All Other	Total		Out of Labor Force	Total
					Per cent	N		
Physics . .	70	16	12	4	100	1,654	56	1,710
Mathematics	56	6	31	4	100	1,223	88	1,311

TABLE III.a

AGE, SEX AND OCCUPATION - 1960

(Per cent Occupation Same - 1962)

Occupation - 1960	Sex	Age		
		Under 35	35-44	45 and over
Physics	Men	71 (650)	71 (669)	61 (266)
	Women	71 (21)	53 (19)	20 (20)
Total		1,645		
Out of Labor Force		56		
NA + ...		9		
Total Physics		1,710		
Mathematics . . .	Men	54 (562)	62 (290)	54 (98)
	Women	54 (139)	61 (71)	61 (51)
Total		1,211		
Out of Labor Force		88		
NA + ...		12		
Total Mathematics		1,311		

TABLE III.b

AGE BY DEGREE BY OCCUPATION - 1960

(Per cent Occupation Same in 1962)

Occupation - 1960	Age	Highest Academic Degree			
		No Degree	Bachelors	Masters	Ph.D.
Physics	Under 35	52 (46)	63 (310)	75 (168)	90 (144)
	35-44	42 (65)	69 (230)	72 (162)	82 (226)
	45 and over	34 (63)	55 (69)	73 (55)	71 (94)

Total	1,632
Out of Labor Force + . . .	56
Other Degree and NA + . . .	22
Total Physics	1,710

Occupation - 1960	Age	No Degree	Bachelors	Masters	Ph.D.
Mathematics . .	Under 35	36 (81)	50 (383)	61 (179)	89 (54)
	35-44	46 (37)	53 (145)	66 (120)	81 (58)
	45 and over	33 (33)	65 (51)	62 (37)	64 (25)

Total	1,203
Out of Labor Force + . . .	88
Other Degree and NA + . . .	20
Total Mathematics	1,311

TABLE III.c

AGE BY SEX BY DEGREE, MATHEMATICIANS - 1960

(Per cent Occupation Same - 1962)

Sex	Age	Highest Academic Degree			
		No Degree	Bachelors	Masters	Ph.D.
Men . .	Under 35	32 (53)	47 (291)	62 (161)	89 (54)
	35-44	37 (24)	51 (106)	67 (102)	81 (57)
	45 and over	29 (21)	56 (25)	68 (28)	61 (23)
Women .	Under 35	28 (43)	66 (82)	53 (17)	(0)
	35-44	62 (13)	61 (38)	61 (18)	- (1)
	45 and over	42 (12)	73 (26)	44 (9)	- (2)

Total	1,206
Out of Labor Force	88
Other Degree and NA	17
Total Mathematicians	1,311

TABLE IV

EMPLOYMENT MOBILITY, 1960-1962

Occupation 1960	Same Job	Different Job	Total		Out of Labor Force	NA	Total
			%	N			
Physics	81	19	100	1,650	56	4	1,710
Mathematics	80	20	100	1,220	88	3	1,311

TABLE IV.a

SEX BY AGE BY EMPLOYMENT MOBILITY, 1960-1962

(Per cent Same Job)

Occupation 1960	Sex	Age		
		Under 35	35-44	45 and over
Physicists	Men	74 (649)	83 (669)	92 (266)
	Women	70 (20)	74 (19)	95 (20)

Total = ...	1,643
Out of Labor Force	56
NA	11
Total Physicists	1,710

Occupation 1960	Sex	Under 35	35-44	45 and over
Mathematicians	Men	69 (562)	77 (290)	87 (98)
	Women	72 (139)	86 (71)	95 (51)

Total = ...	1,211
Out of Labor Force	88
NA	12
Total Mathematicians . . .	1,311

TABLE IV.b

AGE BY DEGREE BY EMPLOYMENT MOBILITY, 1960-1962

(Per cent Occupation Same in 1962)

Occupation - 1960	Age	Highest Academic Degree			
		No Degree	Bachelors	Masters	Ph.D.
Physicists . . .	Under 35	76 (46)	78 (308)	66 (168)	68 (142)
	35-44	88 (65)	89 (230)	85 (162)	77 (226)
	45 and over	99 (63)	91 (69)	93 (55)	87 (94)

Total	1,628
Out of Labor Force	56
Other Degree and NA	26
Total Physicists	1,710

Occupation - 1960	Age	No Degree	Bachelors	Masters	Ph.D.
Mathematicians	Under 35	66 (81)	72 (383)	68 (179)	67 (54)
	35-44	76 (37)	76 (145)	82 (120)	76 (58)
	45 and over	88 (33)	90 (51)	92 (37)	92 (25)

Total	1,203
Out of Labor Force	88
Other Degrees and NA . . .	20
Total Mathematicians . . .	1,311

TABLE V

EMPLOYMENT AND OCCUPATION MOBILITY, 1960-1962, AMONG 1960 PHYSICISTS

Occupation Change	Employment Change	Type	Per cent
No	No	Stable Physicist	58
Yes	No	Organization Professional	23
No	Yes	Itinerant Physicist	11
Yes	Yes	Mobile Ex-physicist	8
Total . .			100

N = ...	1,623
Out of Labor Force, 1962 . . .	56
NA	31
Total	1,710

TABLE V.a

MOBILITY TYPE BY HIGHEST ACADEMIC DEGREE, 1960 PHYSICISTS

Highest Academic Degree	Mobility Type				Total	
	Stable	Organization	Itinerant	Mobile	Per cent	N
No degree	38	51	3	8	100	173
Bachelors	57	26	8	9	100	599
Masters	61	18	12	8	100	377
Ph.D.	66	11	16	6	100	455

Total	1,604
Out of Labor Force	56
Other Degrees and NA	50
Total Physicists	1,710

TABLE V.b

AGE BY HIGHEST ACADEMIC DEGREE BY MOBILITY BEHAVIOR

Age	Academic Degree			
	No Degree	Bachelors	Masters	Ph.D.
	Per Cent Itinerant			
Under 35 . . .	4 (46)	9 (302)	22 (161)	23 (135)
35-44	5 (64)	9 (229)	5 (161)	15 (225)
45 and over .	0 (62)	3 (68)	5 (55)	7 (94)
	Per Cent Organization Professional			
Under 35 . . .	28	25	15	3
35-44	53	24	19	11
45 and over .	66	40	25	23
	Per Cent Stable Physicists			
Under 35 . . .	48	55	53	69
35-44	36	60	67	66
45 and over .	34	53	67	64
	Per Cent Mobile Ex-Physicists			
Under 35 . . .	20	11	10	5
35-44	6	7	9	8
45 and over .	0	4	2	5

TABLE VI

TYPE OF MOBILITY BY ACADEMIC FIELD OF HIGHEST DEGREE ATTAINED, 1960 PHYSICISTS

(% Highest Degree in Physics)

Mobility Type	Per Cent Physics
Itinerant	79 (175)
Stable	68 (928)
Mobile	50 (115)
Organization	31 (368)
N = ...	1,596
Out of Labor Force 1962. . . .	56
NA	58
Total Physicists	1,710

TABLE VII

MOBILITY BEHAVIOR AND ORGANIZATIONAL ROLE

(Per Cent Organizational Roles)*

Mobility Type	Organizational Role					Total	
	Administrator (1)	Coordinator (2)	Total (1) + (2)	Supervisor (3)	Other	%	N
Stable	8	4	12	49	39	100	846
Organization	13	10	23	43	34	100	324
Itinerant	7	4	11	39	49	100	163
Mobile	20	8	28	31	42	100	115

Total	1,448
NA	206
Out of Labor Force	56
Total Physicists	1,710

*Multiple choice item.

TABLE VIII

AGE BY DEGREE BY WORK ACTIVITIES IN 1962 EMPLOYMENT, 1960 PHYSICISTS

Age	Degree	Basic Research	Applied Research	Administer R & D	N
Under 35	No Degree	32	22	4	46
	Bachelors	28	48	10	304
	Masters	40	55	12	165
	Ph.D.	80	27	15	142
35 - 44	No Degree	25	41	11	64
	Bachelors	18	43	27	223
	Masters	29	44	29	159
	Ph.D.	57	38	39	222
45 and Over	No Degree	17	26	6	53
	Bachelors	17	38	27	66
	Masters	25	49	28	53
	Ph.D.	46	39	42	92

Total	1,589
Out of Labor Force	56
Other Degree & NA	65
Total Physicists	1,710

TABLE IX.a

AGE, HIGHEST ACADEMIC DEGREE AND
SALARY RATE, 1960 PHYSICISTS
(% $12,000 or MORE)

Age	No Degree	Bachelor	Masters	Ph.D
Under	13 (46	12 (305)	23 (163)	65 (139)
35 - 44	27 (62)	39 (225)	61 (160)	86 (220)
45 and Over	30 (53)	55 (66)	55 (53)	90 (89)

Total	1,580
Out of Labor Force.	56
Other Degree & NA	74
Total Physicists	1,710

TABLE IX.b

ACADEMIC DEGREE, AGE AND MOBILITY BEHAVIOR OF 1960 PHYSICISTS
(Per cent 1962 Salary Rate of $12,000 or More)

Academic Degree	Age	Type of Mobility			
		Mobile Ex-Physicist	Stable Physicist	Itinerant Physicist	Organization Ex-Physicist
Ph.D.	35 and over	100 (20)	88 (104)	82 (39)	87 (45)
	Under 35	- (7)	67 (90)	65 (31)	- (4)
No Ph.D. . . .	35 and over	66 (38)	44 (357)	64 (36)	41 (185)
	Under 35	21 (57)	12 (268)	28 (64)	14 (111)

Total	1,556
Out of Labor Force	56
NA	98
Total Physicists	1,710

IX

AGING AND WAGE CHANGES: THE USE OF COHORT DATA

Chairman, Edwin D. Goldfield, U. S. Bureau of the Census

CHANGES IN ANNUAL WAGE CREDITS AS WORKERS AGE: A COHORT ANALYSIS[a/]

David J. Farber, Bureau of Apprenticeship and Training
U. S. Department of Labor

1. Introduction

The purpose of this paper is to examine the employment and wage histories of a cohort of wage earners as they aged 7 years during the 1951-57 period. These histories are not complete, and may not be completely representative of the history of the "typical" wage earner. From the records maintained by the Social Security Administration, however, it is possible to trace the experience of a cohort of 61,202 workers--39,018 men and 22,184 women wage earners--with respect to that part of their employment covered by the Social Security program, and to the wage credits they received as a result of such employment.

In this paper, we shall analyze the relationship between cyclical change and the wage credits of the cohort. We shall pay particular attention to the 1954 recession, and to its differing effects on the wage credits of the lower and higher paid male and female cohorts.

Characteristics of the Cohort

The cohort represents a 0.1 percent sampling of wage earners who in 1957 were working in jobs covered by the Social Security program.[1/] Workers were included in the cohort on the basis of two criteria: (1) In 1957, they were working in employment covered by the OASI program; and (2) when working in covered employment in any of the years in the 1951-57 period, they received wage credits[2/] solely from work as wage earners. Selection of the cohort on this retrospective basis guaranteed that the cohort members were alive in the period preceding 1957, and obviated the need for adjusting the wage data to take into account the incidence of mortality. The method of cohort selection, therefore, differs from the more usual method, and conceivably could affect the findings of this study. Our findings with respect to wage differentials and sex differentials, however, are consistent with the findings of other students of wage problems, and suggest that use of this method of cohort selection does not result in findings which would be appreciably different had a different mode of selection been used.

Wage histories of workers included in this study are classified by age in 1957, and by sex. In addition, they are grouped into four groups. Those whose average wage credits per year employed (PYE), 1951-57, were

(1) less than \$1,200, constitute the low-paid cohort
(2) \$1,200-\$2,399, constitute the intermediate low-paid cohort
(3) \$2,400-\$3,599, constitute the intermediate high-paid cohort
(4) \$3,600 or more, constitute the high-paid cohort

About two-thirds of the male cohort, and more than nine-tenths of the female cohort, had average wage credits of less than \$3,600 per year employed during the 1951-57 period. While this proportion varied with age, among males 40 years of age or older, the proportion of lower-paid workers was surprisingly high. For men aged 50-54, 46 percent had average wage credits per year in covered employment (PYE), 1951-57, of less than \$3,600, and for those aged 60-64, a little more than one-half the male cohort had average wage credits of less than \$3,600 PYE, 1951-57. For only one group in the female cohort--age 55-59--was the proportion of workers with credits of \$3,600 or more PYE, 1951-57, as high as 10 percent. Among women 25 to 69 years of age, the proportion with average credits of less than \$1,200 PYE, 1951-57, varied from 35 percent to 45 percent. About 80 percent of all the women are to be found in the two lowest-paid cohorts--i.e., those with average credits of less than \$2,400 PYE, 1951-57. (Tables 1 and 2.)

a/ The data contained in this paper were tabulated by the Social Security Administration in accordance with the author's research design. The interpretation of the data presented is the author's and not necessarily that of the Social Security Administration or of the U. S. Department of Labor.

1/ For a description of the sample, see The Continuous Work History Sample Under Old-Age and Survivors Insurance in the United States of America, by B. J. Mandel, First International Conference of Social Security Actuaries and Statisticians, Brussels, November 1956. Also see Jacob Perlman and Benjamin Mandel, "Sampling the Federal OASI Records," Journal of the American Statistical Association, September 1953.

2/ Wages credited to a worker for work in covered employment in a given year for purposes of benefit computations--up to \$3,600 for 1951-54, and \$4,200 for 1955-57.

Table 1: Number of Wage Earners in the Cohorts, by Sex, Age, and Average Wage Credits Per Year in Covered Employment a/

Age in 1957	Male: Average Wage Credits PYE, 1951-57 Total	Under $1,200	$1,200-$2,399	$2,400-$3,599	$3,600 and Over	Female: Average Wage Credits PYE, 1951-57 Total	Under $1,200	$1,200-$2,399	$2,400-$3,599	$3,600 and Over
Total	39,018	8,989	7,497	9,811	12,721	22,184	10,481	7,215	3,542	946
1-15......	324	320	3	1	---	150	148	2	---	---
16-19......	3,243	2,812	374	55	2	2,290	1,991	289	10	---
20-24......	5,096	2,472	1,882	716	26	3,040	1,658	1,194	188	---
25-29......	4,829	870	1,495	1,861	503	2,371	984	902	455	30
30-34......	4,775	461	769	1,634	1,911	2,397	1,090	751	472	84
35-39......	4,599	419	614	1,267	2,299	2,500	1,064	805	512	119
40-44......	3,957	339	486	1,017	2,115	2,390	953	829	463	145
45-49......	3,542	324	493	844	1,881	2,326	834	824	502	166
50-54......	2,887	275	358	694	1,560	1,788	630	622	382	154
55-59......	2,318	216	332	651	1,119	1,365	457	471	296	141
60-64......	1,731	148	280	465	838	848	325	293	158	72
65-69......	1,064	154	231	334	345	479	216	157	79	27
70 and over	629	174	168	169	118	234	128	73	25	8

a/ 0.1 percent sample of workers with credits solely from work as wage earners in 1957 and when employed any time in the 1951-57 period.

Table 2: Percentage Distribution of Male and Female Age Cohorts by Average Wage Credits Per Year in Covered Employment

Age in 1957	Male: Average Wage Credits PYE, 1951-57 Under $1,200	Under $2,400	Under $3,600	$3,600 and over	Female: Average Wage Credits PYE, 1951-57 Under $1,200	Under $2,400	Under $3,600	$3,600 and over
Total..	23.0	42.2	67.3	32.6	47.2	79.7	95.7	4.3
1-15........	98.8	99.7	100.0	---	98.7	100.0	100.0	---
16-19........	86.7	98.2	99.9	0.1	86.9	99.5	99.9	---
20-24........	48.5	85.4	99.5	0.5	54.5	93.8	100.0	---
25-29........	18.0	49.0	89.6	10.4	41.5	79.5	98.7	1.3
30-34........	9.7	25.8	60.0	40.0	45.5	76.8	96.5	3.5
35-39........	9.1	22.5	50.0	50.0	42.6	74.8	95.3	4.8
40-44........	8.6	20.9	46.6	53.4	39.9	76.6	94.0	6.1
45-49........	9.1	23.0	46.8	53.1	35.9	71.3	92.9	7.1
50-54........	9.5	21.9	45.9	54.0	35.2	70.0	91.4	8.6
55-59........	9.3	23.6	51.7	48.3	33.5	68.0	89.7	10.3
60-64........	8.5	24.7	51.6	48.4	38.3	72.9	91.5	8.5
65-69........	14.5	36.2	67.6	32.4	45.1	77.9	94.4	5.6
70 and over..	27.7	54.4	81.3	18.8	54.7	85.9	96.6	3.4

Average Credits As A Wage Measure

Average annual wage credits of the cohort, as a wage measure, differ substantially from average hourly earnings or average annual wages, the most frequently used measure of wages. For each year from 1951 to 1956, average credits of the cohort reflect both the employment and nonemployment of cohort members. Furthermore, the effects of nonemployment, or of an increase in the number of employed members, of the cohort on the average credits of the cohort can be measured from year to year. Differences in the direction and amount of year-to-year changes in employment levels, on the one hand, and the average credits of employed cohort members, on the other hand, can be isolated, and their relative effects on the average credits of the entire cohort studied.

Average hourly or annual earning data, however, are generally collected only for workers employed in a given time period, and year-to-year comparisons of these data--the usual cross-sectional method of analysis--can refer only to the wages of employed workers. For this reason, particularly when average hourly earnings in a prosperous year, for example, are compared with average hourly earnings in a recession year, the results obtained by such comparisons are difficult to interpret. Woytinsky objected to such cross-sectional wage comparisons as a means of approximating longitudinal analysis, warning that they

> are likely to be misleading in comparisons over time because of a change in the composition of the labor force; they may show, for instance, a rise of wages in depression because of a reduction in the number of less skilled and low-paid workers, whom it is customary to lay off first...3/

2. The Male Cohort

Male Low-Paid Cohort (Table 3)

The average annual wage credits of any of the cohorts are affected by changes in both the level of employment of the cohort and the level of wage credits of the employed members of the cohort. The indices represented in this figure permit us to determine the extent to which changes in either or both of these factors are responsible for the changes in the average annual credits of the members of the cohort.

Among all the members of the lowest-paid male cohort below age 65, average annual credits increased substantially during the 1951-57 period. Among the older and younger members of the cohort, however, there is a difference in the relationship between the extent of changes in employment and changes in annual wage credits of the employed cohort members. Among those who were 44 years of age or younger in 1957, the indices of average wage credits of the cohort increased at a faster rate than the indices of employment. For those 45 to 69 years of age, the indices of employment for most of the age groups tended to rise at a faster rate than the indices of the average annual credits of the cohort. Indeed, for workers who were 60 years of age or older the indices of annual wage credits of employed workers declined sharply below the 1951 level during the later years of the 1951-57 period.

As a result of the failure of the annual wage credits of the older employed males in the cohort to rise, the average wage credits of the entire older male cohort tended to increase to a lesser degree than the wage credits of the younger members of the cohort.

* * *

An important characteristic of the wage credits of this cohort is their sensitivity to cyclical change. Variations in the average annual wage credits of the employed members of the cohorts are generally of much lesser magnitude than the variations in the average credits of the entire cohort. This is due to the greater responsiveness of employment to economic change, as opposed to fluctuations of generally lesser magnitude in average annual wage credits of the employed workers in the cohort. From 1953 to the recession year of 1954, for example, average annual wage credits of employed workers in the cohort declined sharply. Indeed, among all workers except those aged 20-24 in 1957, they declined below the 1951 level. The indices of employment for cohort members through age 54 also declined during this period. As a result, for virtually all cohort members the index of average wage credits for 1954 fell not only below the 1953 level, but also below the 1951 level.

Male Intermediate Low-Paid Cohort (Table 4)

Among male members of the intermediate low-paid cohort, average credits of the cohort members below age 65 increased during the

3/ Encyclopedia of Social Sciences, p. 303. See also Staff Report on Employment, Growth, and Price Levels, Joint Economic Committee, U. S. Congress, December 24, 1959, which at page 142, on the basis of cross-sectional comparisons concludes that from 1947 to 1958 wages tended "to continue to move upward, even during periods of substantial unemployment."

Table 3: Male Low-Paid Cohort*: INDICES OF ANNUAL EMPLOYMENT AND OF AVERAGE ANNUAL CREDITS OF EMPLOYED MEMBERS AND OF ENTIRE COHORT, 1951-57

Age in 1957, No. Employed, and Average Credits	Number of workers and average credits							Index (1951 = 100.0)					
	1951	1952	1953	1954	1955	1956	1957	1952	1953	1954	1955	1956	1957
Total:	Total												
Employed 1/	2,740	3,216	3,698	3,852	4,890	5,861	8,989	117.4	135.0	140.6	178.5	213.9	328.1
Employed, $	576	553	564	516	593	679	871	96.0	97.9	89.6	103.0	117.9	151.2
Cohort, $	176	198	232	221	323	443	871	112.5	131.8	125.6	183.5	251.7	494.9
Age:	20-24												
Employed	825	1,214	1,499	1,463	1,549	1,572	2,472	147.2	181.7	177.3	187.8	190.5	299.6
Employed, $	301	395	495	515	651	814	1,193	131.2	164.5	171.1	216.3	270.4	396.3
Cohort, $	100	194	300	305	408	518	1,193	194.0	300.0	305.0	408.0	518.0	1193.0
Age:	25-29												
Employed	472	448	449	421	505	575	870	94.9	95.1	89.2	107.0	121.8	184.3
Employed, $	570	575	550	544	675	840	1,139	100.9	96.5	95.4	118.4	147.4	199.8
Cohort, $	309	296	284	263	392	555	1,139	95.8	91.9	85.1	126.9	179.6	368.6
Age:	30-34												
Employed	259	258	262	239	280	324	461	99.6	101.2	92.3	108.1	125.1	178.0
Employed, $	686	728	734	617	689	738	886	106.1	107.0	89.9	100.4	107.6	129.2
Cohort, $	385	407	417	320	418	519	886	105.7	108.3	83.1	108.6	134.8	230.1
Age:	35-39												
Employed	241	250	261	254	285	309	419	103.7	108.3	105.4	118.3	128.2	173.9
Employed, $	696	616	660	626	703	789	760	88.5	94.8	89.9	101.0	113.4	109.2
Cohort, $	400	368	411	379	478	582	760	92.0	102.8	94.8	119.5	145.5	190.0
Age:	40-44												
Employed	196	196	194	193	230	249	339	100.0	99.0	98.5	117.3	127.0	173.0
Employed, $	698	707	676	547	656	732	748	101.3	96.8	78.4	94.0	104.9	107.2
Cohort, $	404	409	387	311	445	538	748	101.2	95.8	77.0	110.1	133.2	185.1
Age:	45-49												
Employed	174	173	184	183	239	256	324	99.4	105.7	105.2	137.4	147.1	186.2
Employed, $	742	744	744	602	640	653	697	100.3	100.3	81.1	86.3	88.0	93.9
Cohort, $	398	397	423	340	472	516	697	99.7	106.3	85.4	118.6	129.6	175.1
Age:	50-54												
Employed	156	162	160	157	181	211	275	103.8	102.6	100.6	116.0	135.3	176.3
Employed, $	748	685	704	556	728	754	676	91.6	94.1	74.3	97.3	100.8	90.4
Cohort, $	424	404	410	317	479	579	676	95.3	96.7	74.8	113.0	136.6	159.4
Age:	55-59												
Employed	120	114	115	128	147	162	216	95.0	95.8	106.7	122.5	135.0	180.0
Employed, $	689	664	729	663	687	714	702	96.4	105.8	96.2	99.7	103.6	101.9
Cohort, $	383	350	388	393	468	536	702	91.4	101.3	102.6	122.2	139.9	183.3
Age:	60-64												
Employed	80	79	78	83	108	116	148	98.8	97.5	103.8	135.0	145.0	185.0
Employed, $	808	836	712	600	615	745	737	103.5	88.1	74.3	76.1	92.2	91.2
Cohort, $	437	446	375	336	449	584	737	102.1	85.8	76.9	102.7	133.6	168.6
Age:	65-69												
Employed	83	95	99	99	115	131	154	114.5	119.3	119.3	138.6	157.8	185.5
Employed, $	883	790	847	694	751	744	549	89.5	95.9	78.6	85.1	84.3	62.2
Cohort, $	476	487	544	446	561	633	549	102.3	114.3	93.7	117.9	133.0	115.3
Age:	70 and older												
Employed	104	100	103	107	125	142	174	96.2	99.0	102.9	120.2	136.5	167.3
Employed, $	1,014	840	716	606	633	679	520	82.8	70.6	59.8	62.4	67.0	51.3
Cohort, $	606	483	424	373	455	554	520	79.7	70.0	61.6	75.1	91.4	85.8

*Average credits of less than $1,200 PYE, 1951-57

1/ Detail will not add to total because ages of 24 cohort members were not known for specific years of employment.

Table 4: Male Intermediate Low-Paid Cohort*: INDICES OF ANNUAL EMPLOYMENT AND OF AVERAGE ANNUAL CREDITS OF EMPLOYED MEMBERS AND OF ENTIRE COHORT, 1951-57

Age in 1957, No. Employed, and Average Credit	Number of workers and average credits							Index (1951 = 100.0)					
	1951	1952	1953	1954	1955	1956	1957	1952	1953	1954	1955	1956	1957
Total:	Total												
Employed1/	4,692	4,997	5,243	5,247	5,909	6,507	7,497	106.5	111.7	111.8	125.9	138.7	159.8
Employed, $	1,351	1,475	1,593	1,545	1,777	2,179	2,411	109.2	117.9	114.4	131.5	161.3	178.5
Cohort, $	846	983	1,114	1,081	1,401	1,891	2,411	116.2	131.7	127.8	165.6	223.5	285.0
Age:	20-24												
Employed	826	1,100	1,233	1,200	1,368	1,554	1,882	133.2	149.3	145.3	165.6	188.1	227.8
Employed, $	630	926	1,183	1,322	1,782	2,444	2,743	147.0	187.8	209.8	282.9	387.9	435.4
Cohort, $	277	541	775	843	1,295	2,018	2,743	195.3	279.8	304.3	467.5	728.5	990.3
Age:	25-29												
Employed	990	905	886	922	1,106	1,235	1,495	91.4	89.5	93.1	111.7	124.7	151.0
Employed, $	1,053	1,201	1,341	1,417	1,856	2,392	2,775	114.1	127.4	134.6	176.3	227.2	263.5
Cohort, $	697	727	795	874	1,373	1,976	2,775	104.3	114.1	125.4	197.0	283.5	398.1
Age:	30-34												
Employed	570	575	603	582	631	666	769	100.9	105.8	102.1	110.7	116.8	134.9
Employed, $	1,461	1,632	1,761	1,602	1,774	2,159	2,388	111.7	120.5	109.7	121.4	147.8	163.4
Cohort, $	1,083	1,220	1,381	1,212	1,456	1,870	2,388	112.7	127.5	111.9	134.4	172.7	220.5
Age:	35-39												
Employed	470	469	485	478	512	541	614	99.8	103.2	101.7	108.9	115.1	130.6
Employed, $	1,489	1,689	1,805	1,670	1,895	2,184	2,297	113.4	121.2	112.2	127.3	146.7	154.3
Cohort, $	1,140	1,290	1,426	1,300	1,580	1,924	2,297	113.2	125.1	114.0	138.6	168.8	201.5
Age:	40-44												
Employed	355	387	399	392	418	431	486	109.0	112.4	110.4	117.7	121.4	136.9
Employed, $	1,657	1,718	1,798	1,672	1,755	2,048	2,175	103.7	108.5	100.9	105.9	123.6	131.3
Cohort, $	1,210	1,368	1,476	1,349	1,509	1,816	2,175	113.1	122.0	111.5	124.7	150.1	179.8
Age:	45-49												
Employed	375	392	413	408	434	461	493	104.5	110.1	108.8	115.7	122.9	131.5
Employed, $	1,690	1,801	1,845	1,700	1,797	2,070	2,128	106.6	109.2	100.6	106.3	122.5	125.9
Cohort, $	1,285	1,432	1,546	1,407	1,582	1,936	2,128	111.4	120.3	109.5	123.1	150.7	165.6
Age:	50-54												
Employed	279	303	309	304	314	346	358	108.6	110.8	109.0	112.5	124.0	128.3
Employed, $	1,773	1,734	1,802	1,711	1,939	1,961	2,024	97.8	101.6	96.5	109.4	110.6	114.2
Cohort, $	1,382	1,468	1,555	1,453	1,701	1,895	2,024	106.2	112.5	105.1	123.1	137.1	146.5
Age:	55-59												
Employed	270	276	280	284	301	315	332	102.2	103.7	105.2	111.5	116.7	123.0
Employed, $	1,697	1,839	1,891	1,773	1,815	2,016	2,009	108.4	111.4	104.5	107.0	118.8	118.4
Cohort, $	1,380	1,529	1,595	1,517	1,646	1,913	2,009	110.8	115.6	109.9	119.3	138.6	145.6
Age:	60-64												
Employed	231	242	244	244	257	267	280	104.8	105.6	105.6	111.3	115.6	121.2
Employed, $	1,734	1,757	1,896	1,784	1,892	2,008	1,900	101.3	109.3	102.9	109.1	115.8	109.6
Cohort, $	1,431	1,519	1,652	1,555	1,737	1,915	1,900	106.1	115.4	108.7	121.4	133.8	132.8
Age:	65-69												
Employed	182	187	199	203	215	218	231	102.7	109.3	111.5	118.1	119.8	126.9
Employed, $	1,927	2,072	2,172	1,894	1,687	1,722	1,396	107.5	112.7	98.3	87.5	89.4	72.4
Cohort, $	1,518	1,677	1,871	1,664	1,570	1,625	1,396	110.5	123.3	109.6	103.4	107.0	92.0
Age:	70 and older												
Employed	133	140	139	140	146	156	168	105.3	104.5	105.3	109.8	117.3	126.3
Employed, $	2,285	2,143	1,972	1,655	1,566	1,590	1,305	93.8	86.3	72.4	68.5	69.6	57.1
Cohort, $	1,809	1,786	1,632	1,379	1,361	1,476	1,305	98.7	90.2	76.2	75.2	81.6	72.1

*Average credits of $1,200 - $2,399 PYE, 1951-57

1/ Detail will not add to total because ages of 24 cohort members were not known for specific years of employment.

1951-57 period as a whole. For these age groups, the index indicates that the rate of increase varied inversely with the age of the workers--the older the age group, the lower was the rate of increase. During the 1954 recession, however, the average annual credits of all cohort members aged 30 or older in 1957 declined sharply. Primarily responsible for this decline was the fall in the average annual credits of employed cohort members, the indices of which were from 5 to 14 points lower in 1954 than in 1953. On the whole, the indices of employment for this cohort remained relatively stable from 1953 to 1954, declining by only 1 to 4 points among the 6 younger age groups, and rising slightly for the remaining age groups. Among workers 65 years of age or older, despite increasing employment, the average credits of the age groups declined, primarily as the result of substantial declines in the average annual wage credits of the employed cohort members.

Male Intermediate High-Paid Cohort (Table 5)

The average annual wage credits of the male members of this cohort tended to rise throughout the 1951-57 period as a result of rising employment and rising annual wage credits of the employed members of the cohort. Of particular interest, however, is the change which occurred from 1956 to 1957. Among those younger than age 50, average annual credits moved upward from 1956 to 1957. Among those aged 50 or older in 1957, average annual wage credits declined from 1956 to 1957.

Among workers below age 50, average credits of the cohort members increased from 1956 to 1957--despite a decline in the average annual wage credits of the employed members of the cohort--because of a sharp increase in employment. Among the older age groups, however, the indices of employment rose by only an insignificant amount. As a result, the decline in the average annual wage credits of employed cohort members was sufficiently large to reduce average wage credits of these cohort members to levels below those prevailing in 1956.

Differences between changes in the annual wage credits of employed cohort members and the level of employment also help explain the differing effects of the 1954 recession on the average annual wage credits of the younger and older cohort members. For employed workers below the age of 40, average annual wage credits increased or remained unchanged from 1953 to 1954. Employment among these age groups rose from 1953 to 1954. Thus, average annual wage credits of these younger cohort members during the 1954 recession were higher than they had been in the preceding year.

Among cohort members aged 40 to 65 in 1957, employment increased moderately from 1953 to 1954. Average annual wage credits of the older employed cohort members, however, declined to a more than compensating degree, causing a decline in the average annual wage credits of these cohort members.

Male High-Paid Cohort (Table 6)

For the high-paid male cohort, average annual wage credits of employed cohort members for all age groups below 65 remained stable throughout the 1951-57 period. (The sharp rise in the wage credit indices from 1954 to 1955 in large part reflects the effects of the change in the maximum limit of wages subject to social security taxes.) During 1951-57, employment also tended to rise slightly. As a result, the indices indicate, average wage credits of the cohort members rose at a somewhat faster pace than the average annual wage credits of the employed cohort members.

Unlike the lower-paid cohorts, among members of this cohort the rate of increase in wage credits during the 1954 recession, as measured by the indices, was slowed somewhat for 6 of the 10 age groups--i.e., older workers 35 to 64 years of age in 1957.

Differentials In Wage Credits - Male Cohorts (Tables 7 and 8)

The data indicate that during the 1951-57 period there was a general tendency for the differentials between the average annual wage credits of employed workers and of the members of the highest and the two intermediate-paid cohorts to narrow. Differentials among those aged 55 or older widened somewhat in 1955, when the maximum taxable limit on wages subject to social security taxes was increased to $4,200. The tendency toward a narrowing of differentials, however, resumed during the following 2 years. Among those younger than 55, differentials generally narrowed from 1951-57, primarily as a result of the relatively greater increases in employment of the younger workers as compared to the older workers.

The 1954 recession, as might be expected, widened differentials in average annual wage credits of employed members of the cohorts and in annual wage credits of the cohorts as the level of employment of the lower-paid cohorts declined.

Among male cohort members under the age of 30, the annual changes in differentials probably reflect to some degree the effects of military service on the employment and wage credits of the age group during the Korean War. The wage

Table 5: Male Intermediate High-Paid Cohort*: INDICES OF ANNUAL EMPLOYMENT AND OF AVERAGE ANNUAL CREDITS OF EMPLOYED MEMBERS AND OF ENTIRE COHORT, 1951-57

Age in 1957, No. Employed, and Average Credits	Number of workers and average credits							Index (1951 = 100.0)					
	1951	1952	1953	1954	1955	1956	1957	1952	1953	1954	1955	1956	1957
Total:	Total												
Employed1/	7,643	7,846	8,267	8,504	9,033	9,318	9,811	102.8	108.2	111.3	118.2	121.9	128.4
Employed, $	2,351	2,628	2,910	2,935	3,355	3,566	3,499	111.8	123.8	124.8	142.7	151.7	148.8
Cohort, $	1,831	2,104	2,452	2,544	3,089	3,387	3,499	114.9	133.9	138.9	168.7	185.0	191.1
Age:	20-24												
Employed	353	441	512	523	636	689	716	124.9	145.0	148.2	180.2	195.2	202.8
Employed, $	1,080	1,736	2,257	2,643	3,251	3,682	3,747	160.7	209.0	244.7	301.0	340.9	346.9
Cohort, $	532	1,069	1,614	1,931	2,888	3,543	3,747	200.9	303.4	363.0	542.9	666.0	704.3
Age:	25-29												
Employed	1,294	1,258	1,433	1,613	1,816	1,858	1,861	97.2	110.7	124.7	140.3	143.6	143.8
Employed, $	1,720	2,012	2,563	2,821	3,460	3,806	3,856	117.0	149.0	164.0	201.2	221.3	224.2
Cohort, $	1,135	1,291	1,873	2,320	3,204	3,606	3,856	113.7	165.0	204.4	282.3	317.7	339.7
Age:	30-34												
Employed	1,265	1,332	1,394	1,408	1,461	1,488	1,634	105.3	110.2	111.3	115.5	117.6	129.2
Employed, $	2,174	2,575	2,992	2,989	3,473	3,683	3,637	118.4	137.6	137.5	159.8	169.4	167.3
Cohort, $	1,683	2,099	2,553	2,576	3,105	3,354	3,637	124.7	151.7	153.1	184.5	199.3	216.1
Age:	35-39												
Employed	1,023	1,045	1,085	1,100	1,140	1,166	1,267	102.2	106.1	107.5	111.4	114.0	123.9
Employed, $	2,405	2,759	2,997	2,995	3,476	3,616	3,577	114.7	124.6	124.5	144.5	150.4	148.7
Cohort, $	1,942	2,276	2,566	2,600	3,128	3,328	3,577	117.2	132.1	133.9	161.1	171.4	184.2
Age:	40-44												
Employed	863	885	910	911	948	967	1,017	102.5	105.4	105.6	109.8	112.1	117.8
Employed, $	2,532	2,814	3,018	2,948	3,294	3,522	3,477	111.1	119.2	116.4	130.1	139.1	137.3
Cohort, $	2,149	2,449	2,700	2,641	3,071	3,349	3,477	114.0	125.6	122.9	142.9	155.8	161.8
Age:	45-49												
Employed	730	752	771	777	805	825	844	103.0	105.6	106.4	110.3	113.0	115.6
Employed, $	2,670	2,830	3,038	2,928	3,252	3,412	3,365	106.0	113.8	109.7	121.8	127.8	126.0
Cohort, $	2,309	2,522	2,775	2,696	3,102	3,335	3,365	109.2	120.2	116.8	134.3	144.4	145.7
Age:	50-54												
Employed	623	629	642	648	661	684	694	101.0	103.0	104.0	106.1	109.8	111.4
Employed, $	2,717	2,909	3,030	2,978	3,219	3,362	3,290	107.1	111.5	109.6	118.5	123.7	121.1
Cohort, $	2,439	2,637	2,803	2,781	3,066	3,314	3,290	108.1	114.9	114.0	125.7	135.9	134.9
Age:	55-59												
Employed	595	602	603	604	614	638	651	101.2	101.3	101.5	103.2	107.2	109.4
Employed, $	2,835	2,980	3,111	2,970	3,195	3,361	3,212	105.1	109.7	104.8	112.7	118.6	113.3
Cohort, $	2,591	2,756	2,882	2,756	3,013	3,294	3,212	106.4	111.2	106.4	116.3	127.1	124.0
Age:	60-64												
Employed	434	439	439	438	441	460	465	101.2	101.2	100.9	101.6	106.0	107.1
Employed, $	2,874	3,017	3,124	3,002	3,259	3,384	3,181	105.0	108.7	104.5	113.4	117.7	110.7
Cohort, $	2,682	2,848	2,949	2,828	3,091	3,348	3,181	106.2	110.0	105.4	115.2	124.8	118.6
Age:	65-69												
Employed	309	314	319	317	319	324	334	101.6	103.2	102.6	103.2	104.9	108.1
Employed, $	2,986	3,118	3,185	3,188	3,356	3,226	2,296	104.4	106.7	106.8	112.4	108.0	76.9
Cohort, $	2,762	2,931	3,042	3,026	3,205	3,129	2,296	106.1	110.1	109.6	116.0	113.3	83.1
Age:	70 and older												
Employed	150	155	155	157	163	165	169	103.3	103.3	104.7	108.7	110.0	112.7
Employed, $	3,028	3,101	3,210	3,204	3,253	3,078	2,385	102.4	106.0	105.8	107.4	101.7	78.8
Cohort, $	2,688	2,844	2,944	2,976	3,138	3,005	2,385	105.8	109.5	110.7	116.7	111.8	88.7

*Average credits of $2,400 - $3,599 PYE, 1951-57

1/ Detail will not add to total because ages of 24 cohort members were not known for specific years of employment.

Table 6: Male High-Paid Cohort*: INDICES OF ANNUAL EMPLOYMENT AND OF AVERAGE ANNUAL CREDITS OF EMPLOYED MEMBERS AND OF ENTIRE COHORT, 1951-57

Age in 1957, No. Employed, and Average Credits	Number of workers and average credits							Index (1951 = 100.0)					
	1951	1952	1953	1954	1955	1956	1957	1952	1953	1954	1955	1956	1957
Total:	Total												
Employed 1/	11844	11959	12056	12117	12273	12443	12721	101.0	101.8	102.3	103.6	105.1	107.4
Employed, $	3,499	3,558	3,582	3,577	4,154	4,170	4,169	101.7	102.4	102.2	118.7	119.2	119.1
Cohort, $	3,258	3,345	3,395	3,407	4,008	4,079	4,169	102.7	104.2	104.6	123.0	125.2	128.0
Age:	20-24												
Employed	9	13	14	17	18	23	26	144.4	155.6	188.9	200.0	255.6	288.9
Employed, $	2,564	3,121	3,490	3,585	4,083	4,098	4,017	121.7	136.1	139.8	159.2	159.8	156.7
Cohort, $	888	1,560	1,879	2,344	2,827	3,625	4,017	175.7	211.6	264.0	318.4	408.2	452.4
Age:	25-29												
Employed	353	378	413	435	457	470	503	107.1	117.0	123.2	129.5	133.1	142.5
Employed, $	3,250	3,457	3,530	3,534	4,124	4,159	4,168	106.4	108.6	108.7	126.9	128.0	128.2
Cohort, $	2,281	2,598	2,898	3,056	3,747	3,886	4,168	113.9	127.0	134.0	164.3	170.4	182.7
Age:	30-34												
Employed	1,752	1,789	1,809	1,826	1,843	1,857	1,911	102.1	103.3	104.2	105.2	106.0	109.1
Employed, $	3,399	3,537	3,573	3,571	4,161	4,173	4,178	104.1	105.1	105.1	122.4	122.8	122.9
Cohort, $	3,116	3,311	3,382	3,412	4,013	4,055	4,178	106.3	108.5	109.5	128.8	130.1	134.1
Age:	35-39												
Employed	2,140	2,159	2,173	2,182	2,202	2,218	2,299	100.9	101.5	102.0	102.9	103.6	107.4
Employed, $	3,487	3,551	3,582	3,577	4,159	4,173	4,171	101.8	102.7	102.6	119.3	119.7	119.6
Cohort, $	3,246	3,335	3,386	3,395	3,984	4,026	4,171	102.7	104.3	104.6	122.7	124.0	128.5
Age:	40-44												
Employed	1,995	2,002	2,012	2,017	2,040	2,055	2,115	100.4	100.9	101.1	102.3	103.0	106.0
Employed, $	3,520	3,566	3,584	3,578	4,158	4,173	4,181	101.3	101.8	101.6	118.1	118.6	118.8
Cohort, $	3,320	3,375	3,409	3,412	4,011	4,055	4,181	101.7	102.7	102.8	120.8	122.1	125.9
Age:	45-49												
Employed	1,807	1,816	1,824	1,826	1,840	1,860	1,881	100.5	100.9	101.1	101.8	102.9	104.1
Employed, $	3,543	3,570	3,587	3,581	4,158	4,174	4,175	100.8	101.2	101.1	117.4	117.8	117.8
Cohort, $	3,404	3,447	3,478	3,476	4,067	4,127	4,175	101.3	102.2	102.1	119.5	121.2	122.6
Age:	50-54												
Employed	1,489	1,493	1,496	1,497	1,521	1,548	1,560	100.3	100.5	100.5	102.1	104.0	104.8
Employed, $	3,544	3,576	3,588	3,582	4,149	4,171	4,169	100.9	101.2	101.1	117.1	117.7	117.6
Cohort, $	3,383	3,422	3,441	3,437	4,045	4,139	4,169	101.2	101.7	101.6	119.6	122.3	123.2
Age:	55-59												
Employed	1,075	1,080	1,082	1,082	1,099	1,115	1,119	100.5	100.7	100.7	102.2	103.7	104.1
Employed, $	3,544	3,576	3,591	3,587	4,155	4,172	4,163	100.9	101.3	101.2	117.2	117.7	117.5
Cohort, $	3,405	3,451	3,472	3,468	4,081	4,157	4,163	101.4	102.0	101.8	119.9	122.1	122.3
Age:	60-64												
Employed	789	792	794	795	807	831	838	100.4	100.6	100.7	102.3	105.3	106.2
Employed, $	3,545	3,566	3,584	3,577	4,146	4,155	4,156	100.6	101.1	100.9	117.0	117.2	117.2
Cohort, $	3,338	3,370	3,396	3,393	3,993	4,120	4,156	101.0	101.7	101.6	119.6	123.4	124.5
Age:	65-69												
Employed	325	328	328	329	334	343	345	100.9	100.9	101.2	102.8	105.5	106.2
Employed, $	3,549	3,569	3,584	3,589	4,126	4,158	4,071	100.6	101.0	101.1	116.3	117.2	114.7
Cohort, $	3,343	3,393	3,407	3,423	3,994	4,134	4,071	101.5	101.9	102.4	119.5	123.7	121.8
Age:	70 and older												
Employed	107	106	107	107	108	117	118	99.1	100.0	100.0	100.9	109.3	110.3
Employed, $	3,531	3,577	3,578	3,599	4,149	4,103	4,093	101.3	101.3	101.9	117.5	116.2	115.9
Cohort, $	3,202	3,213	3,244	3,264	3,797	4,068	4,093	100.3	101.3	101.9	118.6	127.0	127.8

*Average credits of $3,600 or more PYE, 1951-57

1/ Detail will not add to total because ages of 24 cohort members were not known for specific years of employment.

Table 7: Male Cohorts: WAGE CREDIT DIFFERENTIALS OF EMPLOYED COHORT MEMBERS
(Average Annual Credits of Employed Members of Lower-Paid Cohorts as a Percentage of Average Annual Credits of Employed Members of Highest-Paid Cohort, 1951-57)

Age in 1957 and subcohort		1951	1952	1953	1954	1955	1956	1957
Total	3 lower-paid cohorts.....	49.1	52.1	55.9	55.7	53.1	56.8	54.9
	Low-paid[a]/..............	16.5	15.5	15.7	14.4	14.3	16.3	20.9
	Intermediate low-paid[b]/..	38.6	41.5	44.5	43.2	42.8	52.3	57.8
	Intermediate high-paid[c]/.	67.2	73.9	81.2	82.1	80.8	85.5	83.9
16-19	3 lower-paid cohorts.....	----	----	----	----	----	17.8	22.6
	Low-paid[a]/..............	----	----	----	----	----	12.3	16.7
	Intermediate low-paid[b]/..	----	----	----	----	----	40.7	57.7
	Intermediate high-paid[c]/.	----	----	----	----	----	69.0	86.6
20-24	3 lower-paid cohorts.....	22.4	26.4	29.7	32.6	38.0	48.7	53.0
	Low-paid[a]/..............	11.7	12.7	14.2	14.4	15.9	19.9	29.7
	Intermediate low-paid[b]/..	24.6	29.7	33.9	36.9	43.6	59.6	68.3
	Intermediate high-paid[c]/.	42.1	55.6	64.7	73.7	79.6	89.8	93.3
25-29	3 lower-paid cohorts.....	39.5	42.9	52.2	58.2	61.4	68.9	69.9
	Low-paid[a]/..............	17.5	16.6	15.6	15.4	16.4	20.2	27.3
	Intermediate low-paid[b]/..	32.4	34.7	38.0	40.1	45.0	57.5	66.6
	Intermediate high-paid[c]/	52.9	58.2	72.6	79.8	83.9	91.5	92.5
30-34	3 lower-paid cohorts.....	52.9	59.5	67.2	66.4	64.7	69.2	68.5
	Low-paid[a]/..............	20.2	20.6	20.5	17.3	16.6	17.7	21.2
	Intermediate low-paid[b]/	43.0	46.1	49.3	44.9	42.6	51.7	57.2
	Intermediate high-paid[c]/.	64.0	72.8	83.7	83.7	83.5	88.3	87.1
35-39	3 lower-paid cohorts.....	55.1	61.1	65.6	64.9	63.7	67.1	65.3
	Low-paid[a]/..............	20.0	17.3	18.4	17.5	16.9	18.9	18.2
	Intermediate low-paid[b]/	42.7	47.6	50.4	46.7	45.6	52.3	55.1
	Intermediate high-paid[c]/	69.0	77.7	83.7	83.7	83.6	86.7	85.8
40-44	3 lower-paid cohorts.....	58.5	62.9	66.7	64.4	60.4	65.0	63.0
	Low-paid[a]/..............	19.8	19.8	18.9	15.3	15.8	17.5	17.9
	Intermediate low-paid[b]/..	47.1	48.2	50.2	46.7	42.2	49.1	52.0
	Intermediate high-paid[c]/.	71.9	78.9	84.2	82.4	79.2	84.4	83.2
45-49	3 lower-paid cohorts.....	59.9	63.0	66.0	62.9	57.8	61.1	59.3
	Low-paid[a]/..............	20.9	20.8	20.7	16.8	15.4	15.6	16.7
	Intermediate low-paid[b]/	47.7	50.4	51.4	47.5	43.2	49.6	51.0
	Intermediate high-paid[c]/.	75.4	79.3	84.7	81.8	78.2	81.7	80.6
50-54	3 lower-paid cohorts.....	61.5	63.0	65.6	63.9	59.8	60.6	57.7
	Low-paid[a]/..............	21.1	19.2	19.6	15.5	17.5	18.1	16.2
	Intermediate low-paid[b]/..	50.0	48.5	50.2	47.8	46.7	47.0	48.5
	Intermediate high-paid[c]/.	76.7	81.3	84.4	83.1	77.6	80.6	78.9
55-59	3 lower-paid cohorts.....	63.8	67.0	69.4	65.4	59.1	62.2	58.3
	Low-paid[a]/..............	19.4	18.6	20.3	18.5	16.5	17.1	16.9
	Intermediate low-paid[b]/..	47.9	51.4	52.7	49.4	43.7	48.3	48.3
	Intermediate high-paid[c]/.	80.0	83.3	86.6	82.8	76.9	80.6	77.2
60-64	3 lower-paid cohorts.....	64.9	67.0	69.3	65.8	59.5	62.2	57.1
	Low-paid[a]/..............	22.8	23.4	19.9	16.8	14.8	17.9	17.7
	Intermediate low-paid[b]/..	48.9	49.3	52.9	49.9	45.6	48.3	45.7
	Intermediate high-paid[c]/.	81.1	84.6	87.2	83.9	78.6	81.4	76.5
65-69	3 lower-paid cohorts.....	66.1	67.8	69.3	65.9	56.7	54.3	40.1
	Low-paid[a]/..............	24.9	22.1	23.6	19.3	18.2	17.9	13.5
	Intermediate low-paid[b]/..	54.3	58.1	60.6	52.8	40.9	41.4	34.3
	Intermediate high-paid[c]/.	84.1	87.4	88.9	88.8	81.3	77.6	56.4
70 and over	3 lower-paid cohorts.....	63.2	61.2	59.5	55.0	46.5	44.9	34.1
	Low-paid[a]/..............	28.7	23.5	20.0	16.8	15.3	16.5	12.7
	Intermediate low-paid[b]/..	64.7	59.9	55.1	46.0	37.7	38.8	31.9
	Intermediate high-paid[c]/.	85.8	86.7	89.7	89.0	78.4	75.0	58.3

Workers with average credits PYE, 1951-57, of:

a/ Less than $1,200

b/ $1,200 to $2,399

c/ $2,400 to $3,599

Table 8: Male Cohorts: WAGE CREDIT DIFFERENTIALS OF COHORT MEMBERS
(Average Annual Credits of Members of Lower-Paid Cohorts as a Percentage of Average Annual Credits of Members of Highest-Paid Cohort, 1951-57)

Age in 1957 and subcohort		1951	1952	1953	1954	1955	1956	1957
Total	3 lower-paid cohorts.....	37.6	41.1	46.1	47.0	47.5	52.5	54.9
	Low-paid[a/]...............	5.4	5.9	6.8	6.5	8.1	10.9	20.9
	Intermediate low-paid[b/]..	26.0	29.4	32.8	31.7	35.0	46.4	57.8
	Intermediate high-paid[c/].	56.2	62.9	72.2	74.7	77.1	83.0	83.9
16-19	3 lower-paid cohorts.....	----	----	----	----	----	12.6	22.6
	Low-paid[a/]...............	----	----	----	----	----	7.5	16.7
	Intermediate low-paid[b/]..	----	----	----	----	----	33.2	57.7
	Intermediate high-paid[c/].	----	----	----	----	----	62.8	86.6
20-24	3 lower-paid cohorts.....	28.1	30.3	36.6	33.1	42.2	46.2	53.0
	Low-paid[a/]...............	11.3	12.4	16.0	13.0	14.4	14.3	29.7
	Intermediate low-paid[b/]..	31.2	34.7	41.2	36.0	45.8	55.7	68.3
	Intermediate high-paid[c/].	59.9	68.5	85.9	82.4	102.2	97.7	93.3
25-29	3 lower-paid cohorts.....	36.7	35.6	43.8	51.6	58.7	66.4	69.9
	Low-paid[a/]...............	13.5	11.4	9.8	8.6	10.5	14.3	27.3
	Intermediate low-paid[b/]..	30.6	28.0	27.4	28.6	36.6	50.8	66.6
	Intermediate high-paid[c/].	49.8	49.7	64.6	75.9	85.5	92.8	92.5
30-34	3 lower-paid cohorts.....	43.6	50.2	58.9	58.0	58.6	63.7	68.5
	Low-paid[a/]...............	12.4	12.3	12.3	9.4	10.4	12.8	21.2
	Intermediate low-paid[b/]..	34.8	36.8	40.8	35.5	36.3	46.1	57.2
	Intermediate high-paid[c/].	54.0	63.4	75.5	75.5	77.4	82.7	87.1
35-39	3 lower-paid cohorts.....	46.5	52.3	57.8	57.5	58.5	62.9	65.3
	Low-paid[a/]...............	12.3	11.0	12.1	11.2	12.0	14.5	18.2
	Intermediate low-paid[b/]..	35.1	38.7	42.1	38.3	39.7	47.8	55.1
	Intermediate high-paid[c/].	59.8	68.2	75.8	76.6	78.5	82.7	85.8
40-44	3 lower-paid cohorts.....	50.3	56.1	60.9	58.7	56.9	62.2	63.0
	Low-paid[a/]...............	12.2	12.1	11.4	9.1	11.1	13.3	17.9
	Intermediate low-paid[b/]..	36.4	40.5	43.3	39.5	37.6	44.8	52.0
	Intermediate high-paid[c/].	64.7	72.6	79.2	77.4	76.6	82.6	83.2
45-49	3 lower-paid cohorts.....	51.4	55.7	60.1	57.5	54.9	59.3	59.3
	Low-paid[a/]...............	11.7	11.5	12.2	9.8	11.6	12.5	16.7
	Intermediate low-paid[b/]..	37.7	41.5	44.5	40.5	38.9	46.9	51.0
	Intermediate high-paid[c/].	67.8	73.2	79.8	77.6	76.3	80.8	80.6
50-54	3 lower-paid cohorts.....	55.1	58.0	61.4	60.2	56.6	59.3	57.7
	Low-paid[a/]...............	12.5	11.8	11.9	9.2	11.8	14.0	16.2
	Intermediate low-paid[b/]..	40.9	42.9	45.2	42.3	42.1	45.8	48.5
	Intermediate high-paid[c/].	72.1	77.1	81.5	80.9	75.8	80.1	78.9
55-59	3 lower-paid cohorts.....	58.4	62.0	64.3	60.9	55.7	60.2	58.3
	Low-paid[a/]...............	11.2	101.	11.2	11.3	11.5	12.9	16.9
	Intermediate low-paid[b/]..	40.5	44.3	45.9	43.7	40.3	46.0	48.3
	Intermediate high-paid[c/].	76.1	79.9	83.0	79.5	73.8	79.2	77.2
60-64	3 lower-paid cohorts.....	61.5	64.5	66.8	63.4	57.7	61.0	57.1
	Low-paid[a/]...............	13.1	13.2	11.0	9.9	11.2	14.2	17.7
	Intermediate low-paid[b/]..	42.9	45.1	48.6	45.8	43.5	46.5	45.7
	Intermediate high-paid[c/].	80.3	84.5	86.8	83.3	77.4	81.3	76.5
65-69	3 lower-paid cohorts.....	60.9	63.3	66.4	63.3	54.9	52.2	40.1
	Low-paid[a/]...............	14.2	14.4	16.0	13.0	14.0	15.3	13.5
	Intermediate low-paid[b/]..	45.4	49.4	54.9	48.6	39.3	39.3	34.3
	Intermediate high-paid[c/].	82.6	86.4	89.3	88.4	80.2	75.7	56.4
70 and over	3 lower-paid cohorts.....	57.0	58.2	56.5	53.1	46.5	42.7	34.1
	Low-paid[a/]...............	18.9	15.0	13.1	11.4	12.0	13.6	12.7
	Intermediate low-paid[b/]..	56.5	55.6	50.3	42.2	35.8	36.3	31.9
	Intermediate high-paid[c/].	83.9	88.5	90.8	91.2	82.6	73.9	58.3

Workers with average credits PYE, 1951-57, of:

a/ Less than $1,200

b/ $1,200 to $2,399

c/ $2,400 to $3,599

credit differentials for these younger workers, however, tended to follow the same patterns as those of the older workers.

Summary - The Male Cohort, 1951-57

Cyclical changes have a differing impact on the annual wage credits of male workers, depending upon their age and their average level of wage credits. In general, there is an inverse relationship between the age of workers and the extent to which their wage credits vary from year to year; the older the worker, the less the year-to-year variation in average annual wage credits.

The extent of fluctuation varies inversely with the average level of wage credits; the wage credits of the lower-paid cohorts are subject to much greater variation than are the wage credits of the higher-paid cohorts.

Changes in the average credits of the cohorts from 1953 to 1954 offer an interesting illustration of the differential effects of the 1954 recession on wage credits. Among the lowest-paid male cohort, the average credits of cohort members aged 24 or younger were greater in 1954 than in 1953. Among cohort members who were 25 years of age or older, however, with the exception of one age group, average credits in 1954 were lower than in 1953.

Among the intermediate low-paid male cohort, on the other hand, the average credits of workers who were 29 years of age or younger were higher in 1954 than in 1953; for those 30 years of age or older, average credits in 1954 were lower than in 1953.

In the case of the intermediate high-paid cohort, however, the average credits of cohort members who were 39 years of age or younger were higher in 1954 than in 1953; the average credits of the older cohort members declined from 1953 to 1954.

For the highest-paid cohort, the "turning" age was 45. The average wage credits of those cohort members who were younger than 45 years of age increased from 1953 to 1954; for those aged 45-64, average wage credits in 1954 were somewhat below 1953 levels.

Varying degrees of nonemployment among the age-wage cohorts substantially affected the level of average wage credits in each of the years, 1951-57. Among the highest-paid male cohorts aged 50-54, for example, the average annual wage credits of the employed cohort members in 1955 were approximately $4,150; for the cohort, the average was approximately $100 less, $4,050. Among the lowest-paid cohort in this age group, the difference between the average wage credits of the employed members and all the cohort members was even greater. For the employed members, average credits in 1955 were about $730; for all the cohort members, the average was approximately $480, a difference of about $250. In general, among all the cohorts, because of the greater intermittency in employment among the lower-paid workers, the disparity between the average wage credits of the employed members of the cohort and of all the members of the cohort was greatest among the lower-paid workers, and was least significant among the higher-paid cohorts.

Among the older workers in the higher-paid cohorts there was a perceptible tendency for nonemployment to affect average wage credits. In 1955, for example, among the 60-64 age group in the highest-paid cohort, average annual credits of the cohort were approximately $150 lower than the average annual credits of the employed members of the cohort. In that year, the difference between these two averages for the 55-59 group was $74, and for the 50-54 year group, $104. Among the younger members of the cohort, the difference between the average credits of the employed members and of all the cohort members of the age group was also relatively large. Among the highest-paid cohort members, therefore, nonemployment tended to affect the wage credits of workers at the youngest and oldest ages; only during the middle years of life did nonemployment have relatively little effect on average annual wage credits. In general, therefore, nonemployment adversely affected the wage credits of the cohort members inversely with the level of wage credits, 1951-57, i.e., the lower the level of average wage credits PYE, 1951-57, the greater was the effect of nonemployment on average wage credits within a given year.

Another general tendency during the 1951-57 period was the influence of increased employment on the average annual wage credits of all the cohort members. Among all the cohort members the increased employment during the 1951-57 period resulted in greater increases in the average wage credits of the cohort members than in the average credits of the employed members of the cohort.

Especially worthy of note are the data for 1953 and 1954. The 1954 recession had significant effects on the average annual credits of the employed members of the three lower-paid cohorts and on the average annual credits of all the members of these cohorts. For workers aged 30 or older in 1957, the 1954 recession resulted in significant declines in the average annual credits of the cohort members and in the average annual credits of the employed members of these cohorts. For the highest-paid cohort, however, the effects of the 1954 recession on wage credits were virtually nil.

Among the very oldest and very youngest members of the cohort, however, the effects of age appear to be at least as significant as those of cyclical change in terms of their effects on wage credits. Among workers below the age of 30, average wage credits in each of the years, 1951-57, tended to rise among members of all of the wage cohorts. Among cohort members 65 years of age or older, however, there was a tendency for average wage credits to fall during the last 3 or 4 years of the 1951-57 period, a tendency that was particularly prevalent among the three lower-paid cohorts.

3. The Female Cohort

Female Low-Paid Cohort (Table 9)

For members of this female cohort below the age of 65 in 1957, employment rates increased greatly from 1951 to 1957. Despite only relatively modest increases in the indices of average annual wage credits of employed cohort members, the sharp rise in employment resulted in a substantial increase in average wage credits of the cohort.

From 1956 to 1957, employment among women aged 65 or older, however, did not increase at a sufficiently rapid rate to overcome the substantial fall in annual wage credits of the employed cohort members. As a consequence, for these oldest age groups the index of average wage credits declined from 1956 to 1957.

Unlike the low-paid male cohort, the low-paid female cohort generally suffered no declines in average annual wage credits during the 1954 recession. Although the indices of average annual wage credits of the employed members declined somewhat during the 1954 recession, principally among those aged 30 to 44, in general the increase in the employment index more than offset the declines in the index of average wage credits of the employed members of the cohort. The index of average annual wage credits for most of the age groups in the cohort, therefore, rose from 1953 to 1954. Among those aged 25 to 34 and 40 to 44, however, the index of average wage credits declined somewhat from 1953 to 1954.

Female Intermediate Low-Paid Cohort (Table 10)

The index of average annual wage credits of the members of this female cohort rose sharply during the 1951-57 period. In general, the rate of increase tended to rise with increasing age to a peak at age 35 to 39. The rate of increase varied inversely with increasing age among older members, reaching its lowest point among cohort members age 65 or older.

Among the cohort members, the index of average annual wage credits of the employed women aged 40 to 59 tended to rise at a somewhat faster pace than the index of employment. As a consequence, the index of average annual wage credits of these women rose substantially from 1951 to 1957. Among those in the younger age groups, the indices of employment and of annual wage credits of the employed women tended to rise at different rates. In the case of both indices, however, these rates increased more rapidly than those of the older women and the average annual credit index increased more rapidly for the younger than the older groups.

The effects of the 1954 recession on the credits of this cohort--as compared to the lowest-paid female cohort--were not substantial. The indices of employment of this cohort generally rose from 1953 to 1954. In most instances, also, the decline in the indices of annual wage credits of women employed in 1954 was not as severe in this cohort as among their lower-paid sisters. For this reason, the substantial rise in the indices of employment resulted in a significant increase in the indices of average annual wage credits of this cohort.

Female Intermediate High-Paid Cohort (Table 11)

The indices of average annual wage credits of this female cohort rose at a faster rate than did the two component indices of employment and average annual wage credits of the employed cohort members. The indices of average annual wage credits for the cohort rose uninterruptedly, each year, 1951-57. The 1954 recession did not affect the indices of employment, wage credits of employed cohort members, or average wage credits of this female cohort.

From 1956 to 1957, the index of the average credits of employed women aged 25 to 29 declined somewhat, for reasons which are not readily apparent. The rise in the index of employment was not sufficient to offset the decline in the index of average annual wage credits of the employed members, and as a consequence, for this age group, the index of the average annual wage credits of the cohort for 1956 was below the 1957 level. It is possible that the fall in the index may reflect a decline in the proportion of full-year workers in this age group, since the index of employment for the age group rose. For those women 65 years of age or older, the indices of employment, average annual wage credits of

Table 9: Female Low-Paid Cohort*: INDICES OF ANNUAL EMPLOYMENT AND OF AVERAGE ANNUAL CREDITS OF EMPLOYED MEMBERS AND OF ENTIRE COHORT, 1951-57

Age in 1957, No. Employed, and Average Credits	Number of workers and average credits							Index (1951 = 100.0)					
	1951	1952	1953	1954	1955	1956	1957	1952	1953	1954	1955	1956	1957
Total:	Total												
Employed1/	3,419	3,854	4,417	4,644	5,909	7,440	10481	112.7	129.2	135.8	172.8	217.6	306.6
Employed, $	589	585	598	593	616	711	728	99.3	101.5	100.7	104.6	120.7	123.6
Cohort, $	192	215	252	263	347	505	728	112.0	131.2	137.0	180.7	263.0	379.2
Age:	20-24												
Employed	433	667	913	949	1,085	1,260	1,658	Not computed					
Employed, $	296	385	482	550	654	866	866						
Cohort, $	77	155	265	315	428	658	866						
Age:	25-29												
Employed	517	512	541	500	582	665	984	99.0	104.6	96.7	112.6	128.6	190.3
Employed, $	587	626	612	630	615	748	734	106.6	104.3	107.3	104.8	127.4	125.0
Cohort, $	308	326	336	320	364	506	734	105.8	109.1	103.9	118.2	164.3	238.3
Age:	30-34												
Employed	463	504	548	547	639	783	1,090	108.9	118.4	118.1	138.0	169.1	235.4
Employed, $	636	607	636	594	643	720	736	95.4	100.0	93.4	101.1	113.2	115.7
Cohort, $	270	281	320	298	377	517	736	104.1	118.5	110.4	139.6	191.5	272.6
Age:	35-39												
Employed	423	456	465	498	617	740	1,064	107.8	109.9	117.7	145.9	174.9	251.5
Employed, $	589	577	616	597	613	731	738	98.0	104.6	101.4	104.1	124.1	125.3
Cohort, $	234	247	269	279	355	508	738	105.6	115.0	119.2	151.7	217.1	315.4
Age:	40-44												
Employed	396	418	474	497	582	722	953	105.6	119.7	125.5	147.0	182.3	240.7
Employed, $	651	627	663	626	690	760	805	96.3	101.8	96.2	106.0	116.7	123.7
Cohort, $	271	275	330	326	421	576	805	101.5	121.8	120.3	155.4	212.5	297.0
Age:	45-49												
Employed	360	391	434	416	535	650	834	108.6	120.6	115.6	148.6	180.6	231.7
Employed, $	631	629	616	653	677	763	771	99.7	97.6	103.5	107.3	120.9	122.2
Cohort, $	272	295	321	326	434	595	771	108.5	118.0	119.9	159.6	218.7	283.5
Age:	50-54												
Employed	287	304	314	324	412	493	630	105.9	109.4	112.9	143.6	171.8	219.5
Employed, $	661	670	639	665	687	759	739	101.4	96.7	100.6	103.9	114.8	111.8
Cohort, $	301	323	318	342	449	594	739	107.3	105.6	113.6	149.2	197.3	245.5
Age:	55-59												
Employed	202	216	243	242	299	353	457	106.9	120.3	119.8	148.0	174.8	226.2
Employed, $	648	683	673	701	701	755	722	105.4	103.9	108.2	108.2	116.5	111.4
Cohort, $	286	323	358	371	459	583	722	112.9	125.2	129.7	160.5	203.8	252.4
Age:	60-64												
Employed	144	161	173	177	225	266	325	111.8	120.1	122.9	156.2	184.7	225.7
Employed, $	642	611	645	675	696	713	690	95.2	100.5	105.1	108.4	111.1	107.5
Cohort, $	284	303	343	368	482	584	690	106.7	120.8	129.6	169.7	205.6	243.0
Age:	65-69												
Employed	110	111	121	127	166	197	216	100.9	110.0	115.5	150.9	179.1	196.4
Employed, $	675	737	763	741	724	708	633	109.2	113.0	109.8	107.3	104.9	93.8
Cohort, $	344	379	427	436	556	646	633	110.2	124.1	126.7	161.6	187.8	184.0
Age:	70 and older												
Employed	80	86	86	88	105	114	128	107.5	107.5	110.0	131.2	142.5	160.0
Employed, $	815	750	762	657	633	679	600	92.0	93.5	80.6	77.7	83.3	73.6
Cohort, $	509	504	512	452	519	605	600	99.0	100.6	88.8	102.0	118.9	117.9

*Average credits of less than $1,200 PYE, 1951-57

1/ Detail will not add to total because ages of 6 cohort members were not known for specific years of employment.

Table 10: **Female Intermediate Low-Paid Cohort***: INDICES OF ANNUAL EMPLOYMENT AND OF AVERAGE ANNUAL CREDITS OF EMPLOYED MEMBERS AND OF ENTIRE COHORT, 1951-57

Age in 1957, No. Employed, and Average Credits	Number of workers and average credits							Index (1951 = 100.0)					
	1951	1952	1953	1954	1955	1956	1957	1952	1953	1954	1955	1956	1957
Total:	Total												
Employed1/	4,471	4,972	5,386	5,646	6,241	6,752	7,215	111.2	120.5	126.3	139.6	151.0	161.4
Employed, $	1,358	1,495	1,637	1,663	1,808	2,022	2,095	110.1	120.5	122.5	133.1	148.9	154.3
Cohort, $	842	1,030	1,222	1,301	1,564	1,892	2,095	122.3	145.1	154.5	185.7	224.7	248.8
Age:	20-24												
Employed	400	614	775	885	1,046	1,134	1,194	Not Computed					
Employed, $	559	865	1,230	1,491	1,834	2,270	2,354						
Cohort, $	187	445	798	1,105	1,607	2,156	2,354						
Age:	25-29												
Employed	684	707	743	722	782	823	902	103.4	108.6	105.6	114.3	120.3	131.9
Employed, $	1,294	1,555	1,662	1,788	1,886	2,105	2,002	120.2	128.4	138.2	145.7	162.7	154.7
Cohort, $	981	1,219	1,369	1,431	1,635	1,921	2,002	124.3	139.6	145.9	166.7	195.8	204.1
Age:	30-34												
Employed	514	541	566	576	631	682	751	105.3	110.1	112.1	122.8	132.7	146.1
Employed, $	1,467	1,563	1,699	1,727	1,841	2,063	2,108	106.5	115.8	117.7	125.5	140.6	143.7
Cohort, $	1,004	1,126	1,280	1,325	1,547	1,873	2,108	112.2	127.5	132.0	154.1	186.6	210.0
Age:	35-39												
Employed	511	559	604	629	682	738	805	109.4	118.2	123.1	133.5	144.4	157.5
Employed, $	1,423	1,504	1,676	1,635	1,834	2,007	2,103	105.7	117.8	114.9	128.9	141.0	147.8
Cohort, $	903	1,044	1,258	1,278	1,554	1,840	2,103	115.6	139.3	141.5	172.1	203.8	232.9
Age:	40-44												
Employed	556	612	645	678	735	788	829	110.1	116.0	121.9	132.2	141.7	149.1
Employed, $	1,383	1,510	1,691	1,618	1,801	1,985	2,078	109.2	122.3	117.0	130.2	143.5	150.3
Cohort, $	928	1,115	1,316	1,323	1,597	1,887	2,078	120.2	141.8	142.6	172.1	203.3	223.9
Age:	45-49												
Employed	579	629	680	703	747	793	824	108.6	117.4	121.4	129.0	137.0	142.3
Employed, $	1,458	1,578	1,704	1,700	1,844	1,993	2,109	108.2	116.9	116.6	126.5	136.7	144.7
Cohort, $	1,024	1,205	1,406	1,450	1,672	1,918	2,109	117.7	137.3	141.6	163.3	187.3	206.0
Age:	50-54												
Employed	448	479	508	514	559	589	622	106.9	113.4	114.7	124.8	131.5	138.8
Employed, $	1,465	1,636	1,708	1,701	1,777	1,968	2,028	111.7	116.6	116.1	121.3	134.3	138.4
Cohort, $	1,055	1,260	1,395	1,406	1,597	1,864	2,028	119.4	132.2	133.3	151.4	176.7	192.2
Age:	55-59												
Employed	355	378	397	414	433	456	471	106.5	111.8	116.6	122.0	128.5	132.7
Employed, $	1,470	1,669	1,768	1,743	1,829	1,953	2,006	113.5	120.3	118.6	124.4	132.9	136.5
Cohort, $	1,108	1,339	1,490	1,532	1,681	1,891	2,006	120.8	134.5	138.3	151.7	170.7	181.0
Age:	60-64												
Employed	232	250	257	265	278	288	293	107.8	110.8	114.2	119.8	124.1	126.3
Employed, $	1,588	1,698	1,792	1,767	1,867	1,926	1,868	106.9	112.8	111.3	117.6	124.4	117.6
Cohort, $	1,257	1,449	1,572	1,598	1,771	1,942	1,868	115.3	125.1	127.1	140.9	154.5	148.6
Age:	65-69												
Employed	126	132	136	143	148	150	157	104.8	107.9	113.5	117.5	119.0	124.6
Employed, $	1,698	1,808	1,835	1,842	1,763	1,706	1,425	106.5	108.1	108.5	103.8	100.5	83.9
Cohort, $	1,363	1,520	1,590	1,678	1,662	1,630	1,425	111.5	116.7	123.1	121.9	119.6	104.5
Age:	70 and older												
Employed	63	65	65	67	66	72	73	103.2	103.2	106.3	104.8	114.3	115.9
Employed, $	1,606	1,736	1,819	1,724	1,734	1,550	1,530	108.1	113.3	107.3	108.0	96.5	95.3
Cohort, $	1,386	1,546	1,620	1,582	1,568	1,529	1,530	111.5	116.9	114.1	113.1	110.3	110.4

*Average credits of $1,200 - $2,399 PYE, 1951-57

1/ Detail will not add to total because ages of 6 cohort members were not known for specific years of employment.

Table 11: Female Intermediate High-Paid Cohort*: INDICES OF ANNUAL EMPLOYMENT AND OF AVERAGE ANNUAL CREDITS OF EMPLOYED MEMBERS AND OF ENTIRE COHORT, 1951-57

Age in 1957, No. Employed, and Average Credits	Number of workers and average credits							Index (1951 = 100.0)					
	1951	1952	1953	1954	1955	1956	1957	1952	1953	1954	1955	1956	1957
Total:	Total												
Employed1/	2,876	3,000	3,106	3,184	3,365	3,499	3,542	104.3	108.0	110.7	117.0	121.7	123.2
Employed, $	2,268	2,553	2,795	2,897	3,128	3,304	3,368	112.6	123.2	127.7	137.9	145.7	148.5
Cohort, $	1,842	2,162	2,451	2,604	2,972	3,264	3,368	117.4	133.1	141.4	161.3	177.2	182.8
Age:	20-24												
Employed	88	112	136	156	173	186	188	127.3	154.5	177.3	196.6	211.4	213.6
Employed, $	993	1,740	2,364	2,660	2,990	3,335	3,552	175.2	238.1	267.9	301.1	335.8	357.7
Cohort, $	465	1,037	1,710	2,207	2,751	3,300	3,552	223.0	367.7	474.6	591.6	709.7	763.9
Age:	25-29												
Employed	382	396	414	418	441	451	455	103.7	108.4	109.4	115.4	118.1	119.1
Employed, $	1,970	2,415	2,681	2,864	3,182	3,434	3,372	122.6	136.1	145.4	161.5	174.3	171.2
Cohort, $	1,654	2,102	2,439	2,631	3,084	3,404	3,372	127.1	147.5	159.1	186.5	205.8	203.9
Age:	30-34												
Employed	406	420	432	439	454	466	472	103.4	106.4	108.1	111.8	114.8	116.3
Employed, $	2,273	2,531	2,811	2,903	3,237	3,386	3,411	111.4	123.7	127.7	142.4	149.0	150.1
Cohort, $	1,955	2,252	2,573	2,700	3,114	3,343	3,411	115.2	131.6	138.1	159.3	171.0	174.5
Age:	35-39												
Employed	430	443	459	463	486	500	512	103.0	106.7	107.7	113.0	116.3	119.1
Employed, $	2,272	2,589	2,837	2,978	3,180	3,313	3,401	114.0	124.9	131.1	140.0	145.8	149.7
Cohort, $	1,908	2,240	2,543	2,793	3,019	3,235	3,401	117.4	133.3	141.1	158.2	169.5	178.2
Age:	40-44												
Employed	373	392	405	423	446	461	463	105.1	108.6	113.4	119.6	123.6	124.1
Employed, $	2,302	2,570	2,824	2,902	3,150	3,320	3,452	111.6	122.7	126.1	136.8	144.2	150.0
Cohort, $	1,855	2,176	2,470	2,651	3,034	3,306	3,452	117.3	133.2	142.9	163.6	178.2	186.1
Age:	45-49												
Employed	409	427	439	449	473	497	502	104.4	107.3	109.8	115.6	121.5	122.7
Employed, $	2,398	2,640	2,860	2,885	3,077	3,245	3,371	110.1	119.3	120.3	128.3	135.3	140.6
Cohort, $	1,954	2,246	2,501	2,580	2,899	3,213	3,371	114.9	128.0	132.0	148.4	164.4	172.5
Age:	50-54												
Employed	316	327	332	337	357	376	382	103.5	105.1	106.6	113.0	119.0	120.9
Employed, $	2,405	2,634	2,847	2,886	3,060	3,256	3,372	109.5	118.4	120.0	127.2	135.4	140.2
Cohort, $	1,989	2,255	2,474	2,546	2,860	3,205	3,372	113.4	124.4	128.0	143.8	161.1	169.5
Age:	55-59												
Employed	253	259	262	267	285	295	296	102.4	103.6	105.5	112.6	116.6	117.0
Employed, $	2,461	2,664	2,806	2,901	3,097	3,252	3,281	108.2	114.0	117.9	125.8	132.1	133.3
Cohort, $	2,103	2,331	2,484	2,717	2,982	3,241	3,281	110.8	118.1	124.4	141.8	154.1	156.0
Age:	60-64												
Employed	137	140	142	144	150	154	158	102.2	103.6	105.1	109.5	112.4	115.3
Employed, $	2,598	2,734	2,869	2,940	3,018	3,141	3,180	105.2	110.4	113.2	116.2	120.9	122.4
Cohort, $	2,253	2,423	2,578	2,679	2,865	3,061	3,180	107.5	114.4	118.9	127.2	135.9	141.1
Age:	65-69												
Employed	62	62	63	65	70	78	79	100.0	101.6	104.8	112.9	125.8	127.4
Employed, $	2,505	2,710	2,874	2,981	3,067	3,134	2,843	108.2	114.7	119.0	122.4	125.1	113.5
Cohort, $	1,966	2,127	2,292	2,453	2,718	3,094	2,843	108.2	116.6	124.8	138.2	157.4	144.6
Age:	70 and older												
Employed	19	21	21	22	25	25	25	110.5	110.5	115.8	131.6	131.6	131.6
Employed, $	2,547	2,799	3,028	3,109	3,147	3,112	2,722	109.9	118.9	122.1	123.6	122.2	106.9
Cohort, $	1,936	2,351	2,544	2,736	3,147	3,112	2,722	121.4	131.4	141.3	162.6	160.7	140.6

*Average credits of $2,400 - $3,599 PYE, 1951-57

1/ Detail will not add to total because ages of 6 cohort members were not known for specific years of employment.

the employed women, and average annual wage credits behaved similarly.

Female High-Paid Cohort (Table 12)

The average annual wage credits of the highest-paid female cohort increased steadily throughout the 1951-57 period, including the recession year of 1954. The indices of employment rose mo: e rapidly for those female cohort members 45 years of age or older than for the younger groups. Because the indices of annual wage credits of the employed cohort members, on the whole, rose more or less uniformly for all age groups, the indices of average annual wage credits were substantially higher for the older age groups than for the younger age groups. Among the two oldest age groups, the decline in the indices of average annual wage credits of the employed cohort members from 1956 to 1957 was accompanied by relative stability in the indices of employment; the index of average annual wage credits, therefore, declined during this 2-year period.

Differentials In Wage Credits--Female Cohorts (Tables 13 and 14)

For the lowest-paid female cohort, there was a general tendency for the differentials in average annual wage credits to narrow. Differentials in average annual wage credits of the employed members of the cohort, however, tended to widen slightly or to remain constant.

For the intermediate low-paid female cohort, the differential between average annual wage credits of the employed members of this cohort and those of the higher-paid cohort tended to remain constant, except among the very oldest and the very youngest age groups. The differential in average wage credits tended to decline, however, as a result of the greater relative increase in employment among the intermediate low-paid cohort members.

(Because of the relatively few women in the highest-paid cohort differentials for the female cohorts were computed by using the weighted average credits of women in the intermediate high-paid and the high-paid cohorts as a base.)

Aging and Wage Credits

The emphasis in this analysis has centered on changes in wage credits and employment levels. Implicit in the discussion is the conviction that "aging" is not a status which workers achieve at the end of the working lifetime, but rather an ongoing process, beginning at birth and terminating only at death. In terms of this latter conception--akin to the physiological view of aging as a development, maturation, and gradual erosion of the living organism--we have sought to trace the changing employment and wage patterns of workers as they aged through time.

Our data suggest that "aging" and wage changes are not simple, monolithic processes. Instead, we have found that the changes in employment and in the wage credits of the employed cohort members can and have moved in opposing directions, depending on the age, sex, and customary level of wage credits. The 1954 recession, for example, affected the wage credits of women not at all; for men, the higher the level of wage credits PYE, 1951-57, the older were the workers whose wage credits were adversely affected.

The data also suggest that for some men, aging in the immediate preretirement years does not result in a climactic rise of wages. Instead, 8.5 percent of the men in our cohort aged 60-64 had average credits per year employed, 1951-57, of $1,200 or less; for 24.7 percent of the men, credits per year employed, 1951-57 averaged $2,400 or less. It also appears that for most of these workers, relative wage credits were quite low throughout the 1937-57 period, and throughout their entire working lifetimes (Table 15). For most workers, wage credit levels tend to be relatively constant throughout their lifetimes. For men and women workers in the lowest portions of the wage credit structure, our data suggest, the low level of credits is the result not only of relatively low wage credits when they are employed but of relatively low levels of employment. For these workers, an increased level of living, therefore, requires not only an increase in wage credits when employed, but also a substantial rise in employment.

Table 12: Female High-Paid Cohort*: INDICES OF ANNUAL EMPLOYMENT AND OF AVERAGE ANNUAL CREDITS OF EMPLOYED MEMBERS AND OF ENTIRE COHORT, 1951-57

Age in 1957, No. Employed, and Average Credits	Number of workers and average credits							Index (1951 = 100.0)					
	1951	1952	1953	1954	1955	1956	1957	1952	1953	1954	1955	1956	1957
Total:	Total												
Employed1/	733	737	744	753	851	932	946	100.5	101.5	102.7	116.1	127.1	129.1
Employed, $	3,374	3,500	3,567	3,578	4,086	4,132	4,136	103.7	105.7	106.0	121.1	122.5	122.6
Cohort, $	2,614	2,727	2,805	2,848	3,676	4,071	4,136	104.3	107.3	109.0	140.6	155.7	158.2
Age:	20-24												
Employed	---	---	---	---	---	---	---	Not computed					
Employed, $	---	---	---	---	---	---	---						
Cohort, $	---	---	---	---	---	---	---						
Age:	25-29												
Employed	21	21	22	22	24	28	30	100.0	104.8	104.8	114.3	133.3	142.9
Employed, $	3,028	3,398	3,497	3,574	4,082	4,132	4,144	112.2	115.5	118.0	134.8	136.5	136.9
Cohort, $	2,120	2,379	2,564	2,621	3,266	3,857	4,144	112.2	120.9	123.6	154.1	181.9	195.5
Age:	30-34												
Employed	74	75	76	77	82	83	84	101.4	102.7	104.1	110.8	112.2	113.5
Employed, $	3,203	3,462	3,529	3,586	4,058	4,159	4,134	108.1	110.2	112.0	126.7	129.8	129.1
Cohort, $	2,822	3,091	3,193	3,287	3,961	4,109	4,134	109.5	113.1	116.5	140.4	145.6	146.5
Age:	35-39												
Employed	99	99	100	102	109	113	119	100.0	101.0	103.0	110.1	114.1	120.2
Employed, $	3,354	3,514	3,572	3,577	4,079	4,107	4,122	104.8	106.5	106.6	121.6	122.5	122.9
Cohort, $	2,790	2,923	3,002	3,066	3,736	3,900	4,122	104.8	107.6	109.9	133.9	139.8	147.7
Age:	40-44												
Employed	121	121	121	121	134	142	145	100.0	100.0	100.0	110.7	117.4	119.8
Employed, $	3,364	3,485	3,577	3,572	4,066	4,136	4,164	103.6	106.3	106.2	120.9	122.9	123.8
Cohort, $	2,807	2,908	2,985	2,981	3,758	4,050	4,164	103.6	106.3	106.2	133.9	144.3	148.3
Age:	45-49												
Employed	124	127	129	133	151	166	166	102.4	104.0	107.3	121.8	133.9	133.9
Employed, $	3,396	3,490	3,568	3,569	4,087	4,156	4,157	102.8	105.1	105.1	120.3	122.4	122.4
Cohort, $	2,537	2,670	2,773	2,860	3,718	4,156	4,157	105.2	109.3	112.7	146.6	163.8	163.9
Age:	50-54												
Employed	119	119	119	119	135	152	154	100.0	100.0	100.0	113.4	127.7	129.4
Employed, $	3,398	3,510	3,570	3,582	4,111	4,122	4,163	103.3	105.1	105.4	121.0	121.3	122.5
Cohort, $	2,626	2,712	2,759	2,768	3,604	4,068	4,163	103.3	105.1	105.4	137.2	154.9	158.5
Age:	55-59												
Employed	104	104	106	108	125	141	141	100.0	101.9	103.8	120.2	135.6	135.6
Employed, $	3,461	3,516	3,571	3,580	4,098	4,124	4,136	101.6	103.2	103.4	118.4	119.2	119.5
Cohort, $	2,553	2,593	2,685	2,742	3,633	4,124	4,136	101.6	105.2	107.4	142.3	161.5	162.0
Age:	60-64												
Employed	52	52	52	52	62	72	72	100.0	100.0	100.0	119.2	138.5	138.5
Employed, $	3,487	3,554	3,593	3,586	4,071	4,121	4,077	101.9	103.0	102.8	116.7	118.2	116.9
Cohort, $	2,518	2,567	2,595	2,590	3,506	4,121	4,077	101.9	103.1	102.9	139.2	163.7	161.9
Age:	65-69												
Employed	13	13	13	13	22	27	27	100.0	100.0	100.0	169.2	207.7	207.7
Employed, $	3,486	3,554	3,582	3,600	4,164	4,142	3,949	102.0	102.8	103.3	119.4	118.8	113.3
Cohort, $	1,678	1,711	1,725	1,733	3,393	4,142	3,949	102.0	102.8	103.3	202.2	246.8	235.3
Age:	70 and older												
Employed	6	6	6	6	7	8	8	100.0	100.0	100.0	116.7	133.3	133.3
Employed, $	3,597	3,600	3,600	3,600	4,188	4,058	4,042	100.1	100.1	100.1	116.4	112.8	112.4
Cohort, $	2,698	2,700	2,700	2,700	3,664	4,058	4,042	100.1	100.1	100.1	135.8	150.4	149.8

*Average credits of $3,600 or more PYE, 1951-57

1/ Detail will not add to total because ages of 6 cohort members were not known for specific years of employment.

Table 13: Female Cohorts: WAGE CREDIT DIFFERENTIALS OF EMPLOYED COHORT MEMBERS (Average Annual Credits of Employed Members of Lower-Paid Cohorts as a Percentage of Average Annual Credits of Employed Members of Two Highest-Paid Cohorts, 1951-57)

Age in 1957 and subcohorts		1951	1952	1953	1954	1955	1956	1957
Total	2 low-paid subcohorts....	41.1	40.1	39.7	39.0	36.9	38.3	36.4
	Low-paid[a]/...............	23.6	21.4	20.3	19.6	18.5	20.4	20.6
	Intermediate low-paid[b]/..	54.5	54.6	55.6	54.9	54.3	58.1	59.3
16-19	2 low-paid subcohorts....	30.6	8.3	7.0	7.3	17.9	24.0	26.6
	Low-paid[a]/...............	0.8	3.8	6.0	6.6	14.3	17.6	19.7
	Intermediate low-paid[b]/..	149.9	36.7	19.0	11.2	35.9	55.5	74.4
20-24	2 low-paid subcohorts....	42.5	35.3	34.9	37.8	41.2	45.9	41.9
	Low-paid[a]/...............	29.8	22.1	20.4	20.7	21.9	26.0	24.4
	Intermediate low-paid[b]/..	56	49.7	52.0	56.1	61.3	68.1	66.3
25-29	2 low-paid subcohorts....	48.9	47.3	44.8	45.3	41.5	43.1	39.2
	Low-paid[a]/...............	29.0	25.4	22.5	21.7	19.0	21.5	21.5
	Intermediate low-paid[b]/..	63.9	63.1	61.1	61.7	58.2	60.5	58.5
30-34	2 low-paid subcohorts....	44.4	41.2	40.3	39.1	36.8	38.4	36.8
	Low-paid[a]/...............	26.3	22.7	21.8	19.8	19.1	20.6	20.9
	Intermediate low-paid[b]/..	60.7	58.5	58.2	57.5	54.8	58.9	59.9
35-39	2 low-paid subcohorts....	42.2	39.4	41.0	38.1	37.4	39.5	37.5
	Low-paid[a]/...............	23.8	20.9	20.8	19.3	18.3	21.1	20.9
	Intermediate low-paid[b]/..	57.5	54.5	56.5	53.0	54.7	58.0	59.5
40-44	2 low-paid subcohorts....	42.1	41.3	41.9	39.2	38.9	39.8	38.6
	Low-paid[a]/...............	25.4	22.5	22.1	20.5	20.5	21.6	22.2
	Intermediate low-paid[b]/..	54.0	54.2	56.4	53.0	53.5	56.5	57.4
45-49	2 low-paid subcohorts....	43.4	42.9	42.4	43.1	40.8	41.5	40.2
	Low-paid[a]/...............	24.0	22.2	20.4	21.5	20.3	22.0	21.6
	Intermediate low-paid[b]/..	55.4	55.7	56.4	55.9	55.4	57.4	59.1
50-54	2 low-paid subcohorts....	43.0	44.0	42.8	42.4	39.1	40.4	38.3
	Low-paid[a]/...............	24.7	23.4	21.0	21.7	20.4	21.7	20.5
	Intermediate low-paid[b]/..	54.7	57.0	56.2	55.4	52.9	56.1	56.3
55-59	2 low-paid subcohorts....	42.6	45.1	44.7	43.9	40.0	40.5	38.6
	Low-paid[a]/...............	23.5	23.5	22.2	22.6	20.5	21.4	20.3
	Intermediate low-paid[b]/..	53.4	57.4	58.4	56.3	53.5	55.3	56.4
60-64	2 low-paid subcohorts....	43.1	43.0	43.5	42.7	40.4	39.0	36.1
	Low-paid[a]/...............	22.6	20.7	21.1	21.7	20.9	20.7	19.9
	Intermediate low-paid[b]/..	55.9	57.4	58.5	56.8	56.1	55.9	54.0
65-69	2 low-paid subcohorts....	45.6	46.2	44.4	42.9	36.5	33.6	30.9
	Low-paid[a]/...............	25.2	25.8	25.5	24.0	21.7	20.9	20.3
	Intermediate low-paid[b]/..	63.5	63.3	61.3	59.7	53.0	50.3	45.6
70 and over	2 low-paid subcohorts....	41.6	39.4	38.6	34.8	31.4	30.4	30.8
	Low-paid[a]/...............	29.1	25.2	24.2	20.4	18.8	20.3	19.7
	Intermediate low-paid[b]/..	57.4	58.3	57.7	53.6	51.4	46.4	50.3

Workers with average credits PYE, 1951-57, of:

a/ Less than $1,200

b/ $1,200 to $2,399

Table 14: Female Cohorts: WAGE CREDIT DIFFERENTIALS OF COHORT MEMBERS
(Average Annual Credits of Members of Lower-Paid Cohorts as a Percentage of Average Annual Credits of Members of Two Highest-Paid Cohorts, 1951-57)

Age in 1957 and subcohorts		1951	1952	1953	1954	1955	1956	1957
Total	2 low-paid subcohorts....	28.0	29.5	31.1	31.3	31.1	33.9	36.4
	Low-paid[a]...............	9.6	9.4	10.0	9.9	11.1	14.7	20.6
	Intermediate low-paid[b]..	42.0	45.1	48.4	49.0	50.0	55.1	59.3
16-19	2 low-paid subcohorts....	1.0	1.0	3.3	10.5	13.1	16.1	26.6
	Low-paid[a]...............	----	0.3	3.1	9.2	9.4	10.3	19.7
	Intermediate low-paid[b]..	5.0	5.2	5.2	18.3	32.0	44.9	74.4
20-24	2 low-paid subcohorts....	27.9	28.3	29.8	31.6	36.6	41.4	41.9
	Low-paid[a]...............	16.6	14.9	15.5	14.3	15.6	19.9	24.4
	Intermediate low-paid[b]..	40.2	42.9	46.7	50.1	58.4	65.3	66.3
25-29	2 low-paid subcohorts....	41.1	39.8	38.1	37.1	35.2	37.5	39.2
	Low-paid[a]...............	18.3	15.4	13.7	12.2	11.7	14.7	21.5
	Intermediate low-paid[b]..	58.3	57.5	55.9	54.4	52.7	55.9	58.5
30-34	2 low-paid subcohorts....	31.4	30.2	30.3	29.6	29.5	33.1	36.8
	Low-paid[a]...............	12.9	11.8	12.0	10.7	11.6	14.9	20.9
	Intermediate low-paid[b]..	48.1	47.3	48.0	47.5	47.7	54.1	59.9
35-39	2 low-paid subcohorts....	28.9	29.0	31.5	30.3	31.1	34.9	37.5
	Low-paid[a]...............	11.3	10.4	10.2	10.1	11.2	15.1	20.9
	Intermediate low-paid[b]..	43.5	44.1	47.9	46.3	49.2	54.7	59.5
40-44	2 low-paid subcohorts....	31.5	32.9	34.7	33.0	33.5	36.1	38.6
	Low-paid[a]...............	13.0	11.7	12.7	11.9	13.1	16.5	22.2
	Intermediate low-paid[b]..	44.6	47.4	50.8	48.5	49.7	54.1	57.4
45-49	2 low-paid subcohorts....	35.1	36.4	38.3	38.9	37.2	38.4	40.2
	Low-paid[a]...............	13.0	12.5	12.5	12.3	14.0	17.3	21.6
	Intermediate low-paid[b]..	48.8	51.3	54.7	54.7	53.8	55.7	59.1
50-54	2 low-paid subcohorts....	35.0	37.5	38.5	38.1	35.9	37.2	38.3
	Low-paid[a]...............	13.9	13.5	12.4	13.1	14.5	17.2	20.5
	Intermediate low-paid[b]..	48.6	52.8	54.6	53.9	51.7	54.0	56.3
55-59	2 low-paid subcohorts....	36.0	40.1	41.6	41.5	36.8	37.4	38.6
	Low-paid[a]...............	12.7	13.4	14.1	14.0	14.3	16.5	20.3
	Intermediate low-paid[b]..	49.3	55.4	58.5	57.6	52.4	53.6	56.4
60-64	2 low-paid subcohorts....	37.9	40.5	41.7	41.7	39.0	38.0	36.1
	Low-paid[a]...............	12.2	12.3	13.3	13.9	15.7	17.2	19.9
	Intermediate low-paid[b]..	53.8	58.7	60.8	60.3	57.8	57.3	54.0
65-69	2 low-paid subcohorts....	46.9	49.4	48.6	48.2	37.3	31.9	30.9
	Low-paid[a]...............	18.2	18.8	19.9	19.2	19.2	19.2	20.3
	Intermediate low-paid[b]..	72.0	75.2	74.1	74.0	57.5	48.5	45.6
70 and over	2 low-paid subcohorts....	42.2	39.1	38.3	34.5	28.2	28.8	30.8
	Low-paid[a]...............	24.0	20.7	19.8	16.6	15.9	18.1	19.7
	Intermediate low-paid[b]..	65.4	63.5	62.8	58.0	47.9	45.8	50.3

Workers with average credits PYE, 1951-57, of:

a/ Less than $1,200

b/ $1,200 to $2,399

Table 15: Male Cohort: AVERAGE CREDITS PER ELAPSED YEAR, 1937-50 and 1951-57

Age in 1957	Average Credits Per Year Employed, 1951-57							
	Under $1,200		$1,200-$2,399		$2,400-$3,599		$3,600 & over	
	1937-1950 (1)	1951-1957 (2)	1937-1950 (1)	1951-1957 (2)	1937-1950 (1)	1951-1957 (2)	1937-1950 (1)	1951-1957 (2)
Total..	$ 88	$ 352	$ 313	$1,390	$ 711	$2,701	$1,507	$3,666
1-15.......	---	49	---	524	1,146	2,486	---	---
16-19.......	---	170	---	606	---	964	---	1,144
20-24.......	1	431	5	1,213	20	2,189	41	2,449
25-29.......	67	463	128	1,317	226	2,469	379	3,233
30-34.......	165	479	299	1,516	430	2,715	696	3,638
35-39.......	224	483	393	1,565	619	2,774	1,085	3,649
40-44.......	267	463	513	1,558	901	2,834	1,553	3,680
45-49.......	284	463	652	1,617	1,081	2,872	1,850	3,739
50-54.......	291	470	704	1,640	1,215	2,904	2,001	3,719
55-59.......	346	460	736	1,656	1,411	2,929	2,140	3,742
60-64.......	401	481	819	1,673	1,456	2,990	2,112	3,681
65-69.......	344	528	843	1,617	1,596	2,913	2,184	3,681
70 and over.	600	488	1,035	1,535	1,571	2,854	2,068	3,554

Table 16: Female Cohort: AVERAGE CREDITS PER ELAPSED YEAR, 1937-50 and 1951-57

Age in 1957	Average Credits Per Year Employed, 1951-57							
	Under $1,200		$1,200-$2,399		$2,400-$3,599		$3,600 & over	
	1937-1950 (1)	1951-1957 (2)	1937-1950 (1)	1951-1957 (2)	1937-1950 (1)	1951-1957 (2)	1937-1950 (1)	1951-1957 (2)
Total..	$ 107	$ 357	$ 303	$1,421	$ 650	$2,666	$1,124	$3,268
1-15.......	---	32	---	939	---	---	---	---
16-19.......	---	141	---	543	56	1,087	---	---
20-24.......	3	395	2	1,236	13	2,146	---	---
25-29.......	76	413	144	1,508	222	2,669	338	2,993
30-34.......	181	400	333	1,466	586	2,764	805	3,514
35-39.......	188	376	355	1,426	696	2,720	1,014	3,363
40-44.......	186	429	381	1,463	725	2,706	1,107	3,379
45-49.......	179	431	411	1,541	796	2,681	1,133	3,267
50-54.......	157	438	436	1,515	830	2,672	1,264	3,243
55-59.......	153	443	494	1,578	898	2,720	1,305	3,209
60-64.......	163	436	572	1,637	997	2,720	1,367	3,139
65-69.......	169	489	565	1,553	1,002	2,499	794	2,619
70 and over.	241	529	622	1,537	871	2,650	1,918	3,223

MR. FARBER'S REJOINDER TO COMMENTS OF THE DISCUSSANTS

Messrs. Miller and Wernick apparently agree that, in Mr. Miller's words, "the results of.../my/ study are often treated as though they represent the entire universe instead of a segment of it." Mr. Wernick expresses the hope that findings uncovered in my paper be reconciled with other existing data. To some extent this was attempted in the discussion of wage credit differentials, a subject to which I shall address myself in a moment.

In a larger sense, however, I believe this type of criticism to be irrelevant. The universe in a cohort study is multi-dimensional -- it has both a temporal and a worker-population paramater. Changes in the level of employment of the cohort, and in the wage credits of the employed members -- and the interactional effects of these two variables on the average credits of the entire cohort -- are measured for each of the seven successive years in the 1951-57 period. Since membership in the cohort was based on employment in 1957 in work covered by the Social Security Act, but the employment and wage credit histories (concededly only in "covered" employment) of the cohort encompass the entire 1951-57 period, defining just these two time dimensions of the universe presents difficulties. Discussion of these problems at this meeting would be time consuming, and because of the undeveloped state of the art of cohort analysis, would probably contribute relatively little to our knowledge. While I do not pretend to have entirely defined the universe of a cohort sample, I am certain that a theoretical discussion of this magnitude could not possibly be presented within the confines of my allotted time -- certainly not if I wished to present any of the findings disclosed by my study.

On this subject I am content merely to note briefly some of the theoretical difficulties, perhaps even the theoretical impossibility, of reconciling inferences based on cross-sectional trends with findings based on longitudinal or cohort data. While I am aware that such inferences, in fact, have been made, I am not persuaded that the results of a longitudinal analysis can reasonably be expected to be consistent with trends disclosed by cross-sectional comparisons. If longitudinal inferences can be successfully drawn from cross-sectional comparisons there is little need to undertake any longitudinal analysis. Furthermore, as we shall see in a moment, comparison of cross-sectional income or wage data for different time periods involves problems of defining the universe which are even more difficult and more complex than those involved in defining the universe for a cohort sample.

* * *

Some of the problems which the discussants found in the analysis, I believe, result from a misunderstanding of the research design on which the study is based. Basically, my paper compares the employment and wage credit histories of identical age groups in each of the 4 wage subcohorts. Thus, in the comparisons, age and aging, are held constant. The relative impact of changes in two variables -- employment of the cohort and the wage credits of the employed cohort members -- on the wage credits of the entire cohort are distinguished from the effects of age and aging, the precise changes in these two variables is measured separately, and the effects of these changes -- both jointly and singly -- on the average credits of the entire subcohort are indicated.

Much of Mr. Miller's criticism, it appears to me, is therefore irrelevant. It is based on the erroneous assumption that each of the tables -- and each of the indices for a given age group -- are to be considered separately, and their reasonableness evaluated against findings from cross-sectional data. Precisely because, as Mr. Miller points out, Social Security wage data are circumscribed by limitations of coverage and by taxable limits, my paper compares the joint effects of changes in the employment and wage experience of the higher-paid subcohorts with those of the lower paid subcohorts. The significance of this study, therefore, should be judged on the basis of the inter-cohort comparisons, not, as Mr. Miller assumes, on a comparison of the longitudinal data for one cohort, or of one of the indices for a given cohort, with trends disclosed by separate cross-sectional studies of employment or wages.

These comparisons can be made most conveniently by referring to the changes disclosed by the indices in the several tables. In 1954, for example, for 25-29 year old men in the highest paid cohort, the index of employment was 123.2; the index of wage credits of employed members of the cohort was 108.6. For men in this cohort aged 50-54, the 1954 index of employment was 100.5; and of average credits of employed cohort members, 101.1. Because the employment index of the younger cohort for 1954 rose 6 index points above the 1953 level, while the index of credits of employed members of this age-wage cohort remained stable, the index of average credits of this age cohort rose from 127.0 in 1953 to 134.0 in 1954. For 50-54 year old men in the highest paid cohort, however, the indices of average credits of the entire cohort were unchanged, because the indices of employment and of the credits of the employed members were stable in both 1953 and 1954 (Table 6).

If we compare these data with those for low paid males of identical age by disaggregating the effects of changes in the two variables, an entirely different picture emerges. For the 25-29 year old group, the indices reveal a decline in the number of employed subcohort members which, when coupled with a decline in the average credits of its employed members forced the index of average credits of the entire cohort some 6.8 index points below, the 1953 level. For the lowest paid men aged 50-54, the decline in the employment index was not as severe as in the case of the 25-29 year old men, but because the index of average credits of the employed subcohort fell from 94.1 in 1953 to 74.3 in 1954, the index of average credits of the entire cohort declined by more than 21 index points. For men aged 50-54, therefore, the 1954 recession had an adverse effect on the wage credits of the lowest paid cohort, but had no adverse effects on the highest paid cohort.

Thus, disaggregation -- one of Mr. Wernick's desiderata -- is not only possible in the analysis of cohort data, but is actually accomplished in the study. Far from being a mere identification, as Mr. Miller believes, age of a particular subcohort is used to demonstrate, for example, that the age at which a worker "ages" or becomes "old," in terms of demand for his services, tends to vary with his status in the wage structure. Higher paid workers, the data indicates, were adversely affected by the 1954 recession at an older age than were the lower paid workers.

My critics also challenge the finding that wage credit differentials narrowed in the 1951-57 period. Mr. Miller would attribute this finding to the limitations of the maximum taxable limit on wages subject to Social Security taxation, and to a certain naivete in the analysis. Mr. Wernick's critique suggests that I cite sources for the statement that this finding is consistent with the findings of other labor economists. So be it: Citation No. 1. In the Evolution of Wage Structure, by Lloyd Reynolds and Cynthia Taft, published in 1956, the authors note that

> "Most types of wage differential have tended to narrow, not only in the United States but in other countries." (p. 194)

Citation No. 2. Clark Kerr, in Dunlop's the Theory of Wage Determination, published in 1957, also indicates that occupational differentials have been much reduced, and firm wage differentials have also been reduced or occasionally eradicated, as have industry wage differentials. Lastly, I cite Mr. Miller's Technical Paper No. 8, Trends in the Income of Families and Persons in the United States: 1947 to 1960, published in 1963 by the U. S. Bureau of the Census. Tables A and B below, based on Tables A and B in Technical Paper No. 8, indicate that in 1951 families in the lowest income stratum (the classification scheme is Mr. Miller's) were 5-3/4 times as numerous as those in the highest income stratum. In each succeeding year, this ratio declined, until by 1957, the lowest income families were somewhat less than 2-1/2 times as numerous as those in the highest income stratum (Table A). A similar trend is revealed by the data for unrelated individuals, as shown in Table B. Mr. Miller's data, therefore, although they relate to income rather than to wage trends, tend to support the conclusion that wage differentials in fact did narrow during the 1951-57 period.

Table A: Number and Proportion of Families, By Total Money Income in Constant (1959) Dollars

Families, by Total Money Income	Year						
	1951	1952	1953	1954	1955	1956	1957
Number: (000)							
Total	40,442	41,020	------	41,934	42,843	43,445	43,714
$10,000 and over	2,165	2,374	------	3,152	3,563	4,350	4,187
Under $3,000	12,470	12,034	------	12,012	11,074	10,247	10,296
$3,000 - $4,999	13,360	12,935	------	11,915	11,502	10,768	10,678
$5,000 - $9,999	12,447	13,677	------	14,885	16,704	18,080	18,553
As Percent of Number of "$10,000 and over" Families							
Under $3,000	576.0	506.9	------	381.1	310.8	235.6	245.9
$3,000 - $4,999	617.1	544.9	------	378.0	322.8	247.5	255.0
$5,000 - $9,999	574.9	576.1	------	472.2	468.8	415.6	443.1

Based on Table A, Trends in the Income of Families and Persons in the United States: 1947 to 1960. Technical Paper No. 8, U. S. Government Printing Office, Washington, D. C., 1963, pp. 6-11.

Table B: Number and Proportion of Unrelated Individuals by Total Money Income in Constant (1959) Dollars

Unrelated Individuals	Year						
	1951	1952	1953	1954	1955	1956	1957
Number: (000)							
Total	9,015	9,774	-----	9,623	9,766	9,658	10,313
$5,000 or more	450	680	-----	684	750	918	1,125
Under $1,000	3,757	3,542	-----	3,950	3,720	3,549	3,669
$1,000 - $2,999	3,108	3,707	-----	3,325	3,481	3,379	3,603
$3,000 - $4,999	1,700	1,845	-----	1,664	1,815	1,812	1,916
As Percent of Number of "$5,000 or more" Individuals							
Under $1,000	834.9	520.9	-----	577.5	496.0	386.6	326.1
$1,000 - $2,999	690.7	545.2	-----	486.1	464.1	368.1	320.3
$3,000 - $4,999	377.8	271.3	-----	243.3	242.0	197.4	170.3

Based on Table B, Trends in the Income of Families and Persons in the United States: 1947 to 1960. Technical Paper No. 8, U. S. Government Printing Office, Washington, D. C., 1963, pp. 6-11.

I conclude with a brief reference to Mr. Miller's dismissal of Social Security data, of cohort data in general, and to his insistence that these wage data should be inflated to reflect average total wages rather than average taxable wage credits. I believe this to be a species of guessing which can distort real data. In Technical Paper No. 8 Mr. Miller abides by his own precepts, and utilizes the Consumer Price Index to estimate "real" income for the income groups into which he classifies the families and unrelated individuals in his study. Although he cautions that "the same index was used for all groups because separate price indices have not been developed for various income levels," he nonetheless utilizes the Consumer Price Index, regardless of the possibility that prices of food or of medical services may have increased at a faster rate for low income groups than for the high income groups, or that an increase in these prices may have a more significant effect on the "real" income of the low income group than on the high income group. This procedure, it seems to me, provides the appearance, but not the substance, of precision. Mrs. Selma Goldsmith, noted that "Income for a single year is not a satisfactory measure of income inequality"; that failure to develop cost-of-living indices appropriate for various income groups means "that we cannot measure with precision changes in the distribution of real income." Finally, in judging the value of cohort analysis, I urge Mr. Miller to reconsider his judgment in the light of one other comment made by Mrs. Goldsmith:

> "When we compare income shares of a given quintile or the top 5 percent in two periods, we are not comparing what has happened to an identical group of families, because the families comprising the quintile may be quite different in the two periods. For certain purposes, as for example, in interpreting the change in the income share of the top quintile... over, say, a five-to-ten-year time span, it would be extremely helpful to know the extent to which the families comprising the top sector differed in the terminal periods."1/

To end this rejoinder, I refer again to the problem raised by Mr. Miller. But this time, I ask him, "What is the universe when we compare cross-sectional income data for two different time periods?"

Despite Mr. Miller's observation concerning the limitations of unadjusted cohort wage data based on Social Security records, precise and disaggregated measurement of changes through time for an identical sample are possible. Can Mr. Miller claim the same for his cross-sectional data?

It appears to me that Mr. Perlman's comments are completely correct. No set of statistics is without limitations. Our knowledge of wage and employment changes is sufficiently meager -- and our insights so few -- that it takes a brave man indeed to point the finger and say, "This is useless."

1/ The quotations attributed to Mrs. Goldsmith may be found in her article, "Changes in the Size Distribution of Income," American Economic Review, May 1957, p. 511.

THE SOCIAL SECURITY CONTINUOUS WORK HISTORY SAMPLE AS A SOURCE OF COHORT WAGE DATA

S. D. Hearn, Social Security Administration

Purpose and Background

This paper is intended as an explanation of the continuous work history sample of the old-age, survivors insurance, and disability program which is currently being administered under the Social Security Administration. More specifically, the explanation of the sample will be directed toward the use of such sample data as a source for cohort wage analysis.

Under present legislation, over nine out of ten jobs are covered under the program. The exceptions apply to work under federal civil service and under railroad retirement, incidental employment, and to self-employed doctors. Since employers are required to report periodically the taxable earnings of each of their workers, and since self-employed must report their earnings on an annual basis, a very large quantity of data is available to the Social Security Administration. From this vast accumulation of data, a sample of records is selected for statistical analysis on a continuing annual operation.

When the reports for a calendar year are received and posted to the earnings record tape files, maintained by the Social Security Administration, a 1 percent sample of tape records is set aside for statistical processing. The results of such statistical processing are available in small part from the annual handbooks which have been released by the Social Security Administration.1/ Many more tabulations are prepared for analyses by the statisticians, economists, and actuaries of the Administration, as well as by those of other Government agencies and public and private organizations.

Data are available to varying degrees of detail, based on size of sample. In general, the greater the size of sample, the less detail is available. For example, there are 100%, 20% 1%, and 0.1% data. The extent of detail in the varying sample sizes is shown in Appendix 1. Although the principal purpose of the continuous work history sample is for internal studies in the Social Security Administration, the sample was designed to serve many purposes. Many other Government agencies have found the data useful in connection with their own programs as well as for general research purposes. Although a number of studies have been completed involving analyses of the economic and personal characteristics of workers employed over a period of time,2/ and the use of OASI data for medical research purposes,3/ little use has been made of the continuous work history sample for wage studies based on cohort analyses, until the past few years, when Mr. David J. Farber of the U.S. Department of Labor began his studies.

Basic Forms and Data

Cohort studies are possible, using OASI data, because of the compulsory initial registration and periodic reporting requirements under the social security system. Current wage and employment information for a worker, as reported by his employer, can be associated with the personal characteristics of the worker and the environmental characteristics of the employer.

On the worker's application form for an account number, he provides date of birth, sex, race, and other identifying information including mother's and father's name, and his place of birth. Another compulsory registration form is required to be filled out by employers when they register with the Internal Revenue Service. In this case, the employer must identify his place and nature of business. If he has more than one place of business, he is asked to provide, on a voluntary basis place and nature of industrial activity for each of his places of business. Although these data tend to remain static, provision is made for recording changes in our records.4/

1/ The most recent handbook is: Handbook of Old-Age and Survivors Insurance Statistics, Employment, Wages, and Insurance Status of Workers in Covered Employment 1955, U.S. Department of Health, Education, and Welfare, Social Security Administration, BOASI, Baltimore 35, Maryland

2/ For example, see "The Continuous Work History Sample: The First 12 Years," by Jacob Perlman Social Security Bulletin, April 1951.

3/ For example, see "Methods of Studying the Relation of Employment and Long-term Illness-Cohort Analysis," by Thomas F. Mancuso and Elizabeth Coulter, American Journal of Public Health, November 1959.

4/ Periodically, large single unit employers and most multi-unit employers are contacted to determine whether changes have occurred since their previous report, in location and nature of activity at their establishments. Workers at any time may report changes in name, date of birth, or any other items previously reported on their account number application form.

In addition to these two registration forms, data are required to be reported periodically by employers and by the self-employed. In general, employers report for each employee the amount of earnings, up to the taxable limit, and the calendar quarter of those earnings. The self-employed report both their annual earnings from wages, also up to the taxable limit, and the amount of annual earnings from self-employment. Thus, no quarterly data are available from reports by the self-employed.

A third set of forms apply to the termination of the individual's working career. Benefits are payable, under the social security program, in the event of permanent disability, retirement, or death. To obtain such benefits, workers (and their families) provide information on date of termination (i.e., disability, retirement, or death), family composition, recent and estimated future earnings, military service, etc. In some disability and retirement cases the working career may be resumed, at least on a partial basis. A cohort study of such cases may be of significance to those interested in measuring impact on the labor force of previously retired and incapacitated persons.

To summarize, detailed work history records are maintained from the time the individual indicates his intention of wishing to work (by applying for a social security account number[5/]) through his employment with each of his employers in covered employment and/or self-employment, to the end of his employment as indicated by the benefit application for disability, or retirement, or death (in which case generally, a survivor files for a benefit.)

These files are maintained in many segments to facilitate the accounting and claims operations involved. To make statistical analyses of the characteristics of workers it is necessary to collate data from the many files in order to produce a single record for a worker containing this wide variety of information.

This type of collation was done for the continuous work history sample beginning with employment in the years 1937 through 1940. Annually thereafter summarized data for individuals in the sample were added to the original data to facilitate statistical summarizations for analysis. The latest continuous work history file contains summarized data for each individual in the sample who had covered employment in any year of the twenty-four year period 1937 through 1961.[6/] Appendix 1 shows the corresponding data which will become available in the 1937-62 summarization.

Method of Sample Selection

To select this sample a digital technique was employed. Since each worker is identified permanently by a nine digit social security account number, this number was used as the basis for selection. The first three digits of the account number indicate the State in which the number was issued. The next two digits, together with the first three digits, are used in controlling the assignment of blocks of numbers to the local social security offices which issue the individual numbers to the public. To fall in the 1% continuous work history sample, the individual's last four digits of the account number must have the following combination:

> The sixth digit of the number must equal "2" or "7" (which provides the 20% sample) and the eighth and ninth digits must equal "05," "20," "45," "70," or "95."

To assure accurate representation of the sample among persons who recently obtained account numbers, for each block of 500 numbers provided by the central office to any local office, 100 represent account numbers with the digits "2" and "7" in the sixth digit position, and the remaining 400 represent other than the digits "2" and "7" in that digit position. Thus, for each issuance of 500 numbers, 100, or 20 percent represent those with the required sample digits. It is conceivable that as of the end of any one year there may not be the exact 20 percent, and consequently, 1 percent sample representation for the account numbers issued during the year. However, because of the large volume of issuances there is little likelihood of the expected sample count being significantly different from the actual sample count of numbers issued.

The 1 percent sample is further subdivided by taking an 0.1 percent sample selected on the basis of the seventh digit of the account number. Account numbers in the 1% sample with a "5" in that seventh digit position are classified as 0.1 percent sample cases. Thus, the continuous work history sample involves three sample selections: the 20 percent sample, the 1 percent sample and the 0.1 percent sample.

5/ Unfortunately, social security account numbers are now being used for identification purposes where employment may not be intended. E.g., identification of bank account holders by social security numbers is required by Internal Revenue Service.

6/ Because of processing lag, most of the earnings from self-employment in 1961 will not be included until the following year's sample file is produced. It is estimated a total of 110 million living workers are represented in this file, of which 67.6 million were employed as a wage worker in 1961.

Limitations of the Data

In considering potential uses for the continuous work history sample (and especially for cohort studies), it is important to keep in mind the limitations of the data. One of the principal limitations is the inability to pinpoint, in some cases, the reasons for gaps in an individual's work history.

In 1937, the first year in which employment and earnings had to be reported, covered employment included employment by wage and salary workers under age 65 who were engaged in non-agricultural industry and commerce. This comprised roughly half of all paid employment jobs. Since 1937, coverage under the program has been expanded, by legislative enactment, until beginning with 1955 over 90 percent of persons in paid employment include persons in covered employment.7/ Records of employment and earnings for the groups covered in 1951 for the first time (such as the non-agricultural self-employed) are not available in the social security files for years prior to 1951. To generalize, since only covered employment is reported, persons working in consecutive covered and non-covered employments would appear to have gaps in their employment history.

Another problem relating to gap in employment is caused by unreported deaths. Currently, about two-thirds of deaths are reported to the Social Security Administration in connection with filing for survivor benefits and/or because of reports by undertakers. Unfortunately, it is difficult to estimate the degree of underreporting of deaths in the earlier years of the program.

Another area in which impact on the data cannot be measured accurately is the multiple account number situation. Workers with more than one account number create a sampling problem because they have a better chance of having one of their account numbers fall into the sample selection. This also contributes to the gap problem because the individual's employers may be reporting his employment and taxable earnings under different account numbers. Screening procedures are used to minimize the number of persons with multiple account numbers. A special study of the problem indicated that the number of workers in the work history sample was overstated by less than 1 percent.8/

During a given year another gap in the covered employment record may appear due to the regulations for reporting earnings up to but not in excess of the taxable limit. Thus, if a worker earns his limit in the 2nd calendar quarter of a given year, his employer will not report him in the third and fourth quarter tax returns even though he may continue his employment with that employer.

Suggested Types of Cohort Studies

Given the limitations described above, at least one type of cohort study is feasible--namely, the retrospective cohort. Starting with the definition of the cohort as workers in a given industry in a recent year, an analysis could be made of changes in employment and wage patterns over a series of prior years.9/ The cohort could be subdivided into age-wage-sex groups to permit analyses of patterns of homogeneous groups. Annual patterns could be studied for all groups, and quarterly patterns could be analyzed for groups which do not earn the maximum taxable wages.

Another type of cohort study may be made to analyze the mobility of one or more groups of workers. For example, if we were to define the cohort as workers in covered employment in both 1957 and 1962, data could be compiled for that period on the extent of employer turnover, the frequency and distance involved in geographic change, and the extent of change in industry. Again this could be done for selected age-wage-sex groups. From a cost and time point of view, this type of study would require much more of an expenditure than the first cohort study mentioned above.

Future Prospects

With the electronic computer becoming the accepted tool for data processing, cohort studies

---

7/ There is a technical distinction in coverage between legal and reported coverage. For example, there are 8.6 million persons legally covered in State and local government and in non-profit employment, but coverage is reported for 5.8 million of that total. The remainder are workers who are eligible for coverage on an elective basis, but for whom coverage has not been arranged. The net effect of this limitation is that current social security records are available for about 85 percent of persons in paid employment.

8/ See The Continuous Work History Sample Under Old-Age and Survivors Insurance in the United States of America, by B. J. Mandel, First International Conference of Social Security Actuaries and Statisticians, Brussels, November 1956, page 18.

9/ These wage and employment data are available back to 1937 for the 0.1 percent work history sample and back to 1951 for the 1 percent sample.

should become more feasible, and subject to fewer restrictions. For example, Internal Revenue Service plans to have most of their tax return records on tape in the not too distant future. Furthermore, there is a good possibility that these tapes may be made available to the Social Security Administration. Since the same method of identification--the social security account number--will be used by both agencies, many of the limitations of the continuous work history sample would be eliminated or minimized. We would know the individual's total instead of only the taxable earnings. We would know the extent of non-covered employment and would have better control over the multiple account and unreported death problems.

Summary

The continuous work history sample is a feasible source of cohort data, provided the limitations of the sample are taken into account. It should be possible in the future to combine for the same worker, data from tape record files of a number of government agencies such as Social Security, Internal Revenue, Census, VA, etc. When this statistician's Utopia occurs, we could truly say that just about all limitations to the data have been removed.

APPENDIX 1

EMPLOYEE STATISTICAL RECORD CONTENTS: 1962

Substantive informational contents of machinable, comprehensive, regular, multipurpose records

Item	1937-62 Workers			1962 Workers			
	All 1/			All 1/	Wage and salary		Self-employed
	Updated Summary 2/	1% CWHS	0.1% CWHS	0.1% Annual Advance Sample 3/	1% Employee-Employer 4/	0.1% Quarterly Advance Sample 5/	1% SE Sample 6/
Estimated completion date of record	10/63	12/63	2/64	12/63	4/64	a	5/64
Cutoff date of file (postings through)	9/63	9/63	9/63	9/63	9/63	b	3/64
Approximate number of records (in thousands):	149,000	1,490	149	73	1,020	66	65
With earnings	131,000	1,300	131	73	1,020	66	65
Without earnings	18,000	190	18	n.a.	n.a.	n.a.	n.a.
Maximum record length:	180	350	905	80	100	80	80
Nonsubstantive items	30	64	70	19	7	7	38
Substantive items 7/ (listed below)	150	286	835	61	93	73	42
Account Number Information							
1- Account number	9	9	9	9	9	9	9
2- Multiple account number indication (for multiple account numbers identified since 1946)	X	1	X		X		
3- Year of issue of account number		2					
Personal Characteristics							
4- Year of birth	2	3	3	3	3	3	
5- Month of birth	1	2	2		1		
6- Sex	1	1	1	1	1	1	
7- Race	X	1	1		1		
Insurance Status Information 8/							
Insurance status, each year:							
8- 1/1/51 - 1/1/59			9				
9- 1/1/60 - 1/1/63		4	4				
Indication of insurance status based on 1956 special provisions, each year:							
10- 1/1/57 - 1/1/59			3				
11- 1/1/60 - 1/1/61		2	2				
Benefit Status Information 9/							
12- Benefit status, each year, 1/1/60 - 1/1/63		4	4				
13- Year of death for workers not previously entitled to old-age benefits or year of entitlement for workers entitled to old-age or disability benefits		2	2				
14- Year of death of OAB 10/		2	2				
15- Benefit formula code (only for living workers entitled to old-age benefits as of 1/1/63 with yr/ent. 1960-62)		1					
16- Indication of "in claims status"	1	1					
Combination of family benefits of OAB's and DIB's in force status 11/:							
17- 12/31/61			2				
18- 12/31/62		2	2				
Amount of family benefits of OAB's and DIB's in force status 11/:							
19- 12/31/61			4				
20- 12/31/62		4	4				
PIA of OAB's and DIB's 11/:							
21- 12/31/61			3				
22- 12/31/62		3	3				
Indication of "in current pay status" of OAB 11/:							
23- 12/31/61		1					
24- 12/31/62		1					
Disability Information							
25- First year of beginning disability		2	2				
26- Last year of disability cessation		2	2				
27- First year of DIB		2	2				
28- Number of disability periods		1					
29- Indication of "freeze only" disability case and reason code		2					
30- Number of disability award periods		1					
31- Indication of disability status	X	X					
Years Employed Information							
32- First year employed	c 1	2	2				
33- Last year employed	c 1	2	2				
34- Last year employed before 1962		2					
Number of years employed:							
35- 1937-1962		2	2				
36- 1951-1962		2					

Item	1937-62 Workers: All 1/ Updated Summary 2/	1% CWHS	0.1% CWHS	1962 Workers: All 1/ 0.1% Annual Advance Sample 3/	Wage and salary: 1% Employee-Employer 4/	0.1% Quarterly Advance Sample 5/	Self-employed: 1% SE Sample 6/
Years Employed Information (Continued)							
37- Pattern of years employed, 1957-62		2					
38- Indication of continuous or intermittent employment, 1937-62		1	1				
39- Number of increment years, 1937-50		2					
Quarters Employed Information							
Pattern of actual QE, each year:							
40- 1951-1955			10				
41- 1956-1960	X		10				
42- 1961	X	2	2			2	
43- 1962	X	2	2	2		2	
44- Actual QE with $50 or more quarterly wages, 1962				1			
Deemed QE, 1962:							
45- Concept I - based on quarterly wages				1			
46- Concept II - based on all wages				1			
47- Concept III - based on all employment				1			
Quarters of Coverage Information							
Cumulative number of QC:							
48- 1937-1962		3	3				
49- 1947-1950	2						
50- 1951-1962	2	2	2				
51- Cumulative number of SE QC, 1951-62	2						
Number of QC, each year:							
52- 1937			1				
53- 1951-1962		12					
54- Number of SE QC, each year, 1951-62	12						
55- Number of farm QC, each year, 1955-62	8						
Pattern of QC, each year:							
56- 1938-1956			38				
57- 1957-1962		24	12				
Quarterly wage QC pattern 12/, each year:							
58- 1953-1960	X						
59- 1961-1962	8						
Type of Work Information							
60- Wage or SE indication, each year, 1951-54		4	4				
61- Farm or nonfarm wage indication, each year, 1955-62		16	16				
Farm or nonfarm SE indication, each year:							
62- 1955-1961		14	14				
63- 1962		1	1				
64- Military or other quarterly wage indication, each year 1961-62						2	
65- Coverage indication, 1962 (military wage, other quarterly wage, farm wage or SE)				2			
66- Dual coverage code 1962					1		
67- Minister waiver form indication		1					
Earnings Information							
Cumulative:							
68- Earnings credits, 1937-62		6	7				
69- Earnings credits, 1951-62		5	5				
70- Taxable wages, 1937-50	8						
71- Taxable earnings, 1951-62	8						
Taxable wages, each year:							
72- 1937-1950			56				
73- 1951-1960			50				
74- 1961			5			5	
75- 1962			5	5	7	5	
Taxable farm wages, each year:							
76- 1955-1961			28				
77- 1962			4	5	X		
78- Taxable nonfarm wages, 1962				5	X		
Taxable earnings, each year:							
79- 1951-1961	77	55	55				
80- 1962	7	5	5	5			

	Item	1937-62 Workers: All 1/: Updated Summary 2/	1937-62 Workers: All 1/: 1% CWHS	1937-62 Workers: All 1/: 0.1% CWHS	1962 Workers: All 1/: 0.1% Annual Advance Sample 3/	1962 Workers: Wage and salary: 1% Employee-Employer 4/	1962 Workers: Wage and salary: 0.1% Quarterly Advance Sample 5/	1962 Workers: Self-employed: 1% SE Sample 6/
	Earnings Information (Continued)							
	Quarterly wages, each quarter - each year:							
81-	1951-1960			160				
82-	1961			16			16	
83-	1962			16	16	28	16	
84-	Quarter in which wages of $4797 were reached, each year, 1961-62						2	
85-	Wages shown on SE schedule (1962) 13/							5
	Estimated nonfarm wages, each year:							
86-	1946-1955			50				
87-	1956-1961 (Method 1) 14/			30				
88-	1956-1961 (Method 2) 15/			30				
89-	1962 (Method 1) 14/			5		7		
90-	1962 (Method 2) 15/			5		7		
91-	Estimated earnings, each year, 1951-55			25				
	Self-Employment Information							
	SE taxable income, each year:							
92-	1951-61		44	44				
93-	(1962) 13/				4			6
	SE net earnings, each year:							
94-	1956-1960			30				
95-	1961		6	6				
96-	(1962) 13/							6
	SE industry:							
97-	1961		4	4				
98-	(1962) 13/							4
	SE State and county:							
99-	1961		4	4				
100-	(1962) 13/							4
	SE occupation (selected groups covered by 1954 amendments):							
101-	1961		2					
102-	(1962) 13/							2
	Indication of use of option for computing farm SEI, each year:							
103-	1959-1961		3					
104-	(1962) 13/							1
105-	IRD (1962) 13/							2
106-	Type of reporting form (1962) 13/							1
107-	Reporting year 13/							2
	Employer Information							
108-	Employer identification number					9		
109-	Establishment number					4		
110-	State and county code					4		
111-	Industry code					4		
112-	Class code					1		
113-	Type of reporting form, 1962					1		
114-	Coverage group code 1962					1		
115-	Number of wage items, 1962					2	2	
116-	Number of wage items, each quarter, 1962						8	
117-	Collector's stamp, 1962					1		
118-	Referendum code for State and local governments					1		

Footnotes

a - Quarterly advance sample completion dates:
1st quarter - 11/62   3rd quarter - 5/63
2nd quarter - 2/63   4th quarter - 8/63

b - Quarterly advance sample cutoff dates:
1st quarter - 9/62   3rd quarter - 3/63
2nd quarter - 12/62   4th quarter - 6/63

c - For years 1937-50, record shows only that year was sometime during that period.

1/ Only about 75% of 1962 SE will be included.

2/ Number of records shown is for entire file. A 20% sample is tabulated for statistical purposes.

3/ Punch card file.

4/ Includes quarterly and annual wage workers only.

5/ Punch card file. Only for quarterly reported workers. File contains quarterly data for 1961 and 1962; when 1st quarter 1963 data is added, 1961 data is dropped.

Footnotes (Continued)

6/ Punch card file. Delinquent items in this file will be used to revise previous years' data.

7/ Certain fields contain more than one item of information. An "X" indicates that the item is coded in some other field.

8/ Insurance status codes relate to law in effect on given date:
1/1/51 - 1/1/60 = 1:2 law
1/1/61 = 1:3 law
1/1/62 - 1/1/63 = 1:4 law

9/ Benefit status information relates only to claims based on the account number in the record. Auxiliary information is not reflected in the benefit status codes.

10/ Individual years 1958-1962 are shown. Prior to 1958 there is only an indication of 1957 or prior.

11/ Information for 12/31/62 based on family benefit record of beneficiaries in force status 12/31/62 for living workers entitled to old-age or disability benefits as of 1/1/63. Information for 12/31/61 carried over from previous year's CWHS, with no revision for delinquently reported claims actions.

12/ For years prior to 1962, for each quarter that is not a QC, there is also an indication of whether earnings were:
$0
<$50 or a quarter in which there were debit earnings and a credit has been applied (does not show whether any earnings remain)

For 1962, for each quarter that is not a QC, there is also an indication of whether earnings were:

$0	20-29
01-09	30-39
10-19	40-49

Revision for delinquent items not complete for years 1953-55.

13/ Year will be other than 1962, if delinquent item in 1962 SE Sample. Informational items for delinquent reports in the SE sample refer to the reporting year coded in the record.

14/ Estimated nonfarm wages - Method 1:

For years 1956-58

Quarter taxable limit reached	Estimated wages assigned
1st	$16,800
2nd	12,600
3rd	7,000
4th	4,900

For years 1959-62

Quarter taxable limit reached	Estimated wages assigned
1st	$19,200
2nd	14,400
3rd	8,000
4th	5,600

15/ Estimated nonfarm wages - Method 2:
Consider as the "limit quarter" the one in which the taxable limit is reached. Determine the wages in the next prior quarter that shows as much or more wages than the "limit quarter." Substitute those wages for the "limit quarter" and all subsequent quarters in the year. If there is no prior quarter with wages greater than the "limit quarter," use the "limit quarter" wages for each following quarter. The total of the quarterly wages thus computed represents the annual estimated wages.

Department of Health, Education, and Welfare
Social Security Administration
Division of Research and Statistics
Statistics Branch
Statistical Processing Methods Section
June 7, 1963
Revised June 24, 1963

DISCUSSION

Herman P. Miller, Bureau of the Census

Mr. Hearn is one of a small but valiant group of men who have spent many years pushing the use of administrative records as a source of general statistical information. As far back as 1947 Saul and others like Jacob Pearlman, Ben Mandel and Irwin Wolkstein have been displaying the statistical wares of the Social Security Administration and urging other government agencies to partake of them. Much has already been accomplished as a result of their efforts, but I am sure that what we have seen is only the beginning. Social Security records are now being used to produce a variety of general purpose statistics such as County Business Patterns. Efforts are also being intensified to increase the use of Internal Revenue Service tax records to produce general statistics. I have no doubt that about 10 years from now, when the IRS files are fully automated, we will see statistics produced each year based on an amalgamation of data from the files of the Census Bureau, the Social Security Administration, and the Internal Revenue Service. At that time, men like Saul Hearn will be able to say with some satisfaction, "I told you so."

The paper that Mr. Hearn has given us today is in the tradition of those given on the same subject in the past. It contains a straight forward factual description of the way in which the OASI work history sample is selected and a very brief discussion of some of the limitations of the data based on that sample. The difficulty that I find with this kind of paper is that there is not much to criticize in it, nor is there much to praise. The story Hearn tells is a good one, but I have heard it before and no longer find it very exciting. This does not mean that it is not worthwhile. I am sure that when the day comes that I want to use this sample, I will be very happy to have the description that was presented today. My chief criticism is that it contains too much description and not enough analysis.

Near the end of his report, Hearn mentions some of the limitations of the data based on the work history sample. Although everything he says is correct, I feel that he passes too lightly over some defects that turn out to be major obstacles to analysis. For example, he states that the limited coverage of OASI in the early years is a problem, but he reassures us by stating that over 90 percent of the workers are now covered. Farber's analysis suggests, I think, that this overall figure may be misleading when the data are used to examine trends for the low-income groups. I shall say more on this point later. The relatively low upper limit of taxable wages is also stated as a limitation by Hearn. Here again it is only when the data are put to use, as they are in the Farber study, that we can see how very serious this shortcoming can be. Incidentally, here is point where I think Hearn could have made a real contribution with relatively little effort. There is no need in analytical studies involving the use of OASI data to be saddled with the limitation imposed by the failure to count wages above the taxable limit. There are methods for extrapolating taxable wages above the upper limit to represent full year totals. My understanding is that these estimates are fairly good. If they are not good, we should be told about it; but if they are good I would like to know why they are not being used. It is not very meaningful to talk about high-wage workers whose earnings start below the median.

I believe that Hearn's paper would have been a much more interesting and useful document if he had taken the time to explore somewhat more fully the implications of the procedures he has used instead of confining himself as he did to factual description. Let me cite one more case in point before turning to the Farber paper. In the last three sentences of his paper, Hearn mentions the possibility of matching OASI data with IRS records. In less than a decade this possibility should become a reality. A matching study of this type would open up large areas for analysis and would eliminate some of the defects that now exist in the wage records. I am sorry that he did not say more on this subject, particularly since I am now working on the problem of using administrative records at the Bureau of the Census.

Dave Farber deserves much credit for presenting in great detail one kind of use that can be made of the work history sample. I am sure that a vast amount of work has gone into this paper as well as considerable statistical dexterity and imagination. He has created a set of statistics which, I am sure, will receive considerable use in years to come. Even more important are the concepts and procedures he has developed.

Much as I credit Farber for the work he has done, I cannot say that I agree with his interpretation of the results. I found a disturbing naivete in his analysis and a tendency to accept things too much at their face value. The results are often treated as though they represent the entire universe instead of a segment of it and insufficient attention is paid to the restrictions that are imposed on the data because of the limited nature of the universe. There is also a tendency to push the figures too hard and to make generalizations that are based on small differences. In my discussion of several specific aspects of this paper, I shall confine my attention to only a few points that I consider fundamental. The fact that I am first at the trough does not give me the right to drain it. I am sure the other discussants will stress other points and that they will have more to say about some of the things that I will only touch upon.

Farber's major purpose in writing this paper was to examine changes in the level of wages for different groups of workers during the years 1951-57. Since this period includes the 1954 recession, he pays particular attention to changes during that period. Although the analysis was made for a dozen different age groups, age does not appear to be his chief concern, as the title suggests. After all, workers don't age very much in a seven-year period. I think age is really used in this study more as a way of identifying cohorts than as a method for analyzing changes over time. But, this criticism is perhaps only incidental.

The universe used in the Farber study are those workers who had some covered employment in 1957 and who had only wage credits in earlier years. He traces the work histories in covered employment of these individuals back to 1951 and divides them into four groups based on covered wages per year employed: under $1,200 (low-paid); $1,200 to $2,400 (intermediate low-paid); $2,400 to $3,600 (intermediate high-paid); and $3,600 or more (high-paid). If we stop here and think for a moment about what he has done, we can detect several of the basic defects in this study.

First of all note that the study is restricted to persons in covered employment in 1957 and their work histories in covered employment in earlier years. Because of this limitation we cannot measure changes in the level of covered employment based on this sample nor can we say anything about changes in total employment even for this panel of workers. Yet, Farber attempts to make generalizations on both of these phenommena on the basis of his sample. For example, he converts the number of workers in each cohort with covered employment in each year into an employment index and makes generalizations based on that index. This does not seem to me to be a valid procedure. Let us consider, for example, his findings for the low-paid cohort of men aged 35-39 years in 1957. There were a total of 419 men in the sample. All of them had covered employment in 1957 but only 241 (58 percent) had covered employment in 1951. On this basis Farber says the employment index rose from 100 in 1951 to 174 in 1957. It did no such thing. All he knows is that 40 percent of the men in this cohort who worked in 1957 had no covered employment in 1951. I find it hard to believe that so large a proportion of the men in this age group really did no paid work at all during the year. After all, they were on the average about 31 years old at the time. I suspect that in the low-paid group there is a considerable amount of uncovered employment that does not get reflected in the figures. There may also be reporting or processing errors of one type or another. Whatever, the reason, a good deal of qualification is needed for Farber's conclusion that there is a great deal of intermittency of employment in the low-paid group. I am sure he is right, but I don't think these figures prove the case.

The limitation of taxable wages to $3,600 for 1951-54 and to $4,200 thereafter also creates serious problems of interpretation and, I believe, has led Farber to some erroneous conclusions. He finds, for example, that "for the high-paid male cohort, the average annual wage credits .... remained stable throughout the 1951-57 period." Little wonder that they did. The minimum needed to qualify for the high-paid category was an annual average of $3,600 in covered wages; but no count was made of wages above $4,200. Therefore, there was no way in which the average could rise substantially because wages above $4,200 were not included. There could have been significant increases in the total wages of this group that would not be reflected in the statistics. This error in interpretation leads Farber to the erroneous conclusion that there was a narrowing of wage differentials between the highest paid and the two intermediate cohorts. All that happened here, in my opinion, is that the average for the two lower groups rose, as might be expected during a period of relatively full employment like 1951-57. But, because of the income limitation, the income of the top group did not rise. As a result there was an apparent narrowing of wage differentials. As I mentioned earlier, there is a way to correct this defect in the data and it seems to me that it should receive serious consideration in the future and perhaps be carried back to the earlier years if possible. Procedures have been developed for extrapolating covered wages to represent annual totals. Assumptions are involved in preparing such estimates but the end product in my opinion would be more useful than the unadjusted totals. Such figures would really permit us to make some generalizations about wage trends for the higher-paid workers which cannot, in my opinion, be made on the basis of the present data.

Before concluding, I should like to say that it would be interesting to speculate some time about what we would do with the work history data, even if they had no limitations, that we cannot now do with the cross-section data. This is not intended as a criticism of Farber since he limited his discussion to a consideration of wage movements for a relatively short period. I have thought about this matter from time to time as I worked with the cross-sectional census results, but I have never come up with a clear-cut answer. I have a feeling that some concentrated thinking along these lines might pay handsome dividends.

In conclusion, I should like to say that I have tried to be fair in my criticism of these papers; but I have not worked very closely with this set of numbers and I could be mistaken in some of my points. If so, I am sure Farber or Hearn will let me know about it. I also hope that I have not created the impression that I did not enjoy these papers or that I do not think that the work history sample or Mr. Farber's use of it are unimportant. I think they are important and useful and they should be continued. I must confess, however, that as I prepared these comments, I asked myself if I would undertake a study of these data, were I requested to do so. The answer, I am sorry to say, was "no" because I could see too many shortcomings. I suspect that I am not alone in this view. This attitude will, I hope, provide some food for thought for those whose responsibility it is to keep this project going. If the limitations of the data are so great as to frighten away prospective analysts, it may be worthwhile to give some thought to improvements that may be made. I see little hope for such improvements for the lowest-paid workers, but perhaps more can be done for the top income groups.

DISCUSSION

Jacob Perlman, National Science Foundation

I am very happy to appear as a discussant at this meeting. As you know, I am also a former employee of the Bureau of Old Age, Survivors and Disability Insurance. In fact, it was in this capacity that I was instrumental, during the early forties, in setting up the continuous work history sample, which is based on the wage records of employees covered by the program.

The continuous work history sample was originated primarily to meet the needs of the Old Age and Survivors Insurance Program. It was to be used for two purposes: first, to provide the statistical basis for the actuarial studies, which were needed in connection with the program; and, second, to provide statistical data to check the existing provisions of the law, as well as to test any new provisions that were suggested for its modification. The provisions of the Old Age, Survivors and Disability Insurance system are geared to continuity of covered employment. This applies to both the eligibility provisions and the formula used in computing monthly benefits. This continuity of employment is reflected by the wage records contained in the continuous work history sample tabulations.

In addition to using the data from the continuous work history sample for analyzing the OASDI program, it was apparent that the tabulations might be useful for general analysis connected with patterns of earnings and employment. The continuous work history sample was set up at a time when the depression of the thirties was still fresh in the minds of people, and it was felt that one could use the data in a manner that would make possible some generalizations pertaining to earnings and employment as affecting the total labor force in the United States.

As Mr. Hearn has pointed out in his excellent paper, in terms of uses of the data for general analysis, there were many limitations inherent in the original tabulations of the continuous work history sample. One drawback was due to the restricted coverage under the OASDI system, which in the early years fell considerably short as regards total employment of members of the labor force. Another factor was the limitation placed on the taxable wages, the maximum of which originally amounted to $3,000. There were still other limitations, such as the tendency on the part of some workers to obtain multiple account numbers, as well as the difficulty of eliminating from the sample persons who died without becoming eligible for benefits.

However, there were some in the Social Security Administration in the early days who had confidence that sooner or later these limitations would decrease in importance, due to changes in the law and improvements in the administration of the system. This faith has been more than justified in later years. Thus, the coverage of the present law extends to nearly the entire labor force, the only important exceptions being Federal employees on a permanent basis and members of the medical profession. Likewise, the limit on taxable wages has also been changed, the maximum figure now being $4,800. Finally, administrative improvements have minimized the problem of multiple numbers and the lack of reporting of deaths.

As a result, the data from the continuous work history sample have become more adapted for use in connection with the general analysis of wages and employment. This situation could be improved even more. For example, I am told that the Federal Civil Service System is now planning to develop a similar sample covering Civil Service employment, and it should be possible to coordinate data from the OASDI program with those from the Civil Service system. Likewise, it is possible by means of estimates to make allowances for the $4,800 limitation on taxable wages under the OASDI law. Clearly, as the years go by, the figures in the continuous work history sample will reflect more and more the complete cycle of employment and earnings of individuals in the labor force, beginning with the time the individual enters employment until he retires or dies.

One of the weaknesses of the Federal statistical system is that there is too much emphasis on the collection of "new" data and not enough on the utilization of figures that are a by-product of existing administrative systems. I am glad to know that the Census Bureau is now taking specific steps to investigate the uses and adaptation of such by-product data, as exemplified by those obtained from the Internal Revenue Service and the Bureau of Old Age, Survivors and Disability Insurance, as a basis for providing useful information that can be obtained without resorting to new statistical surveys.

It should be pointed out that in order to make the best possible uses of by-product data emanating from an administrative system, one has to exercise great skill in setting up the tabulations, so that they will minimize their inherent limitations and meet specific analytical objectives. Mr. Farber's paper is a good example of such an effort, and he has done an admirable job in pointing out the extent to which, for various categories of the covered labor force under OASDI, it is possible to develop general conclusions as to the effect of the short depressions of the fifties on the earnings of workers. To be sure, his analysis has to be considerably qualified, as he has done in his paper. It should be remembered, however, that even statistics obtained from specific surveys of collected data are also subject to limitations, which the analyst needs to consider in reaching conclusions. The job of the analyst in the social sciences is a difficult one, whether it is applied to "collected" data or to figures that are a by-product of an administrative system, and it behooves him under any circumstances to be on the alert and cautious in reaching conclusions. Nevertheless, much can be done in this connection to formulate generalizations relating to the socio-economy.

In conclusion, let me emphasize again that the data emanating from the continuous work history sample are one of the best sources of information that can be used in economic analysis dealing with life patterns of earnings and employment, as well as cumulative earnings of individuals during their working life. Such figures are extremely important in analyzing not only the problems related to the OASDI system but also those that have a bearing on the functioning of the total economy.

DISCUSSION

Murray S. Wernick
Board of Governors of the Federal Reserve System

I agree with most of Mr. Miller's criticisms. But on one critical point I do wish to take issue with him before I present my own comments on Mr. Farber's paper. Mr. Miller stated, in effect, that if the choice were his, he would have answered "no" to the question whether cohort analysis should have been undertaken using OASI data. My answer would have been "yes".

Longitudinal analysis -- taking a selected group of workers and following their labor market activity over a period of time is relatively new and hardly tested technique in labor market analysis; but one which I am sure will expand rapidly. How a person, family or an identical group of workers adjusts to ever varying demands for labor, shifting technology and a changing wage structure, during their life cycle, can in large part only be satisfactorily answered by continuous work histories obtained as events take place. Attempts to duplicate this process by means of recall on the part of a respondent has not been very successful because of the very short time span that can be recalled accurately. Cross sectional analysis presents obvious difficulties in tracing labor market experience over the life cycle of identical groups.

Mr. Farber's spade work has been an important contribution to our better understanding both of the possibilities and the hazards of longitudinal analysis. In his imaginative use of wage and age cohort changes for the period 1951-1957, he has uncovered some rather fascinating relationships of aging to wage and employment trends. But the study also has some serious drawbacks which in its present form limits its use in economic analysis.

1. Changes in employment and wages for each age and wage cohort reflects a mixture of different influences -- aging, short run economic fluctuations and administrative procedures. The mixture varies considerably for each one of the age and wage cohort groups. Since it is impossible for the user to separate the data into these basic components, one is never certain of the extent to which the index for any given cohort reflects changes in wage rates, aging or data deficiencies. This is most apparent in the younger cohorts. The sharp rise in participation rates for the youngest cohorts and up-grading of wages, as a youth changes his status from a part time secondary worker attending school to a full time primary worker swamps the influences of all changes which may have been caused by economic factors. We find that for example, the wage index for the 20-24 year old worker rose to 1193.0 from 1951-1957. Did this group benefit from general wage rate increases which amounted to about 30 per cent in the period? How much were each of the other cohorts effected by sharply rising wage rates? Another example of lack of uniformity among cohorts is found in the highest paid male cohort where administrative limitations on maximum covered wages makes it difficult to determine the influence of economic developments. The stability in average wage credits seems entirely inconsistent with what we know was happening to wages in the period, especially for the full time worker. Knowing this obvious weakness in the high paid male cohort, it is difficult to interpret the significance of the narrowing of the wage differentials presented on Tables 7 and 8 based on the relation of low-paid cohorts to the highest paid cohorts.

2. Although Mr. Farber assures us that the results of his cohort analysis are consistent with other available data, no evidence is given to substantiate this claim. A cursory examination indicates that the increase in employment, i.e., the number of workers in the sample reporting wage credits was 41.1 million in 1951 and 61.2 million in 1957 an increase of 20 million or 50 per cent. As might be expected from the selection of the sample of workers employed in 1957, the rise is substantially higher than indicated in the BLS establishment series or the BLS household series -- BLS nonfarm employment rose by 5 million or only 12 per cent. I assume with some effort and work the employment series can be roughly reconciled. Reconcilation would be invaluable in determining how many workers with earnings experienced drop out of the cohort sample and their effect on employment and wage credit trends in any year.

Since the cohort data have potential value as a measure of change in income, there is also need for some systematic comparison with the data for total and median income of wage and salary workers obtained from the Census Bureau. Some knowledge of the relationship between wage credits earned and with average earnings of employed workers reported in BLS series would also be extremely helpful.

3. I raise the question of reconcilation with some concern because it appears to me that there is very substantial understatement of the impact of the 1954 recession on employment and wages in the cohort data. A major purpose of this paper, Mr. Farber says, is, "to pay particular attention to the 1954 recession and its differing effects on the wage credits of the lower and higher paid male and female cohorts". Mr. Farber using the Woytinsky quote also sharply criticizes the use of cross sectional wage data because it fails to adequately take into account the impact of employment declines on workers' earnings in a recession.

Yet, what is perplexing is that in adding employment of each of the male and female wage cohorts together I find total employment in the sample advanced from 42.9 million in 1953 to

43.9 million in 1954 -- an advance in average annual employment of one million in a recession year. This rise in employment in 1954 is in sharp contrast to all our other employment data for 1954. The BLS private nonagricultural establishment series showed a decrease of 1.4 million. Total nonagricultural employment in the household series, declined by over 1.0 million; while female nonfarm employment also showed a decline of 300,000 in 1954 compared to a rise of 600,000 shown in the cohort sample.

Since one of the major determinants of a recession is a decline in overall employment, I am confused as to how it is possible to better evaluate the impact of declining employment on wages among the cohorts when in fact employment rose in a recession year.

This rather large employment increase shown in the cohort sample in a recession year would seem to imply:

a. The current sample may not be representative of year to year changes in employment, and therefore of the economic impact on wages of declining employment during a recession.

b. Average annual employment data do not adequately reflect short term layoffs -- a typical occurance during recessions. The decline in wages from such short layoffs will be reflected partly in the wage cohort of employed workers but will not show an employment impact in the average wages of the entire cohort.

c. Persons who become and remain unemployed or are forced to retire because of a recession disappear from the sample. This group can be sizable especially in the older worker cohorts. Once older workers become unemployed they are the least likely to find other jobs and a substantial number of the long term unemployed probably never return to the sample.

In any event, I do think more explanation is needed, if the analysis of the data is to support the underlying contention of the author that cohort data give the best indication of the effects of declines in employment on incomes in recessions.

4. I also think that the selection of the cohorts based on employment in 1957 is open to some question. The selection has definite advantages for the young age groups where new entrants come into the labor force and participation rates rise rapidly. Employment in 1957 is thus likely to give undue weight to growth in employment and wages for the younger cohorts. When we get to the other end of the age scale, where the sample represents only a small group of older workers who have managed to continue their attachment to the labor force beyond retirement age, the pattern of employment and wage credits is hardly representative of the large and important group of older workers who withdrew from covered employment as they aged between 1951-1957. Were those who retired in higher wage cohorts than those who remained? Unless we know the answer to this question, the movement in employment and average wage credits for the older worker has only very specialized significance for the few selected who worked in 1957.

On net, I would think that the sample selection also results in a downward basis in the earnings reported by wage cohorts. For example, over 70 per cent of those with average earnings of less than $1,200 and over 50 per cent of those whose earnings were between $1,200-2,390, from 1951-1957 were under 30 years of age. Yet this group accounts for only 1/3 of the number employed in 1957. Many of these young people obviously were not separate spending units and made relatively minor contributions to family income between 1951-1957. Many in the low income groups had voluntary part time jobs and such low income is not necessarily indicative of distress or hardship. In fact, unless some distinction is made between primary and secondary family earners, the cohort distribution by high and low incomes has very limited usefulness. The failure to include the income of workers who had retired or died but earned incomes between 1951-1956 probably causes a further downward bias.

My conclusion would be that the cohort data used in this study are extremely tricky for economic analysis. Changes in average wage credits and employment from cohort to cohort over time have a different meaning for each cohort. While the differences in the cohorts can be compared arithmetically, explaining the significance of the differences is an analytical process. And it is in the latter area, in my judgement, that the next steps must be taken if we are to take full advantage of the vast amount of very useful data being made available to us by Mr. Farber.

X

CONTRIBUTED PAPERS - I

Chairman, Felix E. Moore, University of Michigan

AN EVALUATION OF AN OPTICAL SCANNING FORM FOR A MAIL SURVEY

M. Delaney Eldridge
National Institute of Mental Health

A relatively untried questionnaire technique was utilized in a recent survey of professional personnel in approximately 2,500 mental health establishments. Forms were specially designed to be "read" by a high speed optical scanner with information contained thereon transferred directly onto magnetic tape, thus eliminating traditional coding and card-punching operations. Prior to the full-scale survey, an extensive pretest was conducted late in 1962 in a probability sample of mental health clinics, hospitals for the mentally ill, and institutions for the mentally retarded. The purpose of the pretest was to evaluate:

1. Respondent reaction to the form
2. Respondent ability to utilize the form.
3. Clarity of questions and definitions.

The results reported here deal with the first two aspects of the pretest -- the evaluation of the optical scanning form for the mail survey.

BACKGROUND

The system utilized in this survey is based on a dot-matrix arranged for alpha or numeric answers. The answer sheet is completed by filling in, with a standard lead pencil, one or more dots in predetermined segments of the matrix. A high speed photoelectric document reading machine scans one or both sides of the marked sheets simultaneously and converts the marks to electronic impulses. Response forms, or questionnaires, while based on the standard matrix, are specially designed for each application.

The system was originally developed for test scoring and has been used effectively in that application in all levels down through the third grade. However, there were no data available to demonstrate its effectiveness in survey applications, and it was therefore concluded that a pretest should be conducted and the results reviewed critically.

DESIGN OF QUESTIONNAIRES

Two questionnaires were developed. An attempt was made to keep the questions to a minimum rather than to utilize the full matrix capacity which is the equivalent of 13 punch cards. Administrative data, available in the business office of the establishment, was to be supplemented by personal data available only from the individual professional staff members.

The Establishment Schedule (Figure I) was designed to be completed in the administrative office of each establishment and to provide the following summary data:

- 1-2. Establishment name and address.
- 3. Total number of employees on the payroll.
  - a. Full time
  - b. Part time
- 4. Number of mental health personnel.
  - a. Psychiatrists and other M.D.'s
  - b. Psychologists
  - c. Social Workers
  - d. Professional Nurses
- 5. Number of personnel forms returned.

All items on this form, with exception of street and city address, were designed for coding by the respondent. Space was provided on the form for the respondent to write in the complete name and address of the establishment.

The Personnel Schedule (Figure II) was designed to be circulated by the administrative office to the appropriate professional staff. The number of completed Personnel Schedules was entered in Item 5 of the Establishment Schedule. This number equaled the sum of Item 4 on that form if each professional staff member within the scope of the survey completed a Personnel Schedule. The following items were included:

1. Name
2. Professional classification
3. Professional society affiliation
4. Most advanced level of education
5. Multiple employment(for subsequent matching operations)
6. Personal data
   - a. Date of birth
   - b. Sex
   - c. Type of citizenship
7. Years' experience
   - a. In present psychiatric or mental health specialty
   - b. In non-psychiatric or non-mental health activity
8. Number of hours employed by activity type during a typical week in the establishment
   - a. Total
   - b. Patient care and/or clinical service
   - c. Administration
   - d. Teaching
   - e. Research
   - f. Staff consultation
9. Months of approved residency completed by psychiatrists or psychiatric residents

All items on this form were designed for coding by the respondent with the exception of the identification of other employers, which was treated as a write-in and used to assist in the identification of multiple returns from one respondent.

PRETEST SAMPLE DESIGN

Two hundred two sample units were selected for the pretest from the universe of approximately 2,500 establishments stratified by State and type of institution. Each selected establishment was treated as a cluster with 100% enumeration of the professional staff within the cluster. Each establishment received one establishment form and a number of personnel schedules sufficient for distribution to every professional staff member within the scope of the survey. (A post card was enclosed for reorders in case the number of personnel schedules originally sent was insufficient.)

A random subsample of 15 establishments also received pretest evaluation sheets (Figure III) designed to assist in the evaluation of the respondent's reaction to the form. Each person employed in a subsample facility was asked to fill out the evaluation sheet after completing the Personnel Schedule. One hundred sixty-eight establishment schedules and 4,079 Personnel Schedules were available for analysis. A summary of the returns is as follows:

	Establishments	Professional Personnel
Total	202	—
Schedules Returned	168	4,079
Evaluation Sheet Subsample	(15)	(308)
Closures or Out-of-Scope	15	0
Mergers	1	0
Non-respondents(1)	18	-

ANALYTICAL METHOD

The analysis of the pretest was carried out in two stages:

1. Respondent reaction - a tabulation and analysis of the responses on 308 pretest evaluation sheets returned by the subsample of establishments.
2. Execution of the Schedule - a critical editing of selected items on 168 Establishment Schedules and 4,079 Personnel Schedules for discrepancies in the recording process.

In order to complete a schedule "without error," the respondent was required to conform to the following specifications:

All Items - Use only a lead pencil and blacken no more than one circle in each column.

The Alpha Grid

1. Use no more than eighteen letters for the name. Abbreviate wherever necessary and print the name in the row of boxes provided.
2. Leave a blank box between names or abbreviations.
3. Blacken the circle containing the letter printed in the box below.
4. Blacken the blank circle at the top of each unused box.

(1)Returns received after the cutoff date established for this preliminary analysis - a later cutoff date was set for final survey tabulations.

Numeric Grids

1. Enter the appropriate numeric answer in the box provided beneath the grid using the extreme right column as the unit position.
2. Blacken the circle containing the number printed in the box below the column.
3. Blacken zero circles at the top for each unused box.

Errors have been categorized into three groups: critical, non-critical, and permissible discrepancies. Critical errors are those replies which cannot be resolved on the basis of information contained on the questionnaire. Non-critical errors are those which require some editing prior to passing the document under the reader, but the necessary information is available elsewhere on the form. Permissible errors are those which do not follow the instructions in every respect but are not critical in terms of the "answer" to be read by the optical scanner.

ANALYTICAL FINDINGS

RESPONDENT REACTION

The responses to the pretest evaluation sheets are summarized for each professional dissipline in Tables 1 through 5 and are discussed in detail below. All of the observed differences between disciplines were sufficiently small to be attributed to chance, with exception of the response to Question 4 as indicated in Table 5. (The response to Questions 4 and 6 have relevance only to the content of the questionnaire for the full scale survey and will not be discussed here.)

Question 2 - "Is the questionnaire convenient to use?"

It was concluded that the form did not require an unreasonable effort on the part of the respondent. Eighty-one percent found the form convenient to use. An additional 6% were uncertain, and only 13% found it inconvenient. See Table 1.

Question 3 - "Are there any parts which are not clearly stated or which you do not understand?

This question related to more than just the form-design. Nevertheless, 79% of the respondents believed that they fully understood the form and had no suggestions for clarifying it. (We do not necessarily share their optimism.) See Table 2. Eighty percent of the comments dealt with the form content, definitions, and/or instructions for routing the forms back to us, etc., rather than the form design. Examples of some comments relevant to the mark sensing form are:

"Takes time to understand."

"Being a new idea, it seems more difficult. Probably if I had to fill the same way next month, I think it would be easier."

"An excellent schedule - well designed, easily understood and if it will be used for IBM mark-sensing - extremely efficient."

"I feel now you are eaten up by machines."

"Some confusion related to the dots and blank spaces. May have been due to not reading the instructions carefully."

"The questionnaires seemed clear."

"I only wish the instructions (if possible) were more brief and in larger print."

"This is a machine-oriented form which humans like myself resent."

Question 5 - "How long did it take you to fill out the questionnaire?"

We believe that the definitions and instructions covering the routing of the forms could have been simplified. Despite this shortcoming, 84% of the respondents in the pretest evaluation sub-sample reported that they took less than 20 minutes to complete the form. Almost half of the respondents took less than 10 minutes. See Table 3. It is unlikely that a conventional questionnaire could have been completed in less time.

EXECUTION OF THE SCHEDULE

One hundred sixty-eight Establishment Schedules and 4,079 Personnel Schedules were edited for deviations from the specifications discussed above. In addition to the alpha grids, Items 3 and 4 on the Establishment Schedule and Items 3, 6, 7 and 8 on the Personnel Schedule were critically reviewed. The items selected for editing were those where consistency checks were possible or where a non-response problem was anticipated.

The Alpha Grid

Table 6 shows the number and percentage of schedules having errors in the name grid. There were two error types classified as critical errors: a) blanks, and b) incomplete codes accompanied by incomplete write-ins. Strictly speaking, these are not critical errors on the Establishment Schedules if the respondent completes the name and address box elsewhere on the form. However, they are critical on the Personnel Schedule and we attempted to analyze the two alpha grids in the same manner.

It was originally feared that the respondents would resist the completion of the name grid since it involved coding a fairly complex matrix. Our fears were unwarranted. Only two Establishment Schedule grids (1%) were either blank or incomplete, and 3 out of 4,079 (0.1%) of the grids on Personnel Schedules were blank. In some instances the failure to complete the name grid on the Personnel Schedules may represent the respondents' refusal to divulge their names rather than any resistance to the coding technique.

A better indication of the acceptance of the alpha matrix is the very low frequency of written responses accompanied by a blank code matrix. In these few instances (one percent of the Establishment Schedules and 0.4% of the Personnel Schedules) the respondent did not object to providing the information in the conventional write-in but did fail to transfer the information to the code in the alpha matrix. This "error" can be picked up readily in the editing process since the entire alpha grid remains blank. A few more schedules showed an incomplete code but a complete write-in. See Table 6.

The error that is somewhat harder to find in an editing process is the discrepancy between the code and the write-in. Discrepancies were found in almost 2% of the Establishment Schedules and 1% of the Personnel Schedules. In all cases the write-in appeared to be correct. The errors resulted from the transposition of columns or an attempt on the respondent's part to further abbreviate in order to reduce coding time.

The most common error observed was the omission of the zero mark between names. This occurred in roughly 12% of the Establishment Schedules and 6% of the Personnel Schedules, indicating that the importance of the space code may not have been highlighted sufficiently in the instructions. Fortunately, the omission of the space code is not critical. If the specific program requires a space code, a routine spot-checking procedure can be established to identify the errors.

The Numeric Grid

The results of the edit of numeric fields are summarized in Table 7. It is interesting to note that the error rates for the two items on the Establishment Schedule are consistently low while those for the items on the Personnel Schedule are higher and more variable. Since the same coding technique was used for all the items, this difference must be due to some other factor. Two possible explanations come to mind. It could be due to different respondents, the hospital administrator vs. the professional staff member. Or it could result from inherent difficulties encountered in the items themselves. In light of the comments on the pretest evaluation sheet, the second explanation appears to be the more plausible. The Establishment Schedule figures (where the error rates are low) on total payroll (Item 3) and number of professional staff members (Item 4) were readily available and straightforward. The personal information on date of birth (Item 6), years of professional experience (Item 7) and current assignment by type of activity (Item 8) were not so easily obtained. It is not unusual to encounter persons who are reluctant to report their date of birth. There was a typographical error in the heading of Column 7B which made the item practically unintelligible unless the respondent read the detailed instruction. Many respondents asked for definitions of the activity type categories used in Item 8. We conclude that the error rates on the Personnel Schedules are confounded with difficulties resulting from these problems and cannot be attributed solely to the specially designed form. In the absence of any other information, it is reasonable to assume that the errors on the Establishment Schedules are representative of what one might encounter under typical mail-survey conditions.

Critical errors were practically non-existent on the Establishment Schedules. Incomplete, or blank codes were provided on approximately two percent of the items. A discrepancy between the code and write-in was observed twice on the total number of employees on the payroll and once on the professional staff count. The zero space to the left of the last data field was left blank on 4% of the schedules for Item 3 and 3% of the schedules for Item 4. Fortunately this omission is not a critical one. The write-in was either by-passed completely or partially, and the complete code was entered directly on 6% of the schedules for Item 3, and 4% of the schedules for Item 4.

As mentioned above, the analysis of the error rates for the items on the Personnel Schedule is of limited value within the context of this paper. However, certain recurring errors are worth mentioning. In order to save space, we asked for only the last two digits of the year of birth (Item 6). Thus 1901 would be written "01", 1898 would be "98", etc. A few respondents attempted to enter the four digits in the two coding columns. In many of these cases the year of birth was garbled beyond the point where an intelligent judgment could be made as to the correct year of birth. The code did not agree with the write-in on 1.4% of the schedules.

Thirty-four percent of the respondents to the pretest evaluation sheet indicated that they had some difficulty allocating their hours employed by type of work (Item 8). An additional 2% stated that they had great difficulty. It is therefore not surprising to learn that the critical error rate for this item is close to 7%, or that in 2% of the schedules the component hours did not equal the total hours reported in Column A. The relatively high frequency of direct code entries, i.e., codes with incomplete or no write-ins, probably reflects the familiarity with the coding system that develops on the part of some respondents as they work with the form.

Yes-No Grid

There were relatively few errors found in Item 3 (Professional Society Affiliation). See Table 8. Less than one percent of the respondents left the item blank. The poorly worded instruction "Indicate only one response for each item below" rather than "for each column below" probably accounts for the majority of the 3-1/2% showing an incomplete response.

Number of Errors Per Individual Schedule

Up until now we have dealt with the errors identified in selected items. We might naturally ask: Are these errors concentrated in a relatively few schedules or are they distributed more generally among the respondents? In order to answer this question, an analysis was made of error types identified in the editing process. Thus, if one error was repeated in more than one item for a given schedule, it was considered as only one error. The data are shown in Table 9. There is relatively close agreement in the number of errors identified in this aspect of the scanning process with the number identified in the item-by-item analysis. This would seem to indicate that an error type was not frequently repeated on a given schedule. However, the errors were restricted to approximately 30% of the schedules. Ten percent of the Establishment Schedules had more than one error. Twenty percent had one error (the majority were permissible discrepancies as shown in Tables 6 and 7) and 70% had no errors. The corresponding percentages were very similar for the Personnel Schedules. Sixty-seven percent had no errors, 19% had one error and 14% had two or more errors.

CONCLUSIONS

1. The form, when professionally executed, presents an attractive package and demands the attention of the respondent. The final response rate to the pretest, after two follow-up letters and a night letter to the hard-core non-respondents, was 98%. In a survey such as this, where the motivation to reply is not high, this is an unusually high response rate. Eighty percent of the people found the form convenient. We would not attempt to estimate the acceptance rate had the form been more complex.

2. The length of time required to complete the form was not excessive. If the system becomes more popular, respondents will no longer need the detailed instructions. The time required for filling in the dots should be comparable with that required if the questions were asked on an orthodox questionnaire form.

3. Discrepancies between the code and the write-in were noted in 1% to 2% of the cases. Transposed digits and omissions accounted for the majority of the discrepancies. It is therefore recommended that if an error rate of 2% is considered excessive, the write-in box be utilized as an additional control and basis for verification wherever numerical data are requested.

4. The number of necessary query write-backs is not increased appreciably by this system. The critical errors appear independent of the questionnaire format.

5. The critical errors can be spotted relatively easily by a trained staff of editors. If the technique becomes more widely used, these error types would be expected to be reduced.

6. The primary gain realized from this system is, of course, the bypassing of conventional key punching and verifying. Where these facilities are not readily available, the system has obvious advantages both in terms of expense and timing. The punching is, in effect, performed by the respondent, who appears to handle both the alpha and numeric grids with ease.

TABLE 1 -- Responses to Question 2. Is the Questionnaire Convenient to Use?

Professional Classification	Total	Yes	No	Not Sure
	Number			
Total	308	250	40	18
Psychiatrists	64	52	7	5
Other M.D.'s	25	24	1	0
Psychologists	42	34	8	0
Social Workers	75	55	15	5
Nurses	102	85	9	8
	Percent			
Total	100.0	81.2	13.0	5.8
Psychiatrists	100.0	81.3	10.9	7.8
Other M.D.'s	100.0	96.0	4.0	0.0
Psychologists	100.0	80.9	19.1	0.0
Social Workers	100.0	73.3	20.0	6.7
Nurses	100.0	83.4	8.8	7.8

TABLE 2 -- Responses to Question 3. Are there any parts which are not clearly stated or which you do not understand?

Professional Classification	Total	Yes	No	Not Sure
	Number			
Total	308	52	243	13
Psychiatrists	64	8	51	5
Other M.D.'s	25	5	20	-
Psychologists	42	12	29	1
Social Workers	75	18	55	2
Nurses	102	9	88	5
	Percent			
Total	100.0	16.9	78.9	4.2
Psychiatrists	100.0	12.5	79.7	7.8
Other M.D.'s	100.0	20.0	80.0	--
Psychologists	100.0	28.6	69.0	2.4
Social Workers	100.0	24.0	73.3	2.7
Nurses	100.0	8.8	86.3	4.9

TABLE 3 -- Responses to Question 5. How long did it take you to fill out the questionnaire?

Professional Classification	Total	<10 min.	10-20 min.	20-30 min.	one-half hour +
	Number				
Total	308	125	135	44	4
Psychiatrists	64	23	31	10	0
Other M.D.'s	25	15	8	2	0
Psychologists	42	17	18	7	0
Social Workers	75	23	32	18	2
Nurses	102	47	46	7	2
	Percent				
Total	100.0	40.6	43.8	14.3	1.3
Psychiatrists	100.0	35.9	48.5	15.6	0.0
Other M.D.'s	100.0	60.0	32.0	8.0	0.0
Psychologists	100.0	40.5	42.8	16.7	0.0
Social Workers	100.0	30.7	42.6	24.0	2.7
Nurses	100.0	46.0	45.1	6.9	2.0

TABLE 4 -- Responses to Question 4. The allocation of hours employed by type of work could be obtained with:

Professional Classification	Total	Amount of Difficulty: Little or no	Some	Great
	Number			
Total	308	197	105	6
Psychiatrists	64	39	24	1
Other M.D.'s	25	23	2	0
Psychologists	42	24	18	0
Social Workers	75	38	34	3
Nurses	102	73	27	2
	Percent			
Total	100.0	64.0	34.1	1.9
Psychiatrists	100.0	60.9	37.5	1.6
Other M.D.'s	100.0	92.0*	8.0*	0.0
Psychologists	100.0	57.1	42.9	0.0
Social Workers	100.0	50.7	45.3	4.0
Nurses	100.0	71.5	26.5	2.0

* Significant from other discipline responses.

TABLE 5 -- Responses to Question 6. Would you define any additional terms?

Professional Classification	Total	Yes	No	No Response
	Number			
Total	308	21	232	55
Psychiatrists	64	4	50	10
Other M.D.'s	25	1	19	5
Psychologists	42	4	30	8
Social Workers	75	10	52	13
Nurses	102	2	81	19

Professional Classification	Total	Yes	No	No Response
	Percent			
Total	100.0	6.8	75.3	17.9
Psychiatrists	100.0	6.2	78.2	15.6
Other M.D.'s	100.0	4.0	76.0	20.0
Psychologists	100.0	9.5	71.5	19.0
Social Workers	100.0	13.3	69.4	17.3
Nurses	100.0	2.0	79.4	18.6

TABLE 6 -- Number and Percent of Errors in the Name Grids by Error and Type of Schedule

ERROR TYPE	ESTABLISHMENT SCHEDULES (n = 168)		PERSONNEL SCHEDULES (n = 4,079)	
	Number	Percent	Number	Percent
Critical Errors	2	1.2	3	0.1
Blank	1	0.6	3	0.1
Incomplete Code + Write-in	1	0.6	-	--
Non-Critical Errors	6	3.6	91	2.2
Write-in/No Code	2	1.2	15	0.4
Write-in/ Incomplete Code	1	0.6	26	0.6
Code ≠ Write-in	3	1.8	48	1.2
Permissible Dis- crepancies	21	12.5	266	6.5
Zero Space Not Marked	20	11.9	265	6.5
Code/No Write-in	-	--	1	<0.1
Code/Incomplete Write-in	1	0.6	-	--

TABLE 8 -- Analysis of Execution of Professional Affiliation (Item 3) on Personnel Schedules

Error	Frequency	Relative Frequency
Total	4,079	1.000
Blank	21	.005
Incomplete	138	.034
No error	3,920	.961

TABLE 9 -- Number and Percentage Distribution of Establishment and Personnel Schedules by the Number of Errors Per Sheet 1/

NUMBER OF ERRORS	ESTABLISHMENT SCHEDULE		PERSONNEL SCHEDULE	
	Number	Percentage	Number	Percentage
Total	168	100	4,079	100
0	116	70	2,740	67
1	34	20	761	19
2	12	7	298	7
3	5	3	191	5
4	1	<1	57	1
5	0	--	32	1
6 or more	0	--	0	-

1/ An error is defined as any reponse not in complete accord with the DocuTran directions. Also, if one error was repeated in more than one item it was still considered one error (type).

TABLE 7. Number and Percent of Errors in Selected Numerical Fields by Error and Type of Schedule

ERROR TYPE	ESTABLISHMENT SCHEDULES (n = 168)				PERSONNEL SCHEDULES (n = 4,079)					
	Item 3		Item 4		Item 6		Item 7		Item 8	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Critical Errors	1	0.6	1	0.6	62	1.5	116	2.8	280	6.9
Blank	-	--	1	0.6	26	0.6	28	0.7	8	0.2
Incomplete Code + Write-in	1	0.6	-	--	36	0.9	88	2.1	272	6.7
Non-Critical Errors	5	3.0	6	3.6	150	3.7	190	4.7	254	6.2
Write-in/No Code	3	1.8	2	1.2	24	0.6	32	0.8	32	0.8
Write-in/Incomplete Code	-	--	3	1.8	68	1.7	112	2.7	166	4.1
Code ≠ Write-in	2	1.2	1	0.6	58	1.4	46	1.1	56	1.4
Permissible Discrepancies	17	10.1	12	7.1	215	5.3	311	7.6	275	6.7
Code/No Write-in	5	3.0	3	1.8	125	3.1	160	3.9	150	3.7
Code/Incomplete Write-in	5	3.0	4	2.4	90	2.2	151	3.7	125	3.1
Zero Space Not Recorded	7	4.2	5	3.0	1/	1/	1/	1/	1/	1/

1/ Not observed

FIGURE I

Survey of Professional Personnel Employed in Mental Health Establishments

ESTABLISHMENT SCHEDULE
SIDE 1

U.S. Dept. of Health, Education, and Welfare
Public Health Service
National Institute of Mental Health
Manpower Studies Program

PHS T-134
BUREAU OF BUDGET NO. 68-62246
Approval expiration date is January 31, 1963

The purpose of this form is to identify and provide overall basic data on the establishment for which your personnel schedules are submitted. One completed Establishment Schedule must accompany the personnel schedules when they are returned for processing.

Please read all of the following directions and review the example on Side 2 before beginning to complete this schedule.

Item 1—Write in these blanks the complete name and address of your **establishment**:

Name..

Street Address..

City..State........................

Use only a **soft, black-lead pencil (No. 2 or 2½).** Do not use ink, ball-point pen, or colored pencil.

To fill in this schedule, first print appropriate numbers or letters in the boxes provided. Then blacken the corresponding numbered or lettered circles in the columns above the boxes.

In marking a circle, fill in the entire circle but do not go outside it. Mark **one and only one** circle in each column. If you make an error, erase it thoroughly, and then fill in the correct circle.

Item 2—ESTABLISHMENT NAME. In the row of boxes under the establishment name grid below, print the name of your establishment. Use only eighteen letters and abbreviate when necessary. Leave a blank box between names or abbreviations. Notice the following examples:

CK CO INST J RSCH for Cook County Institute for Juvenile Research
STAR OF SEA HOSP for Star of the Sea Hospital
MONT CO CHD GD CLN for Montgomery County Child Guidance Clinic

Then, in the alphabetic column above each box, blacken the circle containing the letter that you printed in the box. Blacken only one circle in each column. **Blacken the blank circle at the top of each unused box,** including those between names and at the end of the name if you do not need all eighteen letters.

Item 3—TOTAL NUMBER OF EMPLOYEES ON PAYROLL—FULL TIME AND PART TIME. Under full time, include all employees who work a minimum of 35 hours per week. Under part time, include those employees working less than this number of hours per week. Include all personnel, regardless of professional or non-professional status.

PLEASE NOTE: If this establishment operates branches, personnel for both headquarters and branches should be reported on one establishment schedule.

However, please note that for purposes of this survey *out-patient mental health clinics* and their branches are to be reported separately from related hospitals. Thus, a person on both the hospital staff and the out-patient clinic staff should be reported on separate forms—one for the clinic and one for the hospital—and his Personnel Schedule should accompany the Establishment Schedule for each establishment.

First, print the number of personnel in the boxes under the appropriate heading. Then, in the numeric column above each box, blacken the circle containing the number printed in the box. Be sure to print zeros for boxes not used. For example, 9 full-time employees is printed 0009; and two part-time employees is printed 002.

(Continued on Side 2)

2 ESTABLISHMENT NAME

Print establishment name in these boxes.

3 TOTAL NUMBER OF EMPLOYEES ON PAYROLL

Full Time | Part Time

4 NUMBER OF MENTAL HEALTH PERSONNEL

A. Psychiatrists and other Physicians | B. Psychologists | C. Social Workers | D. Professional Nurses

USE A SOFT LEAD PENCIL—MARK ONLY IN SPACES PROVIDED.

Please handle this sheet carefully. Keep it as clean as possible. Do not fold it or bend the corners.

5 NUMBER OF PERSONNEL SCHEDULES FORWARDED

Note: The sum of columns 4A through 4D should equal the figure in Item 5.

SRA DOCUTRAN™ SERVICE
165-13-2

FIGURE I (Continued)

Survey of Professional Personnel
Employed in Mental Health Establishments

ESTABLISHMENT SCHEDULE
SIDE 2

U.S. Dept. of Health, Education, and Welfare
Public Health Service
National Institute of Mental Health
Manpower Studies Program

Item 4—NUMBER OF MENTAL HEALTH PERSONNEL. The four mental health core disciplines shown are:

A. Psychiatrists and Other Physicians—residents in psychiatry, psychoanalysts, and all other M.D.'s.

B. Psychologists—clinical and counseling psychologists, psychometrists, and related personnel designated as psychologists by your establishment.

C. Social Workers—all persons so designated by your establishment.

D. Professional Nurses—those persons employed as nurses who have completed a prescribed course of nursing in a school approved by a state licensing agency and who hold, or are eligible for, a current license to practice nursing in the state where employed.

Include personnel as defined above whether they work full time or part time. Include all professional personnel whether reimbursed or not. Also, include personnel for headquarters and branches. **Exclude** professional personnel on leave of absence. **The sum of columns 4A through 4D should equal the figure you enter in Item 5.**

First, print the number of professional personnel in the boxes under the appropriate heading. Then, in the numeric column above each box, blacken the circle containing the number printed in the box. Be sure to print zeros for boxes not used. For example, a 9 in category 4A should be printed 09; notice that a 9 in category 4D should be printed 009. If a category does not apply to your establishment, put 00 in the boxes and blacken the **top** (00) circles.

Item 5—NUMBER OF PERSONNEL SCHEDULES FORWARDED. Count the personnel schedules received from members of your professional staff and enter the figure in Item 5.

Should some personnel schedules be outstanding because staff members are on vacation or otherwise temporarily unavailable, please complete schedules for those personnel to the best of your ability. Do not delay return of all survey schedules until such persons are available, and **do not** retain the schedules to be sent in later. **The number of personnel schedules (and the figure you enter in Item 5) should equal the sum of the full-time and part-time professional personnel indicated in Items 4A through 4D.** Please resolve and correct any discrepancies.

To fill in Item 5, print the number in the boxes at the bottom of the grid. Then, in the numeric column above each box, blacken the circle containing the number printed in the box. Be sure to print zeros for boxes not used. For example, in this section, 12 should be 012.

DIRECTIONS TO RETURN SCHEDULES

1. Place this completed Establishment Schedule on top of the group of personnel schedules from your establishment. Be sure that the personnel schedules are stacked with Side 1 (name grid) up and that the completed Establishment Schedule is on top of the stack with Side 1 (name grid) up.
2. Band the stack with a *loose* rubber band. Please do not use paper clips, staples, or pins.
3. Place the stack in the special envelope provided and mail to:

   Mental Health Manpower Studies Program
   Training Branch
   National Institute of Mental Health
   Building 31, Room 2A-07
   Bethesda 14, Maryland

   a. If you return less than 10 schedules, please place the cardboard stiffener provided in the envelope with the schedules for protection.

   b. If you have more than 100 schedules, it will be necessary to split the stack between envelopes. Place the Establishment Schedule on top of the first stack; then place a plain sheet of paper on each stack and mark each plain sheet with "Split Pack" and the name of your establishment. For example, if you have two stacks, mark the plain sheets "Split Pack 1 of 2" and Split Pack 2 of 2" with your establishment name.
4. Also, please return any *unused* forms in the same envelope at the bottom of the pack.

SAMPLE FILLED-IN ESTABLISHMENT SCHEDULE

MAKE NO MARKS ON THIS SIDE

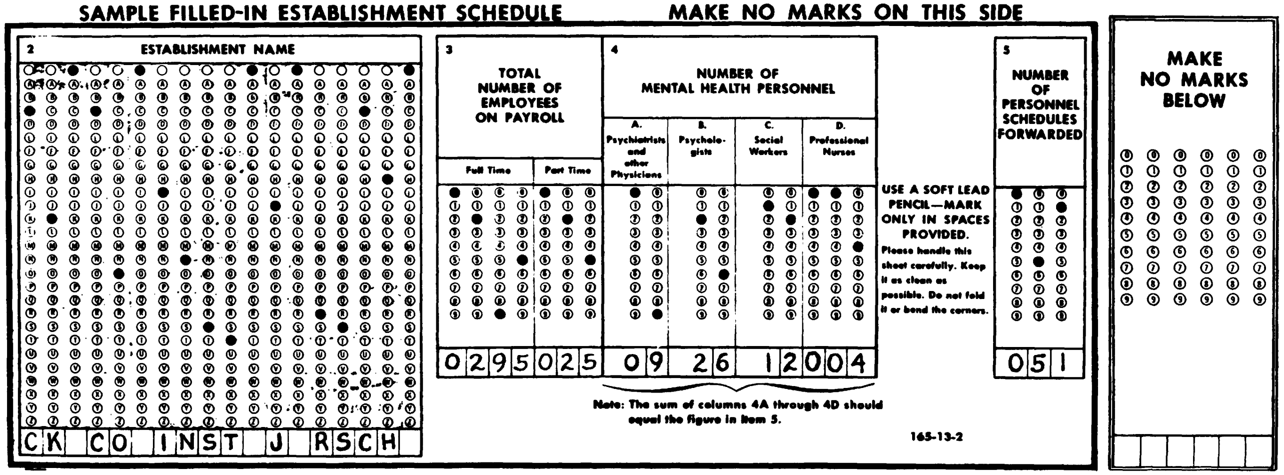


FIGURE II

Survey of Professional Personnel Employed in Mental Health Establishments

PERSONNEL SCHEDULE
SIDE 1

U.S. Dept. of Health, Education, and Welfare
Public Health Service
National Institute of Mental Health
Manpower Studies Program

PHS T-135
BUREAU OF BUDGET NO. 68-6246
Approval expiration date is January 31, 1963

SAMPLE NAME GRID

S M I T H J O H N A

This schedule is specially designed to be "read" by a high-speed, electronic optical scanner in order to facilitate earlier, and more meaningful use of the information gathered. Since the schedule will be processed by machine, it is of the utmost importance that it be filled in completely and accurately.

Please handle this sheet carefully; do not fold it or bend the corners. Make no marks on it except in the spaces provided and keep it as clean as possible.

Read all of the directions and samples before you begin to fill in this schedule.

Use only a soft, black-lead pencil (No. 2 or 2½). Please do not use ink, ball-point pen, or colored pencil because they cannot be read by the machine.

To properly complete the schedule, blacken each appropriate circle with a heavy black mark, filling in the entire circle, but not going outside it. Mark one and only one circle in each column. If you make an error, erase it thoroughly, and then fill in the correct circle.

DIRECTIONS FOR COMPLETING SIDE 1

ITEM 1—NAME. Study the sample name found to the left and then fill in Item 1 with your name. Notice that a box is provided below each column of circles. Starting with the first box on the left, print the letters of your last name, skip a space, print your first name, skip a space, and then print your middle initial. If space is insufficient for first name and middle initial, abbreviate first name or print the initials of first and middle name. Then, in the alphabetic column above each box, blacken the circle containing the letter that you printed in the box. In the column above each unused box that separates your last and first name and middle initial, blacken the top (blank) circle. Blacken only one circle in each column, but be sure that there is one circle blackened in every column of the name grid. (If you have completed this grid as directed and find that there are still unused columns following the last letter, your middle initial, blacken the top row of circles in the remaining columns.)

PLEASE NOTE: This name grid is for identification purposes in processing. No data on individuals will be tabulated.

Immediately to the right of the name grid boxes, you will note that space has been left for you to indicate the name of the establishment for which this schedule is being completed. Please print its name; then, on the line below, print the names of the city and state where your establishment is located.

ITEM 2—PRINCIPAL PROFESSIONAL CLASSIFICATION. Six professional classifications are shown:

Resident in Psychiatry—currently in residence in this establishment.
Psychiatrist—includes psychoanalysts.
Other Physician—all M.D.'s not included in the two categories above.
Psychologist—includes clinical and counseling psychologists, psychometrists, and related personnel designated as psychologists by this establishment.
Social Worker—includes all persons so designated by this establishment.
Professional Nurse—those persons employed as nurses who have completed a prescribed course of nursing in a school approved by a state licensing agency and who hold, or are eligible for, a current license to practice nursing—in the state where employed.

BE SURE TO COMPLETE BOTH SIDES OF THIS FORM.

165-13-1

1 **YOUR NAME**

(Last Name, First Name, Middle Initial)

Print your name in these boxes; start with last name first.

INDICATE ONLY ONE RESPONSE FOR EACH ITEM BELOW

2 PRINCIPAL PROFESSIONAL CLASSIFICATION	3 PROFESSIONAL AFFILIATION A. American Psychiatric Association	B. American Board of Psychiatry and Neurology Certified	C. American Psychological Association	D. American Board of Examiners in Prof. Psychology	E. American Nursing Association	F. National Association of Social Workers	4 MOST ADVANCED LEVEL OF EDUCATION	5 Employed by More Than One Mental Health Establishment?
○ Resident in Psychiatry ○ Psychiatrist ○ Other Physician ○ Psychologist ○ Social Worker ○ Professional Nurse	○ YES ○ NO	○ YES ○ NO	○ YES ○ NO	○ YES ○ NO	○ YES ○ NO	○ YES ○ NO	○ No Degree ○ Assoc/Arts ○ 3-Yr Diploma ○ Bachelors ○ Grad Study No Degree ○ Masters ○ PhD, EdD, ScD, etc. ○ M.D. ○ M.D. Plus PhD, EdD, ScD, etc.	○ YES ○ NO

If yes, fill in blank(s) with name and city.

USE A SOFT LEAD PENCIL ONLY — Please handle this sheet carefully. Keep it as clean as possible. Do not fold it or bend the corners.

NAME OF THIS ESTABLISHMENT

CITY STATE

	NAME	CITY
1		
2		
3		

OTHER ESTABLISHMENTS IN WHICH CURRENTLY EMPLOYED

FIGURE II (Continued)

Survey of Professional Personnel Employed in Mental Health Establishments

PERSONNEL SCHEDULE SIDE 2

U.S. Dept. of Health, Education, and Welfare
Public Health Service
National Institute of Mental Health
Manpower Studies Program

It is possible that more than one classification might apply to you. Nevertheless, blacken the **one** circle that identifies your principal classification **in this establishment.** Do not blacken more than one circle. If you are not in one of the classifications given, please return this schedule to your administrative office.

ITEM 3—PROFESSIONAL AFFILIATION. Read carefully each of the affiliations given (A through F); then blacken the **one** circle in each that denotes your present affiliation status. Be sure to respond to **every** category. (Do not print "Yes" or "No;" just fill in the appropriate circle.)

ITEM 4—MOST ADVANCED LEVEL OF EDUCATION. Carefully review all levels of educational training shown before making any marks. Then blacken the circle beside the level that best identifies your most advanced degree(s) or extent of training. Blacken only one circle in this section.

ITEM 5—EMPLOYED BY MORE THAN ONE MENTAL HEALTH ESTABLISHMENT? Indicate by blackening the appropriate circle ("Yes" or "No") whether you are currently providing services for more than one mental health establishment (hospital, clinic, institution for mentally retarded, etc.). If so, print the name(s) and city(ies) in the spaces provided. If you subsequently receive another schedule from another establishment, please complete a schedule at each establishment.

DIRECTIONS FOR COMPLETING SIDE 2

ITEM 6—PERSONAL DATA. (A) *Date of Birth*. First, print the month, day, and last two digits of your birth year in the boxes provided. Then blacken the circle that denotes the **month** of your birth. Next, blacken the circles to indicate the **day** and the **last two digits of the year** of your birth.

CAUTION: In the event that both columns of day or year are not needed to record the necessary information, as in the case of March 3, the 3 should be entered in the **right-hand** column and the 0 circle blackened in the left-hand column. It will then appear as 03. Study the sample numeric grid found below. Notice that the 3 has been filled in as 03. (B) and (C) *Sex and Citizenship*. Blacken the appropriate circle for each item. It is not necessary to print the answers below the circles.

ITEM 7—YEARS EXPERIENCE. This item is divided into two parts to identify the number of completed years that you have practiced in your present psychiatric or mental health specialty, as related to the number of years for which you have related professional employment not as a psychiatrist or mental health specialist. For example, if you are a psychiatrist, indicate in columns 7A the number of years that you have practiced psychiatry; in columns 7B, indicate the number of years that you practiced in other medical areas. If you are a psychologist, social worker, or a nurse, indicate in columns 7A the number of years you have been employed in a mental health setting; in columns 7B indicate the number of years that you were employed in a non-mental health setting. **Remember**, for any years less than 10, blacken the number circle in the **right-hand** column, then blacken the 0 circle in the left. For example, the number 2 is filled in as 02. If you have completed less than one full year in a given category, put 00 in the boxes and blacken the **top** (00) circles.

ITEM 8—CURRENT ASSIGNMENT—(AVERAGE) NUMBER OF HOURS EMPLOYED IN THIS ESTABLISHMENT BY TYPE OF WORK DURING A TYPICAL WEEK. The purpose of this item is to determine by types of work, the average number of hours per week that you work in this establishment. If your assignment varies from week to week, make your best estimate of a typical week by dividing the total number of hours worked per month by $4\frac{1}{3}$. Accordingly, for each of the categories 8A through 8F, put the number of hours in the boxes below the grid; then, in each column blacken the circle above containing the number appearing in each box. **Remember**, for any numbers less than 10, blacken the number circle in the **right-hand** column, then blacken the 0 circle in the left. For example, 02, 04, 09, and so forth. In those categories that do not apply, put 00 in the boxes and blacken the **top** (00) circles. **Be sure that the sum of columns 8B through 8F equals the figure in column 8A.**

ITEM 9—MONTHS APPROVED RESIDENCY COMPLETED—FOR PSYCHIATRISTS ONLY. If you are a psychiatrist or psychiatric resident, indicate the months of residency completed in an institution approved by the Council on Medical Education and Hospitals of the AMA at the time of residency. Enter that number in the boxes below the grid; then blacken the circles above that contain the number appearing in each box. Where less than 10 months are involved, be sure to blacken the numbered circle in the **right-hand** column and the 0 circle in the left. For example, two months should be filled in as 02.

6 PERSONAL DATA					7 YEARS EXPERIENCE		8 CURRENT ASSIGNMENT						9 PSYCHIATRIST ONLY
A. DATE of BIRTH			B. SEX	C. TYPE OF CITIZENSHIP	A. In Present Psychiatric or Mental Health Specialty	B. In Present Non-Psychiatric or Non-Mental Health Specialty	NUMBER OF HOURS EMPLOYED THIS ESTABLISHMENT BY TYPE OF WORK DURING A TYPICAL WEEK						Months Approved Residency Completed
MONTH	DAY	YEAR					A Total	B Patient Care and/or Clinical	C. Administration	D Teaching	E. Research	F. Consultation	
JAN			Male	U S Native									
FEB													
MAR			Female	Naturalized									
APR													
MAY				Alien									
JUN													
JUL													
AUG													
SEPT													
OCT													
NOV													
DEC													

Note: The sum of columns B through F should equal the figure in column A.

SAMPLE NUMERIC GRID

03

USE A SOFT LEAD PENCIL—MARK ONLY IN SPACES PROVIDED.

Please handle this sheet carefully. Keep it as clean as possible. Do not fold it or bend the corners.

SRA

DOCUTRAN™ *SERVICE*

165-13-1

MAKE NO MARKS BELOW

FIGUPE III

PHS-T139
10-62

FORM APPROVED
BUDGET BUREAU NO. 68-6246

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE
NATIONAL INSTITUTES OF HEALTH

EVALUATION OF THE PRETEST ON
SURVEY OF PROFESSIONAL PERSONNEL EMPLOYED
IN MENTAL HEALTH ESTABLISHMENTS

1. INDICATE YOUR PROFESSIONAL CLASSIFICATION BELOW

Check	Classification
	Psychiatrist, Resident Psychiatrist
	Other M. D.
	Psychologist
	Social Worker
	Professional Nurse

2. IS THE QUESTIONNAIRE CONVENIENT TO FILL OUT?

☐ Yes ☐ No ☐ Not sure

3. ARE THERE ANY PARTS WHICH ARE NOT STATED CLEARLY, AND WHICH YOU DO NOT UNDERSTAND?

☐ Yes ☐ No ☐ Not sure

IF YOU HAVE CHECKED YES OR NOT SURE, LIST ITEM NUMBERS AND EXPLAIN:

4. THE ALLOCATION OF HOURS EMPLOYED BY TYPE OF WORK COULD BE OBTAINED WITH

Check	Obtained
	Little or no difficulty
	Some difficulty
	Great difficulty

IF YOU HAD GREAT DIFFICULTY ALLOCATING YOUR TIME, WHICH WORK CLASSIFICATIONS GAVE YOU TROUBLE?

5. HOW LONG DID IT TAKE YOU TO FILL OUT THE QUESTIONNAIRE?

Check	Length of Time
	Less than 10 minutes
	10 - 20 minutes
	20 - 30 minutes
	More than 1/2 hour

6. WOULD YOU DEFINE ANY ADDITIONAL TERMS?

☐ Yes ☐ No If yes, which terms?

7. ADD ANY OTHER COMMENTS YOU FEEL NECESSARY:

THE PERIODIC HEALTH EXAMINATION RESEARCH PROGRAM: RESULTS AND PROSPECTS

Stanley S. Schor, University of Pennsylvania

The periodic health examination is a medical survey of presumably healthy individuals carried out at predetermined intervals in an effort to diagnose disease in its incipiency, and thus to minimize its seriousness. This concept, long advocated by the medical profession, has been widely accepted by industry. The practice of so-called "preventive maintenance" of machines has proved so successful that its application to humans has seemed a logical extension. The value of these examinations has been judged largely by the kind and amount of disease diagnosed in groups of people in organized programs and by individual experience with a particular disease. For example, the managements of some companies feel that if only one early cancer lesion is discovered in years of examining all executives this, in itself, makes the PHE worthwhile. In view of the increasing amount of time and money being invested in this procedure, however, there is need for a better assessment of its value.

Does it have any effect, for example, on the masses of people in the programs? Does it affect morbidity? Mortality? Is the current type of examination the best? What tests or procedures should be added? Which should be dropped? What is the optimum package for different populations? There are any number of questions which should have been answered by now but have not been.

The United States Public Health Service has been cognizant of this problem for a long time and seven years ago invited representatives of several clinics performing these examinations to Washington to discuss the subject. It became apparent that many of the unanswered questions could only be answered if larger populations were available. To this end, this group of interested people set up a committee to organize cooperative research among the clinics. A director was appointed in April 1961 and several studies were designed for which funds were obtained in August 1961 from the United States Public Health Service.

The first study attempted was an analysis of the people who died while in a PHE program. The physician taking part in a PHE program has more than a casual interest in the obituary page of his local newspaper. There he sometimes finds an old friend whom he has recently examined and possibly given a clean bill of health. When this happens, he starts wondering if there wasn't really something wrong that he should have been able to detect but could not.

It was thought that a review of as many such cases as could be gathered might throw some light on this subject. At the very least, data would be obtained on the proportion of deaths occurring in people having had no diagnoses made associated with the cause of death at their last PHE. In other words, the percentage of people who had died after a PHE giving them a clean bill of health in terms of ultimate cause of death. Data would also be obtained on how this proportion varied when classified by other characteristics such as cause of death, age, interval between last examination and death, etc.

A study such as this might also disclose the fact that certain characteristics are different in a group of people who died while in a PHE program from those in a control group still alive, if in fact they are different. For example, the proportion of heavy cigarette smokers in the dead group might be higher than in the alive control group, or the proportion of overweight people might be higher, etc. This would give one some idea as to which characteristics are important to consider if one wishes to assess risk of death at the periodic health examination of any individual.

A study of this type could also help in evaluating certain tests and procedures in terms of their power to discriminate between people who died and those who did not die in the same interval of time.

Finally, in the process of examining the past PHE records of people who have died, other interesting studies might be suggested.

MATERIALS AND METHODS

Ten Clinics* performing periodic health examinations on people in the same broad socio-economic group - executives of companies and the faculty of a university - supplied data on 350 usable deaths and 350 counterparts still alive and matched with the dead people on clinic, age, date of last examination and time-interval

(*) The ten Clinics supplying the information:
Benjamin Franklin Clinic
Eastman Kodak (Rochester Plant)
General Electric (Schenectady)
The Greenbrier Clinic
Associates of Mass. Memorial Hospitals
Univ. of Pennsylvania Diagnostic Clinic
Univ. of Pittsburgh
Univ. of Michigan, Faculty Program
Standard Oil Company (N.J.)
United States Steel Co.

This research supported by Contr.SAph 78790 USPHS.

in a PHE program.

The data supplied consisted of the findings on every PHE performed on these 700 people, certain items in their family histories, their medical histories, and the diagnoses made by the examining physician at each of these examinations.

These 350 deaths were only the ones classified usable for the purpose of this Study. They were all males, all in a PHE program (did not just present themselves when deemed necessary), all in whom the cause of death could be determined by death certificate and all for whom matched counterparts could be obtained.

Only the data on the last examination on the dead person and the examination performed at the same time on the alive counterpart, were used in this Study. Over time changes, i.e. changes from first examination to last, in all tests and procedures are being analyzed but as yet no patterns are discernible.

The causes of death used were those listed on the death certificates and were combined into 7 categories as follows:

1. Coronary heart disease (including coronary occlusion, coronary thrombosis, myocardial infarction, coronary heart attack, acute coronary insufficiency. International Codes 420x).
2. Other heart disease (including decompensation, myocardial insufficiency, cardiac failure, rheumatic heart disease. 41xx, 421x to 446x).
3. Vascular disease (including cerebral vascular accident, cerebral hemorrhage, emboli, aortic aneurysm, subarachnoid hemorrhage, ruptured aneurysm, etc. 330x to 334x, 45xx, 463x, 464x, 465x, 466x).
4. Cancer, leukemia, and other malignant disease (140x to 207x, 23xx).
5. Post-operative deaths.
6. Accidents and suicides.
7. Other deaths.

To test the validity of causes of death as drawn from death certificates in this Study, a sample of 18 records of autopsies performed on subjects dying in the Philadelphia area was examined. These showed agreement with the death certificate as to cause of death in 17 instances.

In determining whether the cause of death had been detected prior to death, the diagnoses made by the examining physician were compared with those on the death certificate. Agreement was considered to exist if the clinical record included either the specific diagnosis of cause of death, or a diagnosis of a disease closely related to the cause of death.

The first question we wished to answer was:

<u>Question 1</u> - How often was the cause of death detected on last examination?

The disease which caused death was known on the last examination in just about one-half of those who died. This proportion varied with age, cause of death, interval between last examination and death and other characteristics, one of them being cigarette smoking. It was interesting to note that heavy cigarette smokers who die from coronary artery disease tend not to have the disease diagnosed before death, whereas light smokers and non-smokers who die of it, do. One interpretation is that heavy cigarette smokers tend to die from coronary disease <u>before</u> the onset of clinical manifestations of the disease, whereas non-smokers do not. Putting it another way, one might conclude that non-smokers have a better chance of surviving the first coronary than heavy smokers.

If one wishes to use these figures, i.e. the proportion of people in whom the cause of death was known at the last examination, as an assessment of the accuracy of the PHE a severe limitation arises. If a diagnosis of heart disease is made on everyone being examined then, obviously, in 100 percent of the coronary heart disease deaths would the cause of death be known on last examination. Ideally, in order to draw a valid conclusion about the accuracy of the examination the proportion of heart disease diagnoses made incorrectly should be known. This was impossible to ascertain. It was possible, however, to obtain a sample of people similar to those who died who were still alive at the time of the former, and calculate the proportion in whom heart disease was diagnosed.

For this reason, for each person who died while in a PHE program another person who was alive at the time of death of the former, was selected. This control was matched on several characteristics with his dead counterpart. He was the same age, the same sex, the same race, visited the same Clinic, was examined at approximately the same time as the last examination of the dead person, and had been examined over the same interval of time. These alive matched counterparts can be thought of as a representative sample of all people who are alive but having the same age distribution, examination date, etc. as the dead group.

As noted, in one-half of the people who died, the cause of death was known at the last examination. In 1/5 of their alive counterparts, the same disease had been diagnosed at the last examination. Thus, it was only 2-1/2 times as likely for a person who died to have had the cause of death diagnosed as it was to have had the same disease diagnosed in those still alive.

A more detailed discussion of this Study and a more comprehensive set of results appear elsewhere.

The data gathered to answer Question 1 can also be used to answer another question, namely: Are there any characteristics which distinguish the people who died from those who survived the same period of time?

If one compares two groups, one containing all who died with the other containing all who survived, one finds that those dying of coronary heart disease had certain abnormalities with significantly greater frequency than did their surviving counterparts. In addition, these characteristics were not the same when one compares deaths with survivals of people in whom coronary heart disease was diagnosed as they were when coronary heart disease was undetected in both groups.

If a test is given to two individuals, one of whom dies and the other survives, the test results will fall into 1 of 4 categories:

1. The test may indicate abnormality, i.e. be positive in the member who dies and negative or normal in the member who lives. In this case the test has correctly discriminated between them in terms of survival ($\frac{+}{-}$).
2. The test results may be exactly the reverse, i.e. positive in the member who survives and negative in the member who dies. This may be termed "false discrimination". A positive test has not only failed to develop in the member who died, but has developed in the survivor ($\frac{-}{+}$).
3. The test may be negative or normal in both ($\frac{-}{-}$).
4. It may be positive in both ($\frac{+}{+}$).

None of the characteristics tested appeared to discriminate very well between people who died and people who did not die in the same interval.

We hope in the future to obtain discriminant functions for these data but they are not as yet available.

Another question which arose after a cursory glance at the collected data is now being attacked. For people who died of coronary disease within a year after the EKG was read as normal was there anything unusual about these EKG tracings, as compared to people with so-called normal tracings and still alive? This is being studied at the University of Michigan and I have as yet no results to report.

At present we are starting two new studies in evaluating the PHE. One is aimed at determining if the PHE has any effect on mortality rates. The other has as its goal the selection of risk factors for certain diseases.

The study of mortality will be attempted in two major sections. The first will be the establishment of life tables for all people in the PHE programs of the cooperating Clinics for specific causes of death. Life tables will also be constructed for executives and for other population subgroups if warranted by size of sample. These tables will be compared with known tables for similar groups. The second will consist of locating life tables for as many different groups as are now or will shortly be available and comparing them with life tables constructed for various comparable segments of our PHE group.

A table of age specific mortality rates will be computed for the executives of certain railroads who have been in a PHE program since 1950. A comparison of these mortality rates will be made with rates available from other sources for all railroads.

A table of age specific mortality rates will be computed for faculty members who have been in a PHE program. A comparison of these mortality rates will be made with rates available from other sources for all teachers.

A table of age specific mortality rates will be computed for various socio-economic groups as similar as possible to those which are soon to become available from another study.

If possible, mortality rates will be constructed from insurance company records for insured people as similar as possible to people in certain segments of the PHE group.

The study of risk factors is a retrospective-prospective type of study - prospective as of 10 years ago. Incidence rates of certain diseases among groups of people with and without selected characteristics will be compared. The diseases to be studied are: coronary artery disease, hypertension, hypertensive heart disease, and diabetes.

A whole host of risk factors will be tested. As a by-product of this study some light will also be thrown on the question of the relationship between rectal polyps and cancer of the colon.

Some important findings are being uncovered by this cooperative program. But more important than any of these to us is the knowledge that these Clinics can work together and that they realize how important standardization of techniques, records, etc. is. We hope that if nothing else comes out of this effort but standardization, it will have been worthwhile.

We, meaning the research team of Drs. Katharine and Kendall Elsom and Thomas Clark, of the University of Pennsylvania, Dr. James Dunn, formerly of the University of Pittsburgh and now with Western Electric, myself, and the Steering Committee headed by Dr. Norbert Roberts of the Standard Oil Company (N.J.) are indebted to the Long-Term Illness Branch of the Division of Chronic Diseases of the U.S.P.H.S. and especially to Mr. Robert Thorner, for their support, and also to the people at the cooperating Clinics who gave much time and knowledge to this program.

CONGRESSIONAL RE-APPORTIONMENT IN MICHIGAN: A STUDY IN PUBLIC POLICY

Richard A. LaBarge

Introduction

This paper advances a new approach to the problem of representation in modern representative government. It argues that traditional thinking on sound representation requires thorough revision if new challenges to political democracy are to be met. It presents a standard for the resolution of these problems and discusses the implementation of that standard in congressional re-apportionment in Michigan.

In order to understand the requirements of a sound congressional apportionment, we must first direct our attention to the operating principles at work in the political behavior of modern American society. Unless we know something about our constituencies and the bases for their behavior, we cannot pretend to design representative institutions which suit their needs and protect their rights.

Operating Principles

PRINCIPLE I: <u>The organization potential of any given number of voters increases with the density of population</u>.

Time spent by a political worker in interviews with his constituents varies considerably from domicile to domicile. Some people will talk politics for more than an hour; others have no interest at all. Our experience in the field indicates that a mean canvass time of approximately five minutes per domicile is necessary for a good initial contact with one's constituents.

Now let us apply this cost to three typical election precincts, each with 1,000 registered voters in 500 separate domiciles. If the first precinct consists of high-rise apartment dwellings in a densely populated urban community, the transportation time between domiciles is negligible. One can say with some confidence that adequate contact with these voters can be established in approximately 500 times 5 minutes -- roughly 42 man-hours.

If the second precinct consists of less densely populated single-family dwellings, at least two additional minutes will be required for the canvasser to walk between domiciles. In such a precinct -- typical of the suburban communities -- the manpower cost of a canvass rises to approximately 500 times 7 minutes, roughly 58 man-hours.

If the third precinct consists of widely separated farms which require an average motorized transportation time of five additional minutes between domiciles, the manpower cost of a canvass becomes double that of the initial urban precinct: 500 times 10 minutes or approximately 83 man-hours.

When one considers the additional costs of using motorized transport in the rural precincts or the problems involved when inclement weather makes travel between dwellings unpleasant, it is small wonder that the most densely populated areas are the most susceptible to political organization within any given span of time. It is no accident that "machine politics" in the United States has been associated closely with the big cities.

Of course, this proposition is not new. There is a long, honored, and familiar list of dissertations on the need for "checks and balances" to limit political authority in such a way that political organizations in densely populated areas could not establish firm and lasting holds on major government powers. What IS new is a gradual awareness that some of the old forms of "checks and balances" are no longer feasible in a United States characterized by rapid urbanization, growing population density, and an economic shift from rural to urban occupations.

PRINCIPLE II: <u>Within densely populated areas political abuses need to be limited through the use of countervailing power</u>.

In the past, efforts to limit political power have led to representation systems which skewed representation ratios away from the densely populated areas with a relatively low cost for political organization. Sparcely populated areas with a relatively high cost for political organization thus were used as leveling forces which made it easier for popular indignation to "throw the rascals out" whenever abuses of government power became evident. The modern problem lies in the fact that sparcely populated areas have well-nigh disappeared in some sections of the country, and where such areas remain their ratio of population to the total involved has dropped so low that it becomes increasingly more difficult to justify the over-representation accorded to them.

The problem remains, but the remedy passes on. Clearly what is needed are political institutions which make possible internal checks WITHIN the ever more prevalent and ever more extensive areas of high population density. Since the relatively low costs of political organization in these areas bias them toward a highly organized political situation, proper policy for such areas must foster effective competing organizations within them. If such is not the case, the individual independent voter becomes entirely submerged by the one dominant political machine, thus losing his freedom of choice and his ability to "throw the rascals out."

PRINCIPLE III: <u>In normal times, without unusual</u>

circumstances, the maximum expected deviation of partisan vote does not exceed ten per cent from election to election.

Consider the table appended to the end of this paper under the title, "Percentage Penetration of Party X in the Old 16th Congressional District of Michigan." This table presents the actual outer limits of partisan voting established in our constituency during the 1960 Presidential election year, the off-year election of 1962, and the odd-year spring election of 1963. By normalizing on a percentage basis and comparing the performance of the weakest candidates in each of the three elections, we can chart the trend of core Party X vote in each component of the constituency. Similar procedure comparing the performance of the strongest candidates gives us the upper limit of core Party X vote plus independent vote.

Observe that the spread between the strongest and weakest candidates -- that is, the independent vote -- is NOT large in any area. The largest percentage of net ticket-splitters recorded in any area was 12.2%, and net splitting in excess of 10% of the total vote occurred in only 7 of the 78 spread observations for the 27 components of the constituency. The independent voter exercises an influence only at the narrow margin of an election.

Now observe the changes in voting performance from election to election. It makes no difference whether one compares the weakest candidates or the strongest candidates, the conclusion is the same and is evident by inspection alone -- changes in community voting habits are VERY SMALL. The largest single change recorded for any one of the 104 change observations was 8.2%. Thus, at some limit where the performance of the strongest candidate is not too far away from 41% of the total vote, no one could win an election on the minority ticket.

PRINCIPLE IV: A dominant political organization tends to become increasingly more dominant.

Notice that the individual communities in our table have been arranged roughly in descending rank order of Party X voting penetration. In the intermediate range, which is bordered by rough limits of core vote somewhere in the neighborhood of 30% to 35% for each party, there is mixed change behavior. Eleven of these communities show the largest changes in favor of Party X, while seven others show their largest changes in favor of Party Y. The remaining community demonstrates equal maximum changes in both directions.

Outside of this intermediate range there is a terrifying uniformity of change behavior. At the top of the table, in the one or perhaps two communities where Party X predominates, Party X shows substantial and continuous growth in penetration. The converse is true at the bottom of the table, where Party Y dominates. There Party Y shows a substantial and continuous growth in the bottom six communities in this category.

Why should this be so? One might ask: "Isn't this simply an expression of the confidence of the people?" In dominated areas the answer is "No." The following direct canvass statements, taken by this writer from some of the householders in dominated areas of the old 16th Congressional District, illustrate the reasons:

"Well, I usually vote X, but there's no point in working for X. They're going to lose anyway. I have time to spend, but none to waste."

"I used to be an X, but I'm a Y now because that's the only way I can get anything."

"My husband's a fireman, so I have to be a Y."

Intimidation of the single individual is a powerful and conclusive weapon when exercised by the dominant political group. It is the main reason why so few citizens' reform movements are ever successful.

The New Approach

To sum up, we know that densely populated areas have a low cost of political organization, that the independent vote rarely exceeds 10% of the total, that the change of vote from election to election is not as great as 10%, and that progressive disequilibrium toward a one-party system sets in once the core vote of either one of the contenders slips any sizeable degree below 35%. Hence, a 35% to 40% minimum core vote is requisite to the maintenance of two party democracy in any densely populated area. If this condition is not met, the evidence is very strong that complete political dominance by one group can be the only end result for the constituency as a whole. Once competition in such areas is destroyed there are no internal checks to resurrect it. One party government becomes one faction government, and even the primaries of the dominant organization slip under the firm control of machine forces which endorse their preferred candidates. The ultimate step is dictatorship under the ruling bosses of the dominant group.

These developments already have taken place in large parts of several of our major cities. Their continuation presages the passing of political democracy as we now know it for the entire nation.

Application

With the 1960 Census of Population, Michigan became entitled to a nineteenth representative in the Congress of the United States. Where should the additional seat be located? What adjustments, if any, should be made in the boundaries of the eighteen existing districts? What principles should guide the work? The 1962 session of the Michigan State Legislature grappled with these problems but was unable to resolve them. In 1963 the legislative results were different. Three bills were introduced to resolve the problem, and after extensive discussion, negotiation, compromise, and amendment one of them -- Senate Bill 1334 -- became law.

Underlying enactment of this law was a general bi-partisan agreement that equality of population, plus or minus some fairly narrow deviation, should be the guiding factor in establishment of the new districts. Under the old apportionment three of the State's smallest congressional districts were in the City of Detroit, while three of the largest covered suburban communities elsewhere in the same county. This disparity ranged from one congressman for only 268,040 people in the old depopulated 13th Congressional District of Detroit and Highland Park to one congressman for all of 803,436 people in the largely suburban 16th Congressional District. Hence, a vote in some parts of Detroit actually was worth more than three times a vote in Wayne County's southern and western suburbs when it came to electing congressional representatives.

A similar problem existed in Outstate Michigan, where one congressman represented only 177,431 people in the sparcely populated 12th Congressional District of Michigan's Upper Penninsula. By contrast, the giant 6th, 7th, and 18th Congressional Districts all contained more than 600,000 people. (All figures come from the 1960 U. S. Census of Population.)

In order to approximate population equality, the geographic location of the new districts in the Detroit area had to shift toward the suburbs. The central city actually had lost population to those areas during the previous decennium. Elsewhere in the State, the geographic location of the new districts had to shift southward, away from the sparcely populated Upper Penninsula and northern Lower Penninsula.

At that time, a general consensus of legislative leaders agreed upon a 15% deviation from the mean congressional district population of 411,789. This was considered the preferable outer limit to individual district populations. However, deviations of 20% were considered satisfactory if special circumstances warranted the extra margin. Some of the considerations which were thought to justify some flexibility in the outer margins of the population range were the future direction of population growth and migration, efforts to avoid the disruption of established traditions, and preservation of adequate representation for minority groups.

As finally enacted, Senate Bill 1334 established the extensive 350-mile Upper Penninsula as the new 11th Congressional District, with the lowest population in the State. The 1960 Census reports that this sparcely populated area contained 305,984 people, 25.7% below the mean of Michigan congressional districts. Here, the homogeneity, traditions, and sheer expanse of the Upper Penninsula were important considerations in determining the final border. At the other end of the scale came the new 15th Congressional District, covering densely populated areas of Detroit and southern Wayne County. This area had a decade-long history of out-migration to other suburban areas. Its 1960 population was 490,310, 19.1% above the mean population of Michigan congressional districts.

Given the general agreement to aim for population equality, the next question was whether or not other considerations besides population were admissable guides in the drawing of district borders. Again there was general agreement that the gerrymandering of borders to create majority districts for partisan or personal advantage was reprehensible. There had been more than enough history on that score, both within and without the borders of the State. Michigan did not wish to repeat the errors of the past.

The sentiment on this account was so strong that some students of the subject concluded incorrectly in the complete inadmissability of any political motivations whatever. This school pushed aggressively for the use of unrelated districting concepts, such as "squareness," as the critical considerations second only to population equality.

Of course, we have suggested already that there is an important and well-defined distinction between efforts to assure the power of a given political group and efforts to provide a political framework in which the citizen maintains that freedom of choice so essential in representative government. Both efforts are "political." But while the first is reprehensible, the second aims to create institutions which assure that changing conditions in society do not imperil the democratic way of life. Madison, Jay, and Hamilton went to great lengths in The Federalist -- to cite only one of the better known examples from American political tradition -- to outline the types of new political institutions, the "checks and balances" necessary in their day to prevent an undue concentration of political power in the hands of a few. That problem is no less significant today.

Wayne County contains all of the City of Detroit and its eastern, western and southern suburbs. It contains more than a third of the total population of Michigan. It has been and it continues to be the major area of the State which exhibits the central characteristics of concern in this paper -- high population density combined with serious political imbalance. In only one of the six old Wayne County congressional districts did the individual independent voter have a chance to express himself effectively at the polls. In all five of the remaining districts the core vote of one of the parties had slipped well below 40%, and that condition had spelled the end of independent voter control over election results. For all practical purposes the two party system was facing gradual but certain extinction.

The new apportionment could have included five Wayne County congressional districts which would have been completely secure for one political group and one congressional district which would have been completely secure for the other. That apportionment was never seriously considered by anyone who worked on the Wayne County proposal, for it would have made a mockery of the representative process. Instead, the new apportionment created three congressional districts without overpowering domination by any political organization. These are districts where the individual independent voter is once again in control of election results. One hopes that this decision will provide the framework for active, aggressive, informative campaigns -- and for a resurrection of political democracy in areas where that institution was fast on the wane.

PERCENTAGE PENETRATION OF PARTY X
OLD 16TH CONGRESSIONAL DISTRICT OF MICHIGAN

Area	Strongest Candidates			Weakest Candidates			Maximum	
	1960	1962	1963	1960	1962	1963	Change	Spread
Grosse Ile Twp.	70.2	76.6	78.8	66.7	66.9	75.1	8.2	9.7
Belleville	65.5	65.9	71.4	62.0	60.7	65.2	5.5	6.2
West Dearborn	55.4	58.7	n.a.	49.1	47.4	n.a.	3.3	11.3
Canton Twp.	55.2	55.9	56.1	50.1	49.1	49.7	(1.0)	6.8
Wayne	49.8	53.4	56.2	45.8	44.9	50.0	5.1	8.5
Trenton	51.0	56.5	55.0	47.4	44.5	48.1	5.5	12.0
Gibraltar	49.2	50.3	44.1	45.4	41.3	38.1	(6.2)	9.0
Van Buren Twp.	51.9	50.0	46.3	47.8	44.1	42.8	(3.7)	5.9
Allen Park	40.6	46.9	45.7	35.9	34.7	33.5	6.3	12.2
Brownstown Twp.	43.9	45.3	39.7	39.5	37.0	35.8	(5.6)	8.3
Dearborn Twp.	39.2	44.8	40.2	33.5	34.4	35.8	5.6	10.4
Huron Twp.	41.1	44.6	41.5	38.2	38.3	38.5	3.5	6.3
Garden City	37.7	43.4	41.9	32.3	31.9	34.9	5.7	11.5
Nankin Twp.	38.4	43.2	37.2	34.3	34.6	33.2	(6.0)	8.6
Taylor Twp.	37.6	42.3	35.4	34.3	33.4	31.8	(1.9)	8.9
Romulus Twp.	37.3	37.8	38.8	34.2	32.9	35.7	2.8	4.9
Wyandotte	36.1	38.9	41.0	33.5	30.7	32.3	2.8 - (2.8)	8.7
Southgate	35.3	39.8	35.7	32.4	30.2	26.8	4.5	9.6
Riverview	33.1	38.5	31.9	30.3	28.9	26.6	(6.6)	9.6
East Dearborn	31.2	36.8	n.a.	26.5	26.1	n.a.	5.6	10.7
Detroit Ward 18	29.4	31.6	31.6	26.1	24.8	26.9	3.2	6.8
Sumpter Twp.	36.1	33.8	26.3	33.3	29.2	23.4	(7.5)	4.6
Lincoln Park	31.2	35.1	30.8	28.8	26.9	24.2	(4.3)	8.2
Melvindale	32.3	31.2	27.7	28.0	24.4	23.9	(3.6)	6.8
Ecorse	20.8	22.2	20.7	18.7	16.7	15.6	(2.0)	5.5
River Rouge	21.2	21.2	23.6	19.5	17.1	18.4	2.4 - (2.4)	5.2
Detroit Ward 20	19.3	19.5	17.9	17.3	15.3	14.8	(2.0)	4.2
DISTRICT TOTAL	37.1	41.3	40.0	33.8	32.5	33.9	4.2	8.8

THE RAILROAD DISABILITY RETIREMENT PROGRAM

Ira Marshak, U. S. Railroad Retirement Board

The disability retirement program administered by the United States Railroad Retirement Board has been in existence for 28 years. It is the oldest and most comprehensive program of its kind in the United States. For this reason, experience under this program is of interest not only to the Board and to the railroad industry, but also to pension experts and students of social insurance problems, generally.

This paper discusses the historical development of the program, the current provisions of law, and the relationship of the railroad program to the Board's sickness insurance program and to the disability benefit program of the social security system. The last section consists of a statistical summary of current and historical data on the program, including information on medical and other characteristics of beneficiaries. The actuarial data included were supplied by the Board's Actuary.

Development of the Disability Program

Background

The roots of the railroad disability program lie in the nineteenth century. Company pension plans began in this country in the railroad industry, which led all others in providing old-age and disability retirement benefits for employees. The first formal pension system in the industry providing old-age and disability benefits was established in 1874. By 1927, about 80 percent of all railroad employees worked for companies with pension plans but relatively few employees ever received benefits under such plans.

The financial base of the railroad pension structure was not strong enough to support it. With few exceptions, the plans were financed entirely by employers on a pay-as-you-go basis; that is, pension payments were treated as current operating expenses. As the benefit rolls grew, pension payments became more burdensome and generally could not be maintained during periods of adverse economic conditions. Widespread reductions in pension payments during the depression of the early 1930's brought about a movement which led to the development of a nationwide railroad retirement system on a contributory basis.

The Railroad Retirement Act of 1934 established the first contributory retirement system for nongovernmental employees in this country to be administered by the Federal government. It was declared unconstitutional by the Supreme Court and new legislation was enacted in 1935 designed to meet the constitutional objections to the 1934 act. These laws--the Railroad Retirement Act and the Carriers' Taxing Act of 1935--also became the subject of litigation, but the dispute never reached the Supreme Court. Instead, at the request of President Roosevelt, railroad management and labor formed a joint committee to negotiate their differences. The result was the system of retirement and survivor benefits embodied in the Railroad Retirement Act of 1937 and its companion taxing act.

The Railroad Retirement Act of 1937 is the foundation of the present system. Although that act provided for disability retirement, the conditions under which disabled employees could get benefits were severely limited. Unless the employee had reached age 60, he needed 30 years of service to qualify for a disability annuity.

The law was liberalized by amendments in later years in several important ways: (1) the service requirement for annuities based on total disability was reduced, (2) provision was made for annuities based on occupational disability, and (3) the Board received the authority to determine a "period of disability" within the meaning of the Social Security Act.

Current Disability Benefit Provisions

A railroad employee under age 65 who is permanently disabled, either for all employment or for his regular railroad occupation only, may receive a retirement annuity for life. The annuity terminates if he recovers from disability before age 65. Also, it is subject to deductions if his earnings exceed certain limits.

The amount of the annuity is computed according to the same formula used in calculating annuities to persons retiring at age 65 or over. No benefits to wives are payable until the employee reaches age 65. However, a special provision in the Railroad Retirement Act, called the "social security guaranty," may enable an employee to receive additional benefits if he had dependents who would have qualified for them under the Social Security Act. Under the railroad formula, the maximum annuity currently payable is $211. Under the special guaranty, the maximum is $279--10 percent more than the maximum family benefit of $254 payable under the Social Security Act.

An employee is eligible for an annuity based on total disability if his physical or mental condition disables him permanently for all regular employment and if he has at least 10 years of railroad service.

An employee is eligible for an annuity based on occupational disability if he has a "current connection with the railroad industry" and is permanently disabled for work in his regular occupation. If he has reached age 60, he needs 10 years of railroad service; if he is under age 60, he needs 20 years.

An employee has a current connection with the railroad industry if he worked for a railroad in at least 12 out of the last 30 months before his annuity begins. Any other 30-month period may be used if the employee had no regular nonrailroad employment after the 30-month period. Self-employment is not considered "employment" for this purpose.

Determinations of disability are made by Board medical and other personnel specially trained in the field of disability evaluation. Usually, the Board's physicians do not conduct examinations themselves, but base their determinations on medical evidence supplied by physicians of the employer and employee. If this evidence is not sufficient, a supplementary examination is made. In addition to medical evidence, the Board also considers nonmedical factors such as the employee's age, training, education, and work experience.

Because a beneficiary may receive his annuity only if he remains disabled, the Board systematically checks on his continuing eligibility for benefits. A variety of means are employed, including annual questionnaires covering his medical and employment status and exchanges of wage information with the Social Security Administration. As mentioned previously, the receipt of earnings in excess of specified amounts may result in deductions from annuity payments or in an investigation to determine whether the work record is consistent with his disability.

An employee under age 65 may not receive a disability annuity for any month in which he earns more than $100 either in employment-for-hire or in self-employment. If his total earnings for the year are less than $1,250, however, any amounts withheld are restored at the end of the year. If an annuitant earns $1,250 or more in the year, he loses one month's annuity for each $100 of earnings in excess of $1,200. (In calculating the excess, a remainder of $50 or more is counted as $100, while a remainder of less than $50 is not counted.) No deductions are made for any month in which the annuitant earns $100 or less. Penalty deductions may be made if a disability annuitant fails to report earnings promptly.

The railroad disability program is limited to cash benefits and does not include provision for vocational rehabilitation or training of disabled employees. However, in selected cases, disabled employees are referred to the appropriate State agency for possible vocational rehabilitation services.

Coordination with Sickness Benefits

Disability benefits under the Railroad Retirement Act are coordinated with sickness benefits which first became payable in 1947 under another law administered by the Board, the Railroad Unemployment Insurance Act. Many employees awarded disability annuities receive sickness benefits before their annuities begin. If an employee meets the eligibility conditions of both programs, he may still receive sickness benefits, but only to the extent that they exceed the amount of the annuity for the same period.

In the typical case, a disabled employee first applies for sickness benefits and later for a disability annuity. There are three important reasons for this procedure: (1) at the beginning of his disability, it may not be apparent that the employee will be permanently disabled, (2) sickness benefit payments can be started more quickly so that the employee has some income while he is waiting for the retirement annuity to be awarded, and (3) the sickness benefits are usually larger than the annuity.

Coordination with Social Security

In connection with its own disability program, the Board also makes determinations of "periods of disability" (disability freeze), within the meaning of the Social Security Act, for railroad employees with 10 or more years of service. To be eligible for a freeze, an employee must satisfy three requirements of the Social Security Act. He must be (1) totally disabled for at least 6 months, (2) fully insured on the basis of combined railroad and social security employment, and (3) have worked in at least one-half of the 10-year period before he became disabled. The criteria of disability used in establishing a freeze are approximately the same as those for permanent disability for all regular employment as used in adjudicating disability annuities under the Railroad Retirement Act.

An employee for whom a disability freeze period has been established may qualify for a higher annuity under the special guaranty than under the railroad formula. Under the guaranty, monthly benefits to an employee and his family equal 110 percent of the amount, or the additional amount, that the social security system would pay if railroad employment had been covered under that system. To illustrate the application of the provision, assume a disabled railroad employee could receive $150 a month, calculated under the regular railroad retirement formula. Assume, further, that the employee has dependents and that the Social Security Administration would have paid a total of $200 a month to this family if the employee had been in social security coverage. Therefore, the amount of annuity under the guaranty would be $220 a month. This entire payment would be made to the employee, since the Railroad Retirement Act does not provide for direct payments to the wife of an employee under age 65 or to children while the employee is alive. Even if the disability freeze does not provide higher retirement benefits, it may result in higher survivor benefits after the employee dies, since such benefits may also be computed under the guaranty provision.

Statistical Summary

The importance of the disability program may be illustrated by a number of summary figures. In calendar year 1962, about 10,300 annuities, averaging $125, were awarded to employees on the basis of disability. They accounted for one out of four retirement annuities awarded by the Board. The average annuity awarded does not reflect the effect of subsequent increases under the guaranty provision. The average increase for annuities recertified is about $50. Since about one out of five disability annuities is recertified under the guaranty, the completion of recertification actions will increase the average for all disability annuities awarded in the year by about $10, that is, from $125 to $135.

At the end of 1962, 41,900 disability annuitants under age 65 were receiving monthly benefits averaging $126. They made up 10 percent of the retired employees on the rolls and 5 percent of all individuals on the monthly benefit rolls. (There were also 57,600 annuitants age 65 or over who had been awarded annuities before age 65 on the basis of disability.) About $72 million--7 percent of benefits of all types paid by the Board in calendar year 1962--went to disabled employees under age 65.

The relative importance of the disability program may also be shown in terms of actuarial cost figures. According to the last actuarial valuation made as of December 31, 1959, disability annuities to individuals under age 65 accounted for 8.3 percent of the cost of all retirement benefits, and 6.2 percent of all benefits under the railroad retirement and survivor benefit programs.

Medical and Age Characteristics of Employees Awarded Disability Annuities

During the fiscal year ended June 30, 1962, a total of 11,500 applications were evaluated by the Board. In 9,500 cases--83 percent--the employee was found to be permanently disabled either for his regular occupation or for all regular employment, and an annuity was awarded. The distribution by primary disability group and by age of employee is shown in table 1. Classification by disability group is in accordance with the Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death, World Health Organization.

Diseases of the circulatory system were the most common disabling conditions. They accounted for 37 percent of the disability annuities awarded in the year. Other common disabilities were diseases of the bones and organs of movement, and diseases of the nervous system and sense organs. Altogether, these three groups accounted for two out of three annuities awarded in the year.

Employees aged 60-64 formed the largest age group in each major disability group. There was considerable variation, however, in the representation of particular disabilities in the different age groups. Diseases of the circulatory system were the primary cause for 16 percent of the employees under age 45, compared with 41 percent of those aged 60-64. In contrast, mental, psychoneurotic, and personality disorders occurred most frequently at the younger ages. These conditions accounted for 21 percent of the disability claims allowed for the employees under age 45, but for only 4 percent for those aged 60-64.

Table 2 shows the percentage distribution, by primary disability group, of individuals awarded disability annuities over the 10-year period 1952-53 to 1961-62. In evaluating these data, it should be kept in mind that some differences in the figures may be due to lack of consistency in classification by primary disability group. This is particularly noticeable in the group of "symptoms, senility, and ill-defined conditions," a "catch-all" group of cases in which the primary disability could not easily be determined. This group accounted for over 10 percent of the awards in 1956-57 but only 3 percent in 1959-60. The sharp drop in the size of this group after 1958-59 may reflect increased emphasis on classification by specific disability groups.

Notwithstanding these qualifications, there were certain notable changes in the relative frequencies of the various disability groups. The proportion of claims allowed based on diseases of the circulatory system generally declined over the period, falling from almost 42 percent in 1952-53 to less than 38 percent in 1961-62. On the other hand, the proportion of disabilities due to diseases of the bones and organs of movement rose from 12 to 15 percent.

There were also significant changes in the percentage of employees with some of the less common disabilities. The proportion with diseases of the respiratory system rose fairly steadily over the years, from only 2 percent in 1952-53 to 6 percent in 1961-62. The proportion with mental, psychoneurotic and personality disorders rose from 4 to 6 percent while the percentage with infective and parasitic diseases declined.

Awards of Disability Annuities, 1936-62

Table 3 shows selected information on awards of disability annuities since the beginning of the program. The table also includes data on disability annuities being paid at the end of each year to individuals under age 65.

Analysis of the characteristics and trend of awards of disability annuities calls for separate consideration of the periods 1936-46, and 1947 and later, because of the different eligibility requirements in the two periods. Except for a small number of annuities awarded under the superseded 1935 Act, a disability annuity could not be awarded before 1947 unless the employee was totally disabled for all regular employment. Also, as previously noted, if the employee was under age 60, he needed 30 years of service to qualify, the maximum service that could be credited towards annuities before 1967.

A large number of disabled employees filed applications for annuities after passage of the Railroad Retirement Act in June 1937, and 10,500 disability annuities were awarded in 1938. The number of awards declined sharply in the following years and leveled off at about 4,000 a year during the war years 1942-45. Disability annuities awarded during the years 1937-46 accounted for 21 percent of all retirement annuities awarded in the period.

The averages of years of service, ages at retirement, and amounts of annuities were generally very stable during the years before 1947, especially during the latter part of the period. In each of the years 1941-46, an average of 27.8 years of service was credited, the average age when the annuity began was 60.1 or 60.2 years, and the average annuity awarded ranged from $68 to $71.

The liberalization of the disability provisions under the 1946 amendments resulted in a sharp increase in the number of disability annuities awarded, from 5,000 in 1946 to almost 22,000 in 1947. The number dropped to 14,000 in 1948 and then gradually declined to under 10,000 a year by 1951 and has remained at about that level ever since.

The award of annuities based on occupational disability and the elimination of the 30-year service requirement for annuities beginning before age 60, had pronounced effects on the characteristics of annuities awarded after 1946. From 1946 to 1947, the average age at retirement fell from 60.1 years to 57.1, and the average years of service credited from 27.8 to 24.5. Reflecting the decline in average service, the average annuity awarded dropped from $70 to $64.

After the elimination of the initial backlog of employees who could first qualify for annuities under the liberalized conditions of the 1946 amendments, the average age at retirement rose slowly and stabilized at about 58 years. The average years of service remained just under 25 years through 1951. In 1952, this average increased sharply to 26.1 years as a result of the new 10-year service requirement established by amendments in 1951. Average service credited declined slowly in later years and by 1962 had fallen to 23.1 years.

Changes in the average annuity awarded after 1947 reflected mainly several increases in the annuity formula through amendments to the law. Annuities were increased by 20 percent in 1948, 15 percent in 1951, and 10 percent in 1956 and 1959. Largely as a result of these increases, average disability annuities awarded rose from about $70 in most of the years before 1948 to $120 in 1959. Increasing pay rates throughout the period caused a gradual rise in the average monthly compensation on which annuities were based. However, this effect on average annuities was minor compared with the effect of legislative increases in the scale of benefits.

The social security disability insurance benefit program established in 1957 had pronounced effects on the amounts of annuities awarded to disabled railroad employees through application of the social security guaranty. This was especially true after August 1958 when social security benefits could be paid to certain dependents of disability insurance beneficiaries. At the end of 1961, 20 percent of the disability annuitants under age 65 on the rolls were being paid under the guaranty provision. They were receiving on the average, $142, about $55 more than would have been paid to them under a railroad formula.

Rates of Disability Retirement by Occupational Group

Until recently, the disability retirement rates used in the periodic actuarial valuations of the railroad retirement system were perhaps the only published modern disability rates for a large industrial plan. These rates, which are based on experience under the railroad retirement system, have been used frequently for other industrial pension plans containing disability retirement features.

Table 4 shows the rates of "immediate" disability retirement under the Railroad Retirement Act in the years 1957-59, by occupational group. These rates are for employees who were in active railroad service at the time, or shortly before, they became disabled. They do not reflect retirements among employees who left the railroad industry some time before they became disabled, because of the difficulties of determining the number of such employees who may be still alive and not retired. A retirement is defined as immediate if the employee's annuity began in the same calendar year as the year in which he last worked, or in the following year. Immediate retirements account for the majority of all disability retirements (77 percent in 1958-59).

As may be seen in table 4, disability retirement rates were generally highest among maintenance employees. Among employees of all ages, the highest rates were for extra-gang, section and other maintenance of way employees, and the lowest for executives, supervisors and professionals. The adjusted rates shown in the table for all ages combined were standardized to permit more meaningful comparisons of rates by occupation. Based on the adjusted figures, for example, the disability rates for brakemen, switchmen, and hostlers are higher than for engineers and conductors while the reverse is true for the unadjusted rates.

The occupational differences in disability rates reflect the combined effects of socio-economic differences among employees working in the great variety of railroad occupations. These include rates of pay, stability of employment, working conditions, living standards, industrial hazards and level of medical care. The group with the highest disability rates--the extra-gang and section men--is characterized by lower pay rates, higher unemployment and sickness rates, and generally less favorable working and living conditions than occupational groups with lower rates of disability retirement.

Table 1.--Disability annuities awarded in fiscal year 1961-62, by age of employee and by disability group

Disability group	Total		Age in year disability began									
			Under 45		45-49		50-54		55-59		60-64	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Total------------------------	9,497	100	404	100	511	100	1,278	100	2,600	100	4,704	100
Infective and parasitic diseases---	183	2	25	6	16	3	36	3	52	2	54	1
Neoplasms-------------------------	542	6	30	7	38	7	77	6	173	7	224	5
Allergic, endocrine system, metabolic, and nutritional diseases---	291	3	12	3	21	4	37	3	77	3	144	3
Diseases of the blood and blood-forming organs--------------------	22	*	3	1	2	*	3	*	7	*	7	*
Mental, psychoneurotic, and personality disorders-------------------	592	6	86	21	58	11	93	7	169	6	186	4
Diseases of the nervous system and sense organs---------------------	1,377	14	79	20	99	19	213	17	351	14	635	13
Diseases of the circulatory system-	3,559	37	64	16	142	28	409	32	1,004	39	1,940	41
Diseases of the respiratory system-	599	6	14	3	17	3	76	6	188	7	304	6
Diseases of the digestive system---	158	2	8	2	14	3	36	3	41	2	59	1
Diseases of the genito-urinary system------------------------------	51	1	5	1	4	1	9	1	16	1	17	*
Diseases of the skin and cellular tissue---------------------------	19	*	2	*	0	*	3	*	5	*	9	*
Diseases of the bones and organs of movement----------------------	1,440	15	58	14	80	16	198	15	370	14	734	16
Congenital malformations----------	1	0	0	0	0	0	0	0	0	0	1	0
Symptoms, senility, and ill-defined conditions-----------------------	564	6	10	2	8	2	66	5	129	5	351	7
Amputations-----------------------	99	1	8	2	12	2	22	2	18	1	39	1

* Less than 0.5 percent.

NOTE.--Claims involving more than one type of disability are classified by primary disability. Classification by disability group is in accordance with Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death, World Health Organization. Detail may not add to totals shown because of rounding.

Table 2.--Percentage of disability annuities awarded, by disability group, fiscal years 1953-62

Disease group	Fiscal year									
	1952-53	1953-54	1954-55	1955-56	1956-57	1957-58	1958-59	1959-60	1960-61	1961-62
Number	8,535	9,524	8,875	9,642	8,330	9,459	9,982	8,896	9,218	9,497
Percent, total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Infective and parasitic diseases. .	3.9	3.9	3.3	3.5	3.1	3.3	3.1	2.5	2.3	1.9
Neoplasms	4.9	5.1	4.7	5.1	5.4	5.2	5.1	5.7	5.9	5.7
Allergic, endocrine system metabolic, and nutritional diseases. .	3.2	2.8	2.8	2.7	2.7	1.7	2.4	2.9	3.0	3.1
Diseases of the blood and blood-forming organs	.6	.3	.4	.2	.2	.3	.3	.3	.2	.2
Mental, psychoneurotic, and personality disorders.	4.1	4.2	3.9	3.8	4.1	5.7	5.4	5.4	4.8	6.2
Diseases of the nervous system and sense organs	15.6	16.1	16.8	17.5	17.0	16.0	16.1	17.0	15.4	14.5
Diseases of the circulatory system.	41.6	40.7	40.6	39.7	37.1	36.4	36.9	39.3	38.3	37.5
Diseases of the respiratory system.	2.2	2.6	2.8	3.0	3.2	5.2	5.2	5.9	6.2	6.3
Diseases of the digestive system. .	1.5	1.5	1.5	1.8	1.7	1.6	1.2	1.5	1.6	1.7
Diseases of the genito-urinary system.	.4	.3	.4	.4	.4	.5	.5	.5	.5	.5
Diseases of the skin and cellular tissue	.3	.4	.3	.5	.3	.4	.3	.3	.3	.2
Diseases of the bones and organs of movement	11.8	11.4	12.9	13.9	13.1	12.7	12.3	14.2	15.8	15.2
Symptoms, senility, and ill-defined conditions.	8.2	9.4	7.9	6.7	10.4	9.6	9.9	3.2	4.6	5.9
Amputations	1.6	1.4	1.4	1.1	1.2	1.5	1.4	1.2	1.1	1.0

NOTE.--Detail may not add to totals shown because of rounding.

Table 3.--Selected characteristics of disability annuities awarded and in current-payment status[1/]
1937-62

Calendar year	Awarded during year					In current-payment status at end of year[1/]			
	Number	Percent of all employee annuities	Average years of service	Average monthly amount	Average age at retirement	Number	Percent of all employee annuities	Average monthly amount	Amount paid in year (millions)
1937[2/]	2,500	6	29.6	$85	60.3	1,300	3	$85	$1.5
1938	10,500	22	29.2	72	60.4	10,400	13	75	12.4
1939	7,300	28	28.4	66	60.3	11,600	12	72	13.6
1940	6,000	27	28.4	71	60.3	13,000	12	72	14.0
1941	5,000	26	27.8	68	60.2	13,800	11	71	14.3
1942	4,000	27	27.8	68	60.2	13,300	10	70	18.5
1943	4,300	24	27.8	70	60.1	12,900	10	70	13.4
1944	4,200	23	27.8	70	60.1	13,500	9	70	13.5
1945	4,300	18	27.8	71	60.1	13,600	9	70	13.8
1946	5,000	19	27.8	70	60.1	14,100	8	70	14.4
1947	21,800	50	24.5	64	57.1	30,000	15	66	29.1
1948	14,100	43	24.6	78	57.3	36,000	17	71	35.8
1949	13,300	37	24.9	79	57.7	40,000	17	74	41.2
1950	12,300	34	24.6	79	58.0	42,800	17	75	44.6
1951	9,700	30	24.7	82	58.0	43,000	17	88	48.6
1952	8,800	28	26.1	98	58.0	42,500	16	91	53.1
1953	9,500	27	25.5	96	57.9	42,300	15	92	54.1
1954	9,400	26	25.3	97	58.0	42,200	14	92	53.9
1955	10,000	26	24.8	96	58.1	43,000	14	92	54.7
1956	9,200	25	24.4	106	58.1	43,100	13	101	57.4
1957	9,100	23	24.2	107	58.3	42,300	13	101	59.6
1958	10,000	23	23.7	108	58.2	42,400	12	102	60.5
1959	10,200	23	23.8	120	58.4	42,300	11	116	65.2
1960	10,700	22	23.8	130	58.0	42,200	11	122	72.0
1961	9,000	23	23.4	134	58.1	41,100	10	125	70.2
1962	10,300	23	23.1	[3/] 135	57.9	41,900	10	126	72.2

1/ Annuitants under age 65 only.

2/ Includes data for 100 annuities awarded in 1936 under the provisions of the 1935 Railroad Retirement Act.

3/ Includes estimated effect of recertifications made after initial award.

Table 4.--Rates of immediate disability retirement under the Railroad Retirement Act, by occupational group, 1957-59

(Crude rates of disability retirement per 1,000 exposed)

Occupational group	Age: All Ages, Adjusted	Age: All Ages, Unadjusted	Under 50	50-54	55-59	60-64
All employees	9.63	9.63	1.55	7.30	12.51	29.69
Office employees:						
Executives, supervisors, and professionals	4.90	5.37	*	*	5.32	19.18
Station agents, telegraphers, clerks and other office employees	7.51	7.42	1.12	7.01	9.92	21.95
Train-and-engine-service employees:						
Engineers and conductors	6.61	8.88	*	5.45	9.31	20.10
Firemen, brakemen, switchmen and hostlers	8.75	6.77	1.59	3.45	9.51	31.42
Maintenance employees:						
Way, structure and shop craftsmen, helpers and apprentices	10.62	10.85	1.34	8.54	13.67	33.33
Extra-gang, section and other maintenance-of-way employees	17.78	17.75	4.24	12.78	23.40	51.59
Shop and stores employees	13.16	12.61	*	10.53	15.70	43.58
Station and platform employees	11.30	10.15	*	6.69	19.16	33.14
All other employees	8.89	9.35	1.99	8.20	9.73	26.47

* Rate not calculated because of small number of retirements in sample.

NOTE.--Retirement is classified as immediate if the employee retired in the same calendar year as the year he last worked, or in the following year. Adjusted rates are calculated on assumption that age distribution for each occupational group is the same as for all groups combined. Data are based on 4-percent sample.

SOURCES AND VALIDITY OF MEDICAL STATISTICS WITH SPECIAL EMPHASIS ON DIAGNOSES

Barkev S. Sanders, Public Health Service

While we cannot map out the exact boundaries of medical statistics, all of us will agree that incidence and prevalence of disease represent a sizeable segment of such statistics, however defined.

The following are methods that have been used to ascertain prevalence of disease in a population:

1) A census or survey of the population to be studied;
2) Analysis and interpretation of causes of death related to fatality rates;
3) Hospital statistics by diagnosis;
4) Canvassing of physicians, or record-keeping by them, of the kinds of illnesses they treat;
5) Diseases notifiable by law;
6) Registers of persons with certain diseases or defects; and
7) Examination of representative samples of the population by a trained team of physicians with agreed-upon standards for differential diagnoses.

Each of these approaches has advantages and disadvantages. Since time does not permit a systematic consideration of them all, I shall limit myself to the first and last: population survey and clinical examination.

In the United States prior to 1950, the household survey was the preferred method of obtaining morbidity information, and to most people it still is. In this method an adult member of the household generally is asked to report for a specified time period the illnesses and conditions suffered by herself and other family members. The diagnosis, duration, and severity of illness are also reported, the latter chiefly by length of incapacitation and receipt of medical care, including hospitalization.

Even though morbidity surveys, historically speaking, are comparatively recent, the method of counting certain sick or handicapped persons in the population can be traced at least to Biblical times when periodic census became an instrument of statecraft.

In the United States in the 19th Century, some censuses included a count of sick and incapacitated persons with a few broad categories. The deaf, the mute, and the blind were enumerated by many Federal decennial and some State censuses even as late as 1930.

With the introduction of the categorical assistance program for the blind (Aid to the Blind) in the '30s, statisticians were startled to find many more blind persons eligible for aid than expected from the Census count. (1) While it would seem the Census took this disparity to heart and abandoned the counting of handicapped persons, many statisticians have been unwilling to abandon this method for obtaining prevalence of diseases. Therefore, much of our information on prevalence rates comes from specially designed morbidity surveys.

I. Evidence of Limitation of Surveys and Census

Despite the widespread use of this method, numerous examples indicate it is deficient, at least so far as prevalence 1/ of disease is concerned. To illustrate these deficiencies I will cite briefly instances from different types of studies.

Lack of Validation

Table 1, taken from the Commission on Chronic Illness Study in Baltimore, (3) illustrates the extent of discrepancy in morbidity findings derived from clinical examinations and from a household survey. It shows that, even in terms of broad diagnoses (2 place code), on the average, only 17 per cent of diseases diagnosed by a single clinical examination were reported specifically enough in the household survey to be given the same general code number. The percentage of matching varies widely for different diseases, ranging from 99 per cent for asthma, to 0 for syphilis, rheumatic fever, cervicitis, and arthritis.

[Insert Table 1]

It is probable that this matching would be even lower if it were possible to re-examine the sample, or at least, doubtful cases in the sample.

Another shortcoming of the survey method, as I see it, is the significant proportion of false positives, i.e., reporting of diseases which cannot be confirmed by clinical examination. Some statisticians, interested solely in a prevalence rate, have treated this deficiency as an asset, in that it offsets in part the large proportion of false negatives. But health workers, concerned not only with prevalence rates but with other characteristics of individuals with various diseases, find this a further deficiency.

Table 2, also based on the Baltimore Survey, shows that nearly one-half of all conditions reported in the household survey could not be validated by a single clinical examination. The extent of false positives in the Baltimore Study ranges from zero per cent for tuberculosis, diabetes, psychophysiologic disorders, rheumatic fever, angina pectoris, arteriosclerosis, and rheumatoid-and osteo-arthritis to 91.9 per cent

1/ The author has ready for publication a monograph entitled Evaluation of Morbidity Surveys as a Measure of Disease Prevalence, which presents these deficiencies in much detail based on recent surveys, including the survey of Kit Carson County, Colorado, with which the author was associated. See A Health Study in Kit Carson County, Colorado, Public Health Service Publication No. 844; especially Chapter II.

Table 1 - Percentage Match of Evaluation Diagnoses 1/2/3/ with Diagnostic Information Reported in the Household Survey, Baltimore, 1953-1955

Evaluation Diagnoses	Percent of Evaluation Diagnoses Matched by That Reported by Survey	Evaluation Diagnoses	Percent of Evaluation Diagnoses Matched by That Reported by Survey
All diagnoses	17.2	Diseases of kidney	15.3
		Heart Disease	13.5
Asthma	99.2	Rheumatic heart disease and rheumatic fever with heart involvement	13.3
Rheumatoid arthritis	97.8	Hypertension without heart involvement	13.2
Hay fever	73.3	Anemia	13.0
Symptoms referable to limbs and back	58.3	All other diagnoses	12.7
Other allergies	57.5	Benign neoplasms of uterus	12.6
Chronic sinusitis	44.7	Other heart disease	12.0
Blindness and impaired vision	44.2	Other mental, psychoneurotic and personality disorders	10.9
Other diseases of central nervous system	42.5	Psychoneuroses	10.2
Deafness and impaired hearing	42.3	Psychoses	9.5
Osteoarthritis	42.0	Cataract (not causing blindness)	9.0
Diabetes mellitus	37.2	Tuberculosis	7.8
Hemorrhoids	34.6	Neoplasms 4/	7.8
Coronary artery disease and angina pectoris	31.0	Diseases of thyroid	7.7
Diseases of gallbladder	30.9	Other diseases of circulatory system	6.5
Varicose veins of lower extremities	27.0	Hypertensive heart disease	6.1
Other forms of arthritis	22.9	Obesity 5/	3.7
Orthopedic impairments (n.e.c.) except cerebral paralysis	21.9	Benign neoplasms of other sites	3.4
Hernia of abdominal cavity	21.2	Arteriosclerosis	1.2
Vascular lesions of central nervous system	21.0	Psychophysiologic autonomic and visceral disorders	0.7
Low back strain	19.7	Syphilis	--
Cerebral paralysis (n.e.c.)	17.1	Migraine	--
Malignant neoplasms	15.9	Rheumatic fever without heart involvement	--
Other symptoms, senility, and ill-defined causes	14.5	Cervicitis	--
Diseases of female genital organs (except cervicitis)	13.8	Arthritis	--

1/ Based on weighted number of evaluation diagnoses.
2/ Excludes conditions which began during the interval between the household survey and the clinical evaluation. Includes could-report and could-not-report conditions.
3/ Reported in any terms which came in the definition of "high degree agreement". For definitions of "high" and "low degrees of agreement", see Appendix C of the source indicated below.
4/ Includes neoplasms of unspecified nature.
5/ For the special definition of obesity used in the evaluation clinic, see Chapter 12 and Appendix B of the source indicated below.

Source: Derived from the work of the Commission on Chronic Illness. Chronic Illness in the United States, Vol. IV -- Chronic Illness in a Large City - The Baltimore Study, The Harvard University Press, Cambridge, Mass., 1957. Table 113, pp. 304-305.

for diseases of the kidney. While repeated examinations would probably somewhat reduce the proportion of false positives, their net effect in increasing false negatives would be greater. Other validation studies like that of Baltimore, citations 2 to 6 inclusive, confirm in general this low level of diagnostic specificity from surveys.

[Insert Table 2]

Table 2 - Percentage of Survey Diagnoses 1/2/3/ which could not be matched with that from Clinical Evaluation. Baltimore 1953-1955.

Survey Diagnoses	Percentage False Positives in the Survey Using Evaluation Diagnoses as Criterion	Survey Diagnoses	Percentage False Positives in the Survey Using Evaluation Diagnoses as Criterion
All diagnoses	45.6	Hypertensive heart disease	38.4
Rheumatoid arthritis	0.0	Cerebral paralysis (n.e.c.)	43.3
Osteoarthritis	0.0	Heart disease	44.8
Tuberculosis	0.0	Diseases of gall-bladder	45.7
Obesity 4/	0.0	Neoplasms	46.2
Psychoses	0.0	Benign neoplasms of other sites	48.6
Rheumatic fever without heart involvement	0.0	Chronic sinusitis	50.4
Coronary artery disease and angina pectoris	0.0	Diseases of female genital organs (except cervicitis)	50.6
Psychophysiologic autonomic and visceral disorders	0.0	All other diagnoses	52.3
Cataract (not causing blindness)	0.0	Other heart disease	59.6
Arteriosclerosis	0.0	Rheumatic heart disease and rheumatic fever with heart involvement	59.8
Low back strain	1.6	Malignant neoplasms	60.5
Hernia of abdominal cavity	3.7	Vascular lesions of central nervous system	61.1
Diseases of thyroid	5.6	Anemia	64.2
Varicose veins of lower extremities	6.0	Symptoms referable to limbs and back	65.6
Benign neoplasms of uterus	6.2	Hypertension without heart involvement	73.4
Diabetes mellitus	6.6	Other symptoms, senility, and ill-defined causes	80.6
Hemorrhoids	16.7	Migraine	81.8
Deafness and impaired hearing	21.6	Other mental, psychoneurotic and personality disorders	82.8
Orthopedic impairments (n.e.c.) except cerebral paralysis	22.3	Other diseases of circulatory system	91.2
Other allergies	26.2	Diseases of kidney	91.6
Arthritis	27.6	Cervicitis	--
Other forms of arthritis	28.5	Syphilis	--
Psychoneuroses	29.5		
Blindness and impaired vision	34.0		
Hay fever	34.3		
Asthma	36.3		
Other diseases of central nervous system	37.3		

1/ Based on weighted number of evaluation diagnoses.
2/ Excludes conditions which began during the interval between the household survey and the clinical evaluation. Includes could-report and could-not-report conditions.
3/ Reported in any terms which came in the definition of "high degree agreement". For definitions of "high" and "low degrees of agreement", see Appendix C of the source indicated below.
4/ For the special definition of obesity used in the evaluation clinic, see Chapter 12 and Appendix B of the source indicated below.

Source: Derived from the work of the Commission on Chronic Illness. Chronic Illness in the United States, Vol. IV -- Chronic Illness in a Large City - The Baltimore Study, The Harvard University Press, Cambridge, Mass., 1957. Table 121, pp. 324-325.

Incompleteness of Survey Reported Conditions

Now let us consider how complete our diagnostic information is when drawn from surveys. During the 40's and 50's a number of studies were made to determine the effectiveness of multiphasic screening and of periodic health examinations (19) in finding undiagnosed conditions in various population groups.

I shall use findings from Elsom, et al (13) as illustrative of this deficiency.

In their study these authors analyze findings from periodic health examinations of 1513 executives of various firms in and around Philadelphia and one in the South. These examinees represented 96 per cent of all executives eligible for the periodic examinations. In the first examination, 906 conditions, previously unknown, were diagnosed, and in 822 executives who returned for a second examination 16 to 28 months later, an additional 389 new conditions were diagnosed. These 906 and 389 conditions were classified by three of the authors according to (A) potential seriousness of the disease or condition, (B) optimal effect of treatment, and (C) urgency of medical attention required. 2/ Final scoring represented the agreed-upon rating by the three authors who first classified each disease independently in terms of A, B, and C. Ratings for the most frequent diagnoses not known previously are shown in Table 3.

[Insert Table 3]

Of the 1513 persons receiving the initial examination, 612 were found to have previously unrecognized diseases. Of the remaining 901 persons, 428 (28 per cent) were diagnosed as healthy while 473 (31 per cent) did not have undiagnosed diseases discovered.

Of the 612 group, 57 per cent of the diagnosed diseases were regarded as serious, i.e., would result in death or major disability if unchecked; 34 per cent, minor; and 9 per cent insignificant. Effective therapeutic measures were deemed available for 93 per cent of the 906 conditions. Immediate treatment was considered urgent for only a small proportion. Of interest, too, is the fact that over half of the newly-diagnosed diseases were found in 13 per cent of those examined.

These findings are typical of other studies, some listed in references 7-19, showing that any population subjected to screening or physical examination yields many individuals with potentially serious diseases unknown to them and to their physicians. In the Baltimore Study, (3) for instance, the over-all morbidity rate as revealed by physical examination was 2.3 times that reported by the survey--counting false positives.

Elsom's study also shows that in this group of executives more suffer from undiagnosed conditions than from known diseases. The prevalence rate following the examination was more than twice what it would have been in terms of previously diagnosed conditions. If this is true for men in executive positions, plainly it would hold true, to even greater degree, for comparable age groups lower on the economic ladder, and in the sparsely populated areas with fewer health facilities and personnel and higher relative costs for health care.

Non-Replicability of Response

Aside from specific studies in the health field, a much larger volume of general information indicates that, by and large, unsolicited questioning of sample populations shows a low level of replicability in general, though this level varies for different types of information, for different groups, etc. Examples of these are found in the references 19 to 30.

A study by the Public Health Service in Nashville associating air pollution with prevalence of certain diseases is an example. In the household interview, in addition to other questions, a list of 32 diseases, mostly chronic, was used. A sub-group of those interviewed were later invited to the University clinic for medical examination.3/ The interval between the household survey and clinical examination was, on the average, one week.4/ Before examining the patient or taking his history, the physician used the same list of 32 diseases to ask the individual if he had ever had any of them. (For children under 15, questions were put to the mother, in both instances).

Table 4 shows that of 714 replies given to lay interviewer and the doctor in which there was an affirmative answer, only in 33 per cent the answer was consistent, i.e. "Yes" "Yes" to both. In 57 per cent the answer to the physician was Yes while the prior answer to the interviewer had been No. In 10 per cent the answer to the doctor's questioning was No whereas it had been Yes to the lay interviewer.5/ Thus in two-thirds of the cases the reply was reversed within one week, despite the fact that we may assume some correlation if only for the sake of self-consistency. And if there were some way of getting at the truth, it is highly probable that discrepancies would be even greater.

[Insert Table 4]

We can scarcely write off such discrepancies as due to "memory failure", "misunderstanding", or "changes in the disease picture". Moreover, for children under 15 for whom the mother was the respondent, in both instances, consistency of response was lowest of all, suggesting a purposeful distortion of information on the part of many. Only 26 per cent gave the same reply to both interviewer and physician.

Another example of non-replicability is from a study of respiratory symptoms among 144 mail carriers in Great Britain, (21) with interviews conducted by three physicians and three trained sick visitors. Two separate interviews were held approximately six weeks apart, based on the assumption that in four weeks consistency attributable to memory would be negligible. To pin down sources of inconsistencies, interviews

2/ For further elucidation of A, B, and C, see footnote to Table 3.

3/ Results of the clinical findings were not analyzed as far as is known.

4/ The maximum interval was 23 days.

5/ In this analysis all the "No" "No" replies and replies where the response to the doctor or to lay interviewer was different from "Yes" or "No" are excluded.

Table 3 - Ten Most Frequent Diagnostic Categories Among 1,513 Executives, with Grading and Frequency of Occurrence

Diagnosis		No. of Times Diagnosed†		
Class and Grade* A B C	Disease	1st Examination	2nd Examination	Total
3 1 2	Obesity	137	45	182
2 3 2	Hypertension	124	33	157
	Anorectal lesions	94	37	131
4 3 4	Cryptitis	1	0	1
4 3 4	Fissure	4	0	4
4 3 4	Granuloma	1	0	1
3 1 4	Hemorrhoids	76	33	109
4 3 4	Papillitis	3	1	4
4 3 4	Papillae, hypertrophied	8	1	9
4 3 4	Proctitis	1	2	3
	Colonic polyps	87	25	112
1 1 2	Benign	84	25	109
1 1 1	Malignant	3	0	3
	Prostatic lesions	47	20	67
2 1 3	Benign hypertrophy	37	14	51
1 1 1	Carcinoma	1	2	3
1 1 2	Nodule	1	0	1
3 2 3	Prostatitis	8	4	12
2 1 2	Inguinal hernia	35	5	40
1 3 2	Diabetes	28	7	35
3 3 3	Anxiety state, mild	16	9	25
	Arteriosclerotic heart disease	14	4	18
1 3 2	Angina pectoris	4	1	5
1 3 2	Arteriosclerotic heart disease (Unspecified)	6	1	7
1 3 2	Coronary heart disease	2	1	3
1 2 1	Myocardial infarction	2	1	3
2 2 2	Peptic ulcer	11	10	21
	Subtotal	593	195	788
	Total, all Diagnoses	906	389	1,295

* Class A, potential seriousness of disease. Results if not treated: grade 1, deaths; grade 2, major disability; grade 3, minor disability; and grade 4, insignificant disability. Class B, optimal effect of known treatment, likely to result in improvement of: grade 1, eradicated; grade 2, arrested; grade 3, ameliorated; and grade 4, outcome not affected. Class C, urgency of treatment required; grade 1, urgent within days; grade 2, not urgent but promptness still indicated; grade 3, early therapy helpful though not presently required; and grade 4, time not important in terms of inauguaration of therapy or progress of disease on basis of knowledge of disease.

† Figures to the right in columns "1st Examination", "2nd Examination", and "Total" indicate number of times a diagnosis listed under a general heading occurred.

Source: Elsom, K.A. et al: Periodic Health Examination, JAMA, Vol. 172, No. 1, Jan. 1960. p. 56/6

Table 4 - Responses to Doctor and Nonmedical Interviewer on Having Had Specific Diseases, by Response Groups, Nashville, Tenn. 1959*

Response Groups	"Yes" to Both	"Yes" to M.D. "No" to Interviewer	"No" to M.D. "Yes" to Interviewer	Total "Yes-"Yes" "Yes-"No" "No"-"Yes"
(1)	(2)	(3)	(4)	(5)
Number				
Responses of all individuals	239	404	71	714
Women responding for self	126	181	21	328
Men (wife responding)	53	95	20	168
Children under 15 (mother responding)†	35	79	20	134
All others, age 15 and over, not included elsewhere	25	49	10	84
Per cent				
Responses of all individuals	33.5	56.6	9.9	100.0
Women responding for self	38.4	55.2	6.4	100.0
Men (wife responding)	31.5	56.6	11.9	100.0
Children under 15 (mother responding)†	26.1	59.0	14.9	100.0
All others, age 15 and over, not included elsewhere	29.8	58.3	11.9	100.0

* Excludes "No-No" answers, doubtful answers, and failures to answer either the interviewer or the doctor.

† The mother was also the respondent for children under age 15 in the doctor's office; otherwise all persons in the doctor's office were their own respondents.

Source: Derived from Keilin, J.E.: The Use of the Information Statistic as a Measure of Conformity in Comparing Two Sets of Responses. (Processed.) Air Pollution Medical Program.

were taped and comparison of answers to identical questions showed that:

> "In one-third of the disagreements, the subject gave definite yet different answers to the correctly-asked question at the two interviews." (p. 184)

The authors further observe:

> "Much of the inaccuracy with which answers are reported is clearly irreducible by efforts on the part of the observer.

Consistent false positive and false negative answers, and some of the random variation in the two answers to the same question cannot be avoided." (p. 187) 6/

I have been fortunate to obtain from the authors of this British study the tabulations of the first and second responses. In Table 5 I show the replies on the first and second interview to one of the questions, i.e., "Does your breathing ever sound wheezy or whistly?"

[Insert Table 5]

Table 5 - "Yes" or "No" Response to the Question, "Does your breathing ever sound wheezy or whistly?" on two separate occasions.

	Response to 2nd Questioning 1/			
		Yes	No	Total
Response to 1st Questioning	Yes	65	12	77
	No	27	40	67
	Total	92	52	144

1/ Second questioning was 6 weeks after the first.

Source: Personal communication; for description of the study see Fairbairn, A.B.; Wood, C.H.; Fletcher, C.M.: Variability in Answers to a Questionnaire on Respiratory Symptoms. Brit. J. of Prev. & Social Medicine, Vol. 13, No. 4, Oct. 1959, pp. 175-189. (The figures for the four-fold tables are not given in the published article).

In the first interview, 77 of the 144 mail carriers answered Yes and 67 No. In the second questioning 92 answered Yes and 52 No. The full measure of discrepancy is not obtained by comparing 77 and 92. In the two questionings only 65 men said Yes both times. Twelve of the postmen who said Yes first, reversed themselves in the second questioning. On the other hand, 27 who said No first replied Yes the second time.

In studies where the primary purpose is to identify certain individuals in a population with specific attributes or characteristics, I have proposed the following replicability index which I have developed.

This index consists of $\frac{a-b}{a+c} = I$, where a and b are the first row of frequencies in a four-fold table and a and c are the frequencies in the first column. For the question of wheezing, this index would be $\frac{65-12}{92} = .58$. I shall not, however, attempt to discuss the index at this time, except to indicate that its limiting values are 1 where there is complete agreement between the replicates, and its minimum value approaches minus infinity. A pragmatic rule in using the index would be never to use a test which does not yield a positive index when related to some appropiate criterion which is being tested.

6/ See Ref. 21 (Fairbairn et al), Emphasis added.

For a third example I shall use information obtained from identifiable individuals in the Current Population Survey (CPS) sample in April, 1950, and April, 1960, compared with the information on unemployment with respect to these same individuals obtained in the decennial census of 1950 and 1960, respectively.(24) 7/

In Table 6 we can see that both in the 1950 and 1960 comparisons, unemployment was reported consistently for only about one-half of the males. For the other half, a different status, i.e., employed or not in the labor force was reported either in the CPS or in the Census. For women, unemployment was reported consistently for about one-third. For the other two-thirds the status was given either as employed or not in the labor force.

[Insert Table 6]

Psychological Studies of Interviewer Bias

Another indication of bias in surveys is demonstrated by various psychological experiments. Some of these findings are touched on in two recent papers by Rosenthal, and Rosenthal and Fode. (32)(33)

In one of the studies by these authors, conducted at the University of North Dakota, subjects rated numerically a number of photographs. Ten experimenters were used and 206 subjects. The experimenters were matched and assignment of subjects was random. Half of the experimenters, however, were led to expect a mean numerical score of plus 5 and the other half, minus 5 for the photographs that were to be rated. While experiments were conducted in identical manner, Table 7 shows that the means of experimenters with a minus bias for all 5 experimental pairs were appreciably below the means obtained by experimenters given a bias of plus 5. Obviously, the experimenters' belief influenced his subjects to respond differently. This influence was effected by experimenters' voice and gestures, with or without realizing it. Still other experimental studies found that experimenters in their turn are influenced by the replies. In other words, the interviewer does not remain a constant factor during the course of the survey. As a rule, the effect of such biases are overlooked in the interpretation and use of interview results.

[Insert Table 7]

Differentials in Diagnostic Skills of Doctors

Another consideration unheeded in the mass morbidity surveys is the marked differences among physicians in their diagnostic skills, practices,

7/ It should be observed that the figures given in Table 6 are the blown up estimates, apparently the comparisons could not be made in terms of individuals sampled--individuals interviewed.

Table 6 - Cross Classification of Estimated Number of Unemployed Based on Employment Status as Reported to CPS[1] Interviewers, for April, 1950 and 1960, Compared with that Reported to Census Enumerators for 1950 and 1960, Respectively, by Individuals Who Could Be Matched, by Sex of These Individuals.

		1950 Census[2]				1960 Census[3]			
		Male		Female		Male		Female	
		Number	Percent	Number	Percent	Number	Percent	Number	Percent
Estimated unemployment based on replies to CPS of individuals identified in the Census		2,551,000	100.0	874,000	100.0	1,985,000	100.0	1,146,000	100.0
Estimates of employment status of these same individuals as reported to the Census	Unemployed	1,271,000	49.8	262,000	30.0	1,027,000	51.8	360,000	31.4
	Employed	668,000	26.2	179,000	20.5	477,000	24.0	176,000	15.4
	Not in the labor force	612,000	24.0	433,000	49.5	481,000	24.2	610,000	53.2
Estimated unemployment based on replies to Census of identified individuals in CPS		2,057,000	100.0	699,000	100.0	1,897,000	100.0	1,028,000	100.0
Estimates of employment status of these same individuals as reported to CPS	Unemployed	1,271,000	61.8	262,000	37.5	1,027,000	54.1	360,000	35.0
	Employed	515,000	25.0	111,000	15.9	554,000	29.2	268,000	26.1
	Not in labor force	271,000	13.2	326,000	46.6	316,000	16.7	400,000	38.9

1/ CPS - Current Population Survey in April 1950 and April 1960, respectively.
2/ Employment and Unemployment Hearings before the Subcommittee on Economic Statistics of the Joint Committee, Congress of the United States, 87th Congress, 1st Session, Pursuant to Sec. 5(a) of Public Law 304, 79th Congress, Washington, 1962, p. 242.
3/ Measuring Employment and Unemployment, President's Committee to Appraise Employment and Unemployment Statistics, Table K. 4, p. 392.

Table 7 - Mean score ratings given to photographs by subjects in which 5 of the experimenters were told to expect a mean score of plus 5 and 5 other experimenters were told to expect a mean score of minus 5.

Results of Exp. 1: Mean Obtained Ratings for Each E

E Pairs	Sex of Es	High (+5) Bias N	High (+5) Bias M	Low (-5) Bias N	Low (-5) Bias M	Difference of Means
A	M	19	+ 3.47	21	+ 1.81	+ 1.66
B	M	20	+ 6.60	24	- 3.71	+10.31
C	F	21	+ 4.48	22	- 4.23	+ 8.71
D	M	18	+ 2.50	20	+ 1.70	+ 0.80
E	M	21	+ 3.05	20	+ 0.40	+ 2.65
Sums		99	+20.10	107	- 4.03	+24.13
Weighted Ms			+ 4.05		- 0.95	+ 5.00*

* Significant at .007 level.

Source: Rosenthal, R.; and Fode, K.L.: Psychology of the Scientist: V. Three Experiments in Experimenter Bias. Psychological Reports Monograph Supplement 3-V12 1963.

and in their inclination to examine patients for conditions not directly associated with their immediate complaint. Physicians also vary in how they advise their patients about examination findings. This is the information on which the survey interviewer must rely. The extent or importance of these differences is not yet known. I am familiar with only two studies where a systematic attempt has been made to study the qualifications of general practitioners and their diagnostic interest and skills; (34)(35) both show wide variability. 8/9/

Were such studies to be extended to all practitioners (including specialists), the range of variability of skills and insight would be materially greater. Without gauging the extent of these differences, comparisons which have been made between urban 10/ and rural groups have dubious significance for health workers. (36)(37)

8/ Thus, the North Carolina study found that 45 per cent of the physicians gave physical examinations to patients fully clothed. Findings in the Hunterdon County Survey (2) showed that of the doctors (85 per cent) who responded to questionnaires, 40 per cent made various diagnoses without the use of laboratory tests; 25 per cent said that they kept no written records.

9/ These studies were made not by students interested in morbidity surveys and their usefulness but by physicians interested in the quality of general practice and practitioners.

10/ These are NHS studies comparing prevalence of disease in different geographic regions; urban, rural-farm, rural-non-farm; and metropolitan areas. While these studies purport to show the effect of geographic differences in the prevalence of specific diagnoses, actually the differences could be attributed to numerous other variables, such as threshold of complaint; variation in physician skills; cooperation level, etc.

II. Superiority of Clinical Examination

In the health field where comparable diagnostic information is desired for different population groups, I believe that evidence shows the clinical examination of an appropriate sample (7) is decisively preferable to the household survey. To reduce clinical variability, the examining team should be adequately trained and every effort made to develop objective measures and agreed-upon standards for different diagnoses. Re-examination, where essential for accurate diagnosis, would be included.

For historical and other reasons, the clinical examination method has not been widely adopted. One of the objections is the higher unit cost. This should not be considered a barrier today, however, when large-scale surveys with costs running into millions are being conducted, considering the fact that far more reliable diagnostic information would be obtained perhaps at no greater aggregate cost.

A second limitation pointed to is the low rate of participation. In studies of the Commission on Chronic Illness (2)(3) only about two-thirds of the sample selected consented to the examination. The NHS has shown, however, that with the proper approach and preparation this percentage can be increased to 85-95. The participation rate is particularly likely to suffer if the desired procedure is followed and participants are scheduled for more than one examination. While this limitation is serious, it is not decisive, and again, participation can be improved with proper preparation. Moreover, the alternate survey method with its lack of precision has shown no significant difference in morbidity rates of examination participants and non-participants in any study where this has been tested. (2)(3)(4)

The third objection, which one can scarcely take seriously, is that the physical examination is not infallible. We grant that in science, as in life, nothing is perfect, but this does not mean that alternate procedures cannot be ranked according to their comparative precision. In fact, the main end of science, I suppose, is to reduce the margin of error in measurement. If it is true that certain diseases cannot be diagnosed with any degree of accuracy with our present clinical skills, let us ascertain which these are and attempt to develop differential diagnostic criteria. Let us not, however, use this as a justification for accepting the household survey as an effective method for determining prevalence of disease. Is it reasonable to expect accurate results from a method which we can see is subject to lack of quality control; to wide differences among physicians, both in their skill in diagnosis, and their practice in informing the patient about his condition; to distortions intentionally or otherwise introduced by the patient or the respondent; and to inherent variance among interviewers themselves? 11/

11/ See footnote 1.

References

1. Sanders, B.S.: The Blind - Their Number and Characteristics. Social Security Bulletin, Vol. 6, No. 10, Oct. 1943, pp. 17-26.

2. Trussell, R.E., and Elinson, J.: Chronic Illness in the United States, Chronic Illness in a Rural Area. The Hunterdon Study, Harvard University Press, Cambridge, 1959, Vol. III. 440 pp.

3. Commission on Chronic Illness in the United States. Chronic Illness in a Large City. The Baltimore Study, Harvard University Press, Cambridge, 1957, 620 pp.

4. Cobb, S.; Thompson, D.J.; Rosenbaum, J; Warren, J.E., and Merchant, W.R.: On the Measurement of Prevalence of Arthritis and Rheumatism from Interview Data. J. of Chronic Diseases, Vol. 3, No. 20, Feb. 1956. pp. 134-139.

5. Thompson, D.J., and Tauber, J.: Household Survey, Individual Interview, and Clinical Examination to Determine Prevalence of Heart Disease. Am. J. of Public Health, Vol. 47, Sept. 1957, pp. 1131-1140.

6. Eichhorn, R.L., and Morris, W.H.M.: Respondent Errors in Reporting Cardiac Conditions on Questionnaires. Proceedings of the Purdue Farm Cardiac Seminar. Sept. 10-11, 1958, pp. 46-50.

7. A Study of Multiple Screening. Council on Medical Service, American Medical Association, Revised 1955, 91 pp.

8. Journal of Chronic Diseases, Oct. 1955, pp. 363-490.

9. McDonald, G.W.; Remine, Q.R., and Durdick, E.J.: Results of Diabetics Screening Activities, Fiscal Year 1959, Public Health Reports, Vol. 76, pp. 825-831, Sept. 1961.

10. Remine, Q.R.: A Current Estimate of the Prevalence of Diabetes Mellitus in the United States; Ann. New York Academy of Science, Vol. 82, Sept. 1959, pp. 229-235.

11. Wilkerson, H.L.C.; Krall, L.P., and Butler, F.K.: Diabetes in a New England Town. Journal of American Medical Association, Vol. 169, Feb. 28, 1959, pp. 910-914.

12. McDonald, Glen W.; Hozier, J.B.; Fisher, G.F., and Ederma, A.B.: Large-Scale Diabetes Screening Program for Federal Employees. Public Health Reports, Vol. 78, No. 7, July 1963, pp. 553-560.

13. Elsom, K.A.; Schor, S.; Clark, T.W.; Elsom, K.O., and Hubbard, J.P.: Periodic Health Examination. Journal of American Medical Association, Vol. 172, No. 1, Jan. 2, 1960, pp. 55/5-60/10.

14. Lipkind, J.B.: Evaluation of Continuous Diabetes Screening in a Hospital Outpatient Department. Public Health Reports, Vol. 78, No. 6, June 1963, pp. 471-476.

15. Schenthal, J.E.: Multiphasic Screening of the Well Patient. The Journal of American Medical Association, Vol. 172, No. 1, Jan. 2 1960, pp. 51/1-54/4.

16. Roberts, N.J. (1957). Periodic Health-Maintenance Examinations in the Early Detection and Prevention of Disease, Edited by Hubbard, J.P. The Blackiston Division. McGraw-Hill Book Company, Inc., New York, pp. 27-57.

17. Mooney, H.W.: Methodology in Two California Health Surveys. Public Health Monograph No. 70, Public Health Service Publication No. 942, U. S. Department of Health, Education, and Welfare, Public Health Service, 1962, 143 pp.

18. Cochrane, A.L.; Chapman, P.J., and Oldham, P.D.: Observers Errors in Taking Medical Histories. The Lancet, Vol. 260, No. 6662, May 5, 1951, pp. 1007-1009.

19. Siegel, G.S.: Periodic Health Examinations, Abstracts from Literature, Public Health Service Publication 1010.

20. Keilin, E.J.: A Use of the Information Statistics as a Measure of Conformity in Comparing Two Sets of Responses; Division of Air Pollution, Public Health Service (processed).

21. Fairbairn, A.S.; Wood, C.H.; Fletcher, C.M.: Variability in Answers to a Questionnaire on Respiratory Symptoms. Brit. J. of Preventive & Social Medicine, Vol. 13, No. 4, Oct. 1959, pp. 175-189.

22. Douglas, J.W.B., and Blomfield, J.M.: The Reliability of Longitudinal Surveys. The Milbank Memorial Fund Quarterly, Vol. XXXIV, July 1956, No. 3, pp. 227-252.

23. Eckler, A.R.: Extent and Character of Errors in the 1950 Census, The American Statistician, Vol. 7, No. 5, Dec. 1953, pp. 15-21.

24. President's Committee to Appraise Employment and Unemployment Statistics; Measuring Employment and Unemployment; Government Printing Office, Washington, D.C. 1962, 412 pp.

25. Sanders, B.S.: Have Morbidity Surveys Been Oversold? American Journal of Public Health, Vol. 52, No. 10, 1962, pp. 1648-1659.

26. Health Interview Responses Compared with Medical Records. National Health Survey, Public Health Service Publication No. 584, Series D, No. 5.

27. Lilienfeld, A.M., and Graham, S.: Validity of Determining Circumsion Status by Questionnaire as Related to Epidemiological Studies of Cancer of the Cervix. J. of the National Cancer Institute, Vol. 21, No. 4, Oct. 1958, pp. 713-720.

28. Feldman, J.J.: The Household Interview Survey as a Technique for the Collection of Morbidity Data. Journal of Chronic Diseases, Vol. II, No. 5, May 1960, pp. 535-557.

29. Hyman, H.: Do They Tell the Truth? Public Opinion Quarterly, Vol. 8, 1944, pp. 557-559.

30. Parry, Hugh J., and Crossley, H.M.: Validity of Responses to Survey Questions. Public Opinion Quarterly, Vol. 14, 1950, pp. 61-80.

31. Scott, Christopher: Research on Mail Surveys. J. of the Royal Statistical Society, Series A, Vol. 124, Part 2, 1961, pp. 143-205.

32. Rosenthal, R.: On the Social Psychology of the Psychological Experiment: The Experimenter's Hypothesis as Unintended Determinant of Experimental Results. American Scientist, Vol. 51, No. 2, June 1963, pp. 268-283.

33. Rosenthal, R., and Fode, K.L.: Psychology of the Scientist: V. Three Experiments in Experimenter Bias. Psychological Reports, Monograph Supplement 3-V12 1963, pp. 491-511.

34. Peterson, O.L.; Andrews, L.P.; Spain, R.S.; and Greenberg, B.G. (1956): An Analytical Study of North Carolina General Practice. J. Med. Education, Vol. 31, Part 2, 12 pp.

35. Clute, K.F.: The General Practitioner. University of Toronto Press, 1963.

36. NHS Health Statistics, Series C, No. 5. Geographic Regions and Urban-Rural Residence, United States, July 1957-June 1959.

37. NHS Health Statistics, Series C, No. 6, Geographic Divisions and Large Metropolitan Areas, United States July 1957-June 1959.

XI

CONTRIBUTED PAPERS - II

Chairman, Eli S. Marks, Case Institute of Technology

TESTING A SET OF CORRELATION MATRICES FOR EQUALITY

Norman B. Rushforth, Division of Biometry and Developmental Biology Center, Western Reserve University

1. Introduction

A similar type of problem to that of testing the equality of correlation matrices exists when covariance matrices are considered. The likelihood ratio test for testing the equality of covariance matrices was given by Wilks (1946). If population covariance matrices are shown to be equal then the population correlation matrices must be equal. The converse is not necessarily true, however. In many psychological studies the variables are "standardized" by dividing covariances by corresponding standard deviation elements. This is often done when the underlying variables have no natural scale of measurement and the units are somewhat arbitrary.

The problem of testing the equality of a set of correlation coefficients from one population was recently solved by Lawley (1963). For the case of m correlation matrices each based on two variables a test procedure exists (Rao, 1952). This case, which is a problem of testing the homogeneity of a set of correlation coefficients, was solved by using the z transformation (Fisher, 1924). Brander (1933) derived the likelihood ratio statistic for testing the hypothesis of equal correlation matrices for the bivariate case with two populations. He showed the statistic to be a function of the difference between z transformed correlation coefficients. Efforts by the present author to generalize Brander's results to more than two populations and more than two variables were unsuccessful.

The purpose of this paper is to present two statistics B and M to test the equality of m correlation matrices based on p variables. Test procedures based on B and M are consistent. For the m population bivariate case (p = 2), assuming equal population correlations, a logarithmic function of B is asymptotically distributed as a chi square variable with m - 1 degrees of freedom. Thus B may be used to test the homogeneity of a set of correlation coefficients. For the two population (m = 2), trivariate case (p = 3) with equal sample sizes, a similar function of B is distributed asymptotically as a quadratic form in dependent normal variables, but not as a chi-square variable (Rushforth, 1961). The complexity of this distribution led to the abandonment of B as a statistic for testing the equality of p x p correlation matrices. The statistic M is based on z transformed correlation coefficients and is analogous in structure to the likelihood ratio statistic for the multivariate analysis of variance. Under the hypothesis of equal population correlation matrices a logarithmic function of M is asymptotically chi-square with $\frac{(m-1)p(p-1)}{2}$ degrees of freedom.

2. Properties of the B Statistic

The statistic B is defined as:

$$B^{\frac{1}{N}} = \frac{\left|\sum_{j=1}^{m} k_j R_j\right|}{\prod_{j=1}^{m} \left|R_j\right|^{k_j}}$$

where R_j is the sample correlation matrix of the j^{th} population with sample size N_j. $R_j = (r_{ivj})$ i, v = 1...p, j = 1...m. The quantity $k_j = \frac{N_j}{N}$, $N = \sum_{j=1}^{m} N_j$, and therefore $\sum_{j=1}^{m} k_j = 1$.

Thus $B^{1/N}$ is seen to be a ratio of determinants. The numerator is the determinant of the average sample correlation matrix for the m populations, each correlation matrix being weighted by k_j, the proportion of the total sample size attributed to the j^{th} population. The denominator is a product of the individual determinants of the sample correlation matrices each raised to the appropriate k_j^{th} power.

It can be shown (Rushforth, 1961) that the test procedure based on rejecting the hypothesis $P_j = P$, j = 1... m, in favor of the alternative hypothesis $P_j \neq P_w$, j = 1... m, w = 1... m, for some $j \neq w$ when $B^{1/N}$ is too large is consistent. Here $P_j = (\rho_{ivj})$ is the positive definite population correlation matrix with sample estimate $R_j = (r_{ivj})$ and P the common but unknown value of the matrices under the null hypothesis.

For the case of m bivariate populations the statistic $B^{1/N}$ is equal to

$$\left[1 - (\Sigma k_j r_j)^2\right]\left[\prod_{j=1}^{m} (1 - r_j^2)^{k_j}\right]^{-1}.$$ Here the null hypothesis $P_j = P$ is equivalent to $\rho_j = \rho$, j = 1... m. It can be shown (Rushforth, 1961) that under the null hypothesis ln B =

$$(1+\rho^2)\left[\sum_{j=1}^{m} Y_j^2 - (\sum_{j=1}^{m} \sqrt{k_j} Y_j)^2 + \text{No}(1/N)\right]$$

where Y_j is distributed asymptotically as N(0, 1) and $\lim_{N\to\infty} \text{No}(1/N) = 0$.

Since $\left[1 + (\sum_{j=1}^{m} k_j r_j)^2\right]$ converges stochastically to $(1+\rho^2)$ the quantity $\left[1 + (\Sigma k_j r_j)^2\right]^{-}$ ln B

is asymptotically distributed as

$\left[\sum_{j=1}^{m} Y_j^2 - (\sum_{j=1}^{m} \sqrt{k_j} Y_j)^2\right]$. This latter expression may be written as $\underline{Y}'A\underline{Y}$, a quadratic form in Y_j, where $\underline{Y}'$ is a m component vector of independent N (0, 1) variables. The m x m matrix A of rank m - 1 is equal to

$$\begin{bmatrix} 1 - k_1, & -\sqrt{k_1 k_2}, \ldots & -\sqrt{k_1 k_m} \\ -\sqrt{k_1 k_2}, 1 - k_2, & \ldots & -\sqrt{k_2 k_m} \\ & \ldots & 1 - k_m \end{bmatrix}$$

Since $\underline{Y}'$ is a m component vector with covariance matrix $\underline{I}$, the identity matrix, a necessary and sufficient condition for $\underline{Y}'A\underline{Y}$ to be distributed as a chi square variable with m - 1 degrees of freedom is AA' = A where A' is the transpose of A. Since A satisfies this condition then $\left[1 + (\sum_{j=1}^{n} k_j r_j)^2\right]^{-1} \ln B$ is asymptotically distributed as a chi square variable with m - 1 degrees of freedom.

3. Properties of the M Statistic

The statistic M is based on z transformed correlation coefficients and is analogous in structure to the likelihood ratio statistic for the multivariate analysis of variance. The Fisher z transformations for a sample correlation coefficient r and population correlation coefficient ρ are defined as

$$z = \frac{1}{2} \ln \frac{(1+r)}{(1-r)} \quad \text{and}$$

$$\zeta = \frac{1}{2} \ln \frac{(1+\rho)}{(1-\rho)} \quad .$$

Let $x_1 \ldots x_p$ be variables with the same first four moments as a standard normal variable. Let $r_{12}, r_{13} \ldots r_{1p}, \ldots, r_{p-1,p}$ and $\rho_{12}, \rho_{13} \ldots \rho_{1p}, \ldots, \rho_{p-1,p}$ be the sample and population correlation coefficients respectively for these variables. Define $z_{12}, z_{13} \ldots z_{1p}, \ldots, z_{p-1p}$ and $\zeta_{12}, \zeta_{13} \ldots \zeta_{1p}, \zeta_{p-1p}$ as the corresponding Fisher transforms of these correlation coefficients. Consider m populations defined by the parameters $\underline{\zeta}_j' = (\zeta_{12j}, \zeta_{13j}, \ldots \zeta_{1pj}, \ldots \zeta_{p-1pj})$. A test of the equality of m correlation matrices is equivalent to a test of the equality of the m vectors $\underline{\zeta}_j'$, j=1 ... m. For p underlying variables, there exists a p x p correlation matrix for each of the m populations. Since the matrix is symmetric and entries on the main diagonal are unity, $\frac{p(p-1)}{2}$ different correlational elements appear in the matrix. If Fisher transformations are made of the individual entries in the correlation matrix a $\frac{p(p-1)}{2}$ component vector $\underline{\zeta}_j'$ is obtained.

Equal population correlation matrices will give equal vectors. Thus a test of the equality of $\underline{\zeta}_j'$ vectors is equivalent to a test on the corresponding correlation matrices.

Consider the random vector $\sqrt{N}(z_{12} - \zeta_{12}, z_{13} - \zeta_{13} \ldots z_{1p} - \zeta_{1p}, \ldots, z_{p-1p} - \zeta_{p-1p}$. It can be shown (Rushforth, 1961) that $\sqrt{N}(\underline{z} - \underline{\zeta})'$ is asymptotically a multivariate normal distribution with mean vector $\underline{0}$ and covariance matrix Σ.

$$\Sigma = (u_{ijkl}) \; i, j, k, l, = 1 \ldots p, \; i < j, \; k < l,$$

where $u_{ijkl} = NE_L (z_{ij} - \zeta_{ij})(z_{kl} - \zeta_{kl}) =$

$$NE_L\left(\frac{(r_{ij} - \rho_{ij})(r_{kl} - \rho_{kl})}{(1 - \rho_{ij}^2)(1 - \rho_{kl}^2)} + o(1/N)\right)$$

The operator E_L indicates the moment of the appropriate entry of the covariance matrix of the limiting distribution. Now

$$\Sigma = NE_L(z_{ij} - \zeta_{ij})(z_{kl} - \zeta_{kl}) = \frac{1}{2(1 - \rho_{ij}^2)(1 - \rho_{kl}^2)}$$

$$\Big((\rho_{jk} - \rho_{ij}\rho_{ik})(\rho_{il} - \rho_{ik}\rho_{lk})$$
$$+ (\rho_{il} - \rho_{ij}\rho_{jl})(\rho_{jk} - \rho_{jl}\rho_{kl})$$
$$+ (\rho_{jl} - \rho_{il}\rho_{ij})(\rho_{ik} - \rho_{il}\rho_{kl})$$
$$+ (\rho_{jl} - \rho_{jk}\rho_{kl})(\rho_{ik} - \rho_{ij}\rho_{jk})\Big)$$

For the case k = i, l = j the above expression gives rise to the well known result (Hotelling, 1953)

$$NE_L(z_{ij} - \zeta_{ij})^2 = 1$$

Thus $\sqrt{N}(\underline{z} - \underline{\zeta})'$ is asymptotically normal with mean vector $\underline{0}$ and covariance matrix Σ, whose entries are given by $NE_L(z_{ij} - \zeta_{ij})(z_{kl} - \zeta_{kl})$.

If Σ^{-1} exists then $N(\underline{z} - \underline{\zeta})' \Sigma^{-1}(\underline{z} - \underline{\zeta})$ is asymptotically distributed as a chi square variable with $\frac{p(p-1)}{2}$ degrees of freedom.

Since Σ is the covariance matrix of the limiting distribution of $\sqrt{N}(\underline{z} - \underline{\zeta})'$ whose entries are functions of the population correlation coefficients, then $\hat{\Sigma}$ obtained by substituting the maximum likelihood estimates (r_{ij}) for (ρ_{ij}) is the maximum likelihood estimate of Σ and $\hat{\Sigma}^{-1}$ is the maximum likelihood estimate of Σ^{-1}. Now $\hat{\Sigma}^{-1}$ is a consistent estimate of Σ^{-1}. Therefore $N(\underline{z} - \underline{\zeta})' \hat{\Sigma}^{-1} (\underline{z} - \underline{\zeta})$ is asymptotically distributed as a chi square variable with $\frac{p(p-1)}{2}$ degrees of freedom.

For m populations, $\sqrt{N_t}\, \underline{z}'_t$, $t = 1 \ldots m$ are asymptotically $N(\underline{\zeta}'_t, \Sigma_t)$ under the alternative $\underline{\zeta}_u \neq \underline{\zeta}_v$ for some $u \neq v$. The statistic M used to test $\underline{\zeta}'_t = \underline{\zeta}'$ is defined as:

$$M = \frac{|\hat{\Sigma}_c|}{\left|\hat{\Sigma}_c + \frac{B}{N}\right|} \text{ where } B = (b_{ij}) =$$

$$\left[\sum_{t=1}^{m} N_t(z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j)\right] \quad \text{and}$$

$$N = \sum_{t=1}^{m} N_t$$

$$\hat{\Sigma}_c = \frac{\sum_{t=1}^{m} N_t \hat{\Sigma}_t}{\sum_{t=1}^{m} N_t} \text{ is the sample estimate of}$$

$$\Sigma_c = \frac{\sum_{t=1}^{m} N_t \Sigma_t}{\sum_{t=1}^{m} N_t}$$

Under the null hypothesis $\underline{\zeta}'_t = \underline{\zeta}'$ $t=1 \ldots m$ the quantity $-N \ln M$ is asymptotically a chi square variable with $1/2\,(m-1)p(p-1)$ degrees of freedom. Under the alternative hypothesis $-N \ln M$ is asymptotically distributed as a linear function of non-central chi-square variables. This distribution is difficult however, and was not investigated.

It can be shown, (Rushforth, 1961) that a test procedure based on rejecting $\underline{\zeta}'_t = \underline{\zeta}'$ $t=1 \ldots m$ when M is too small in favor of $\underline{\zeta}'_u \neq \underline{\zeta}'_v$ for some $u \neq v$ u, $v = 1 \ldots m$ is consistent.

4. Application of the Statistics B and M

The statistic $\left[\ln B\right]\left[1 + \left(\sum_{j=1}^{m} k_j r_j\right)^2\right]^{-1}$ is used to test the homogeneity of a set of correlation coefficients derived from measurements in a training evaluation study (Rushforth, 1958). In this study pre- and post- training measurements were made on 15 students attending a course in conference leadership.

The table below gives correlation coefficients between the after-training Q-sorts (Stephenson, 1953) of the 15 students and their corresponding before-training Q-sorts. In this situation a test of the homogeneity of correlation coefficients has practical meaning in terms of evaluating the training course. It is of interest to determine whether or not the group was uniform with respect to the pre- and post-training measurements. Are members of the group affected to the same extent by training or do some individuals change much more than others?

Table 1. Correlation Coefficients of Student's Before and After Training Q-sorts ($N_j = 100$, $j = 1 \ldots 15$, $m = 15$)

Student (j)	Correlation Coefficient (r_j)
1	.756
2	.600
3	.770
4	.616
5	.653
6	.786
7	.762
8	.733
9	.738
10	.616
11	.751
12	.754
13	.624
14	.727
15	.733

Under the null hypothesis that the 15 correlation coefficients are from the same population $(\ln B)\left[1 + \left(\sum_{j=1}^{15} k_j r_j\right)^2\right]^{-1}$ is approximately distributed as a chi square variable with 14 degrees of freedom. A convenient method of computing the statistic is first to compute $B^{1/N}$, from which it is easy to determine $\left[N \ln B^{1/N}\right]\left[1 + \left(\sum_{j=1}^{15} k_j r_j\right)^2\right]^{-1}$. Since each of the 15 correlation coefficients r_j was based on a sample size of 100, $N = 1500$ and $k_j = 1/15$ $j = 1 \ldots 15$. In this situation

$$B^{1/N} = \left[1 - (1/15 \sum_{j=1}^{15} r_j)^2 \right] \left[\prod_{j=1}^{15} (1 - r_j^2) \right]^{-1/15}$$

The statistic $\left[1 + (1/15 \sum_{j=1}^{15} r_j)^2 \right]^{-1} \left[N \ln B^{1/N} \right]$ is 18.8, which is compared with 23.68, the 95th percentile value of a chi square distribution with 14 degrees of freedom. Thus the difference of the correlation coefficients between the before- and after- training card sorts was not significant at the 5 per cent level.

The statistic $- N \ln M$ is used to test the equality of three correlation matrices based on five variables. The data are taken from a study of the responses of industrial executives to a semantic differential questionnaire concerning job concepts (Miller, 1960). They consist of the sample correlation matrices of three rater groups assessing the jobs of executives in 5 specialities, using 25 semantic differential scales. The data are presented below in tables 2, 3, and 4.

In testing the hypothesis that these three correlation matrices are from the same population a lengthy computation is required in order to obtain the value of $- N \ln M$. A computational form of this expression may be derived as follows.

$$M = \frac{|\hat{\Sigma}_c|}{|\hat{\Sigma}_c + B/N|} = \frac{|\hat{\Sigma}_c|}{|\Sigma_c| \, |I + \Sigma_c^{-1} B/N|} =$$

$$|I + \hat{\Sigma}_c^{-1} B/N|^{-1}$$

Therefore $- N \ln M = N \ln (|I + \hat{\Sigma}_c^{-1} B/N|)$

Expanding $|I + \hat{\Sigma}_c^{-1} B/N|$ to order N^{-1}

$$- N \ln M =$$

$$N \ln (1 + 1/N \sum_{i=1}^{q} \sum_{j=1}^{q} \hat{\sigma}^{ij} b_{ij} + o(1/N)) \quad i, j =$$

$1 \ldots q$ where $\hat{\sigma}^{ij}$ is the element of the i-th row and j-th column of $\hat{\Sigma}_c^{-1}$ and $q = \frac{p(p-1)}{2}$. Now $1/N \sum_{i=1}^{q} \sum_{j=1}^{q} \hat{\sigma}^{ij} b_{ij} + o(1/N)$ can be made arbitrarily small for sufficiently large N and therefore $\ln M$ may be expanded as a series to order N^{-1}. Thus

$$- N \ln M = \sum_{j=1}^{q} \sum_{j=1}^{q} \hat{\sigma}^{ij} b_{ij} + N\, o(1/N)$$

Now $\hat{\sigma}^{ij}$ is the element in the i-th row and j-th column of the inverse of the pooled covariance matrix $\hat{\Sigma}_c$. The value of $\hat{\Sigma}_c$ is an average of the individual sample covariance matrices. The symmetric matrix $B = (b_{ij})$ is the "between population" covariance matrix. The statistic $- N \ln M$ is equivalent therefore to the trace of the matrix obtained by pre-multiplying B by $\hat{\Sigma}_c^{-1}$. (i.e., the sum of the elements on the main diagonal).

As a first step in the test procedure, each correlation coefficient in the three matrices is transformed to a z variable. The individual sample covariance matrices $\hat{\Sigma}_t$ t=1, 2, 3 for the resulting z vectors are computed. The average covariance matrix $\hat{\Sigma}_c$ is computed from the sample matrices. Here it may be computed as an unweighted average since the sample sizes are approximately the same, $N_1 = 41$, $N_2 = 40$, $N_3 = 41$. The matrix $B = (b_{ij}) = \left[39 \sum_{t=1}^{3} (z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j) \right]$ is calculated from the z values for each matrix, $\bar{z}_i = 1 \ldots q$ being the average z_i for the three samples. The value of 39 was approximately equal to $N_t - 3$, t = 1, 2, 3. It is used here since $(N_t - 3)^{-1}$ is a closer approximation to the variance of z_{it} than N_t^{-1} (Anderson, 1958). The inverse of $\hat{\Sigma}_c$ is post multiplied by B. Operations of determining the inverse of $\hat{\Sigma}_c$ and the subsequent multiplication by B are best effected by means of an electronic computer. The trace of the matrix resulting from this multiplication is found to be 35.1.

Thus the observed value of the statistic $- N \ln M$ is 35.1 compare with the critical value of 31.41 for the 95 percentile value of a chi square variable with 20 degrees of freedom. The hypothesis of equality of correlation matrices is rejected at the 5 per cent level of significance. Therefore no common correlation matrix is assumed to describe the semantic differential evaluations of the five job categories made by personnel, engineering and production executives.

5. Summary

To test the equality of m correlation matrices based on p variables, the two statistics B and M are proposed. A test procedure based on B is consistent. For the m population bivariate case (p = 2), assuming equal population correlations, a logarithmic function of B is asymptotically distributed as a chi square variable with m - 1 degrees of freedom.

The statistic M is based on z transformed correlation coefficients and is analogous in structure to the likelihood ratio statistic for the multivariate analysis of variance. Under the hypothesis of equal population correlation matrices a logarithmic function of M is shown to be asymptotically chi-square with $\frac{(m-1)\,p\,(p-1)}{2}$ degrees of freedom. A test procedure based on M is consistent.

Application of the B statistic is illustrated in testing the homogeneity of a set of correlation coefficients from a training evaluation study. Use of the M statistic is demonstrated in testing the equality of correlation matrices derived in a study of the job concepts of selected industrial executive groups.

References

Anderson, T. W. (1958) An Introduction to Multivariate Statistical Analysis, New York: Wiley & Sons, 374 pp. (p. 78).

Brander, F. A. (1933) "A Test of the Significance of the Difference of the Correlation Coefficients in Normal Bivariate Samples", Biometrika: 25, pp. 102-109.

Fisher, R. A. (1924) "On a distribution yielding the error functions of several well known statistics", Proceedings of the International Mathematical Congress, Toronto, pp. 805-813.

Hotelling, H. (1953) "New Light on the Correlation Coefficient and its Transforms", Journ. Royal Stat. Soc. (Series B): 15, p. 193-232.

Lawley, D. N. (1963) "On Testing a Set of Correlation Coefficients for Equality", Ann. Math. Stat.: 34, 1. 149-151.

Miller, F. B. (1960) "Interim Report to the Social Science Research Center", Ithaca, N. Y.: Cornell University, 30 pp. (mimeographed).

Rao, C. R. (1952) Advanced Statistical Methods in Biometric Research, New York: Wiley & Sons, 390 pp. (p. 233).

Rushforth, N. B. (1958) "Evaluating Student Conference Leadership Training, A Study Utilizing Q-Technique". (M. S. Thesis, Cornell University, 1958) 131 pp.

Rushforth, N. B. (1961) "A Comparison of Sample Correlation Matrices and a Multivariate Analysis of Job Concepts of Selected Industrial Executive Groups" (Ph. D. Thesis, Cornell University, 1961) 140 pp.

Stephenson, W. (1953) The Study of Behavior; Q-Technique and its Methodology, Chicago: University of Chicago Press, 376 pp.

Wilks, S. S. (1946) "Sample Criteria for Testing Equality of Means, Equality of Variances, and Equality of Covariances in a Normal Multivariate Distribution", Ann. Math. Stat.: 17 pp. 257-281.

Table 2. Correlation Matrix of Scores for

Engineering Executives (N_1 = 41) rating

Five Job Categories

	Personnel	Sales	Accounting	Production	Engineering
Personnel	1.00	0.48	0.76	0.48	0.57
Sales		1.00	0.45	0.52	0.38
Accounting			1.00	0.50	0.31
Production				1.00	0.47
Engineering					1.00

Table 3. Correlation Matrix of Scores for

Production Executives (N_2 = 40)

	Personnel	Sales	Accounting	Production	Engineering
Personnel	1.00	0.57	0.54	0.61	0.54
Sales		1.00	0.32	0.62	0.44
Accounting			1.00	0.35	0.49
Production				1.00	0.65
Engineering					1.00

Table 4. Correlation Matrix of Scores for

Personnel Executives (N_3 = 41)

	Personnel	Sales	Accounting	Production	Engineering
Personnel	1.00	0.50	0.46	0.76	0.58
Sales		1.00	0.52	0.56	0.61
Accounting			1.00	0.51	0.50
Production				1.00	0.55
Engineering					1.00

A GENERALIZED CORRELATION COEFFICIENT: A STATISTIC FOR TESTING SIMILARITY BETWEEN LINKED OBSERVATIONS IN TWO SAMPLES

D. S. Burdick and H. H. Winsborough, Duke University

1. Frequently, in samples of observations, certain pairs of observations are suspected to be more (or less) nearly alike than is generally true for the population. For example, when the observations are made in sequence over a period of time, it may happen that successive observations resemble each other more than do observations which are not adjacent in time. Statistics which measure the relation between successive observations are Von Neumann's ratio and the serial correlation coefficient (1, 8).

Geary (3) proposed an obvious generalization of Von Neumann's ratio, which he called the contiguity ratio, to measure similarity of linked observations in a single sample when the linkage pattern is arbitrary. Geary used the contiguity ratio to test whether or not adjacent counties show similarity with respect to a certain attribute. Here an observation of the attribute for one county is linked with an observation for another county if the two counties are contiguous geographically.

The contiguity ratio is also useful in certain sociological situations. If observations are made on each person in a group of people, observations on persons who know each other might be expected to show a greater degree of similarity than observations on persons who are not acquainted. Winsborough, Quarantelli, and Yutzy (9) have discussed some of the applications of the contiguity ratio to sociology.

The case of two samples with links across the samples can also arise in sociological situations. The group of people may be divided into two categories, e.g., men and women. It may be desired to test whether or not men and women who are acquainted exhibit similar characteristics. A statistic for making such tests is introduced in this paper. This statistic is a generalization of the correlation coefficient in the same way that Geary's contiguity ratio is a generalization of the Von Neumann ratio.

In the next section the generalized correlation coefficient r will be introduced, and its mean and variance will be computed under the assumption that for each of the two samples the joint distribution of the observations is symmetric. Section 3 will contain examples of the application of r.

2. Let $x_1, \ldots, x_{n_1}$ be a sample from some population, and let $y_1, \ldots, y_{n_2}$ be another sample from another population. It is assumed that the joint distribution of the x_i's is symmetric and similarly for the y_j's. That is to say, different arrangements of the x_i's (y_j's) have the same likelihood of occurrence. Suppose further that there is a pattern of links between the x_i's and the y_j's. Let L be the number of these links. The generalized correlation coefficient r is defined by

$$r = \frac{1}{L} \sum_{i \ominus j} \sum u_i v_j \tag{2.1}$$

where

$$u_i = (x_i - \bar{x})/s_x, \quad v_j = (y_j - \bar{y})/s_y, \quad n_1 \bar{x} = \sum_{i=1}^{n_1} x_i, \quad n_2 \bar{y} = \sum_{j=1}^{n_2} y_j, \quad (n_1 - 1)s_x^2 = \sum_{i=1}^{n_1} (x_i - \bar{x})^2, \quad (n_2 - 1)s_y^2 = \sum_{j=1}^{n_2} (y_j - \bar{y})^2 .$$

The notation $\sum_{i \ominus j} \sum$ indicates that the summation extends over all pairs (i, j) for which x_i and y_j are linked.

The statistic r is not quite an exact generalization of the standard product-moment correlation coefficient. A sample from a bivariate distribution is expressable in the form of the previous paragraph with $n_1 = n_2 = L$, where each observation x_i is linked with the corresponding observation y_i. In this case the product moment correlation coefficient is usually defined as

$\frac{1}{L-1} \sum_{i \ominus j} \sum u_i v_j$ instead of $\frac{1}{L} \sum_{i \ominus j} \sum u_i v_j$. The difference is a trivial one, but the definition of r that we use permits a somewhat simpler expression for the variance of r.

It is worth mentioning that the range of possible values for r is unlimited. Of course, if the linkage structure is that described in the preceding paragraph, then r must satisfy $-(L-1)/L \leq r \leq (L-1)/L$, but for certain other linkage structures any real number is a possible value for r.

We now wish to derive the mean and variance of r under the assumption that each x_i is independent of each y_j regardless of whether or not x_i and y_j are linked. Then each u_i is independent of each v_j and therefore

(2.2) $E(r) = \frac{1}{L} \sum\sum_{i \ominus j} E(u_i v_j) = \frac{1}{L} \sum\sum_{i \ominus j} E(u_i)\, E(v_j)$

(2.3) $$E(r^2) = E\Big(\big(\frac{1}{L} \sum\sum_{i \ominus j} u_i v_j\big)^2 \Big) =$$

$$\frac{1}{L^2} \Big(\sum\sum_{i \ominus j} E(u_i^2)E(v_j^2) + \sum\sum\sum_{\substack{i \ominus j_1, j_2 \\ j_1 \neq j_2}} E(u_i^2)E(v_{j_1} v_{j_2}) + \sum\sum\sum_{\substack{i_1, i_2 \ominus j \\ i_1 \neq i_2}} E(u_{i_1} u_{i_2})E(v_j^2)$$

$$+ \sum\sum\sum\sum_{\substack{i_1 \ominus j_1, i_2 \ominus j_2 \\ i_1 \neq i_2 \\ j_1 \neq j_2}} E(u_{i_1} u_{i_2})E(v_{j_1} v_{j_2}) \Big) .$$

The summations above with unequal subscripts extend over all ordered pairs of the subscript which satisfy the linkage conditions. Thus if u_1 is linked to v_1 and to v_2, the term $u_1^2 v_1 v_2$ will occur twice: once as $u_1^2 v_1 v_2$ and again as $u_1^2 v_2 v_1$.

In order to evaluate (2.2) and (2.3) we evaluate $E(u_i)$, $E(u_i^2)$, $E(u_{i_1} u_{i_2})$, and $E(v_j)$, $E(v_j^2)$, $E(v_{j_1} v_{j_2})$. By definition of the u_i's and v_j's we have

$$\sum_{i=1}^{n_1} u_i = 0 = \sum_{j=1}^{n_2} v_j, \quad \sum_{i=1}^{n_1} u_i^2 = n_1 - 1, \quad \sum_{j=1}^{n_2} v_j^2 = n_2 - 1.$$

Since $\sum_{i=1}^{n_1} u_i = 0$, we have $0 = E(0) = E(\sum_{i=1}^{n_1} u_i) = \sum_{i=1}^{n_1} E(u_i)$. But since the joint distribution of the x_i's is symmetric, the joint distribution of the u_i's is also symmetric, and therefore $E(u_i)$ is the same for each i. Thus $0 = \sum_{i=1}^{n_1} E(u_i) = n_1 E(u_i)$ which implies that $E(u_i) = 0$. In a similar manner it can be shown that $E(v_j) = 0$, $E(u_i^2) = (n_1 - 1)/n_1$, and $E(v_j^2) = (n_2 - 1)/n_2$. To evaluate $E(u_{i_1} u_{i_2})$ observe $\sum_{i=1}^{n_1} u_i^2 + \sum\sum_{i_1 \neq i_2} u_{i_1} u_{i_2} = (\sum_{i=1}^{n_1} u_i)^2 = 0$.

Taking expectations and again making use of the symmetry of the joint distribution, we have $n_1(n_1 - 1)E(u_{i_1} u_{i_2}) = -(n_1 - 1)$ or $E(u_{i_1} u_{i_2}) = -\frac{1}{n_1}$.

Similarly, $E(v_{j_1} v_{j_2}) = -\frac{1}{n_2}$.

Substituting the results of the preceding paragraph into (2.2) and (2.3) yields

(2.4) $E(r) = \frac{1}{L} \sum\sum_{i \ominus j} E(u_i)E(v_j) = 0$

(2.5) $$E(r^2) = \mathrm{var}(r) = \frac{1}{L^2} \Big(\sum\sum_{i \ominus j} (n_1 - 1)(n_2 - 1)/n_1 n_2 +$$

$$\sum\sum\sum_{\substack{i \ominus j_1, j_2 \\ j_1 \neq j_2}} -(n_1 - 1)/n_1 n_2 + \sum\sum\sum_{\substack{i_1, i_2 \ominus j \\ i_1 \neq i_2}} -(n_2 - 1)/n_1 n_2 +$$

$$\sum\sum\sum\sum_{\substack{i_1 \ominus j_1, i_2 \ominus j_2 \\ i_1 \neq i_2 \\ j_1 \neq j_2}} 1/n_1 n_2 \Big).$$

To complete the evaluation we must count the number of terms in each of the sums in (2.5). The number of terms in the first sum is clearly L. The number of terms in the other sums can be expressed in terms of the following quantities: m_i = the number of links involving x_i, f_j = the number of links involving y_j. The number of ordered pairs of links to x_i is then $m_i(m_i - 1)$. Thus for each i the number of terms which occur in the second sum in (2.5) is $m_i(m_i - 1)$. The total number of terms in the second sum is therefore $\sum_{i=1}^{n_1} m_i(m_i - 1) = \sum_{i=1}^{n_1} m_i^2 - L$ since $\sum_{i=1}^{n_1} m_i = L$. Similarly, the total number of terms in the third sum is $\sum_{j=1}^{n_2} f_j(f_j - 1) = \sum_{j=1}^{n_2} f_j^2 - L$.

To obtain an expression for the number of terms in the fourth sum in (2.5) observe that if x_i and y_j are linked, the number of links which involve either x_i or y_j is $m_i + f_j - 1$. The number of links which involve neither x_i nor y_j is therefore $L - m_i - f_j + 1$. The total number of terms in the fourth sum is then $\sum\sum_{i \ominus j} (L - m_i - f_j + 1)$

$= L^2 + L - \sum_{i \leftrightarrow j} \sum m_i - \sum_{i \leftrightarrow j} \sum f_j$. In the sum $\sum_{i \leftrightarrow j} \sum m_i$ the term m_i appears once for each link involving x_i. Since there are m_i such links in all, we have $\sum_{i \leftrightarrow j} \sum m_i = \sum_{i=1}^{n_1} m_i^2$. Similarly, $\sum_{i \leftrightarrow j} \sum f_j = \sum_{j=1}^{n_2} f_j^2$. Thus, the number of terms in the fourth sum can be expressed as $L^2 + L - \sum_{i=1}^{n_1} m_i^2 - \sum_{j=1}^{n_2} f_j^2$. As a check on our results the total number of terms in all the sums should be L^2. We have $(L) + (\sum_{i=1}^{n_1} m_i^2 - L) + (\sum_{j=1}^{n_2} f_j^2 - L) + (L^2 + L - \sum_{i=1}^{n_1} m_i^2 - \sum_{j=1}^{n_2} f_j^2) = L^2$ as expected.

Substituting these results into (2.5) yields

$$E(r^2) = \frac{1}{n_1 n_2 L^2} \left((n_1 - 1)(n_2 - 1)L - (n_1 - 1)\left(\sum_{i=1}^{n_1} m_i^2 - L\right) - (n_2 - 1)\left(\sum_{j=1}^{n_2} f_j^2 - L\right) + L^2 + L - \sum_{i=1}^{n_1} m_i^2 - \sum_{j=1}^{n_2} f_j^2 \right) = \frac{1}{L^2} \left(L + L^2/n_1 n_2 - \frac{1}{n_1} \sum_{j=1}^{n_2} f_j^2 - \frac{1}{n_2} \sum_{i=1}^{n_1} m_i^2 \right).$$

We can summarize the results of this section by

(2.6) $E(r) = 0$.

(2.7) $\mathrm{Var}(r) = \frac{1}{L^2} \left(L + L^2/n_1 n_2 - \frac{1}{n_2} \sum_{i=1}^{n_1} m_i^2 - \frac{1}{n_1} \sum_{j=1}^{n_2} f_j^2 \right)$

The generalization of the assumptions made in deriving the mean and variance of r is worth emphasizing. The assumption of symmetry of the joint distributions is satisfied whenever the two samples are random samples from any two populations. This assumption is also satisfied when all within sample permutations of the observations are considered equally likely to have occurred. It is understood that the permutations in question do not affect the linkage structure, i.e., if x_i and y_j are linked originally and after permuting x_i becomes x_i' and y_j becomes y_j', then x_i' and y_j' will be considered linked for the purpose of computing r.

In the case where $L = n_1 = n_2 = n$, $m_i = f_j = 1$ the statistic r is just the product-moment correlation coefficient multiplied by $(n-1)/n$ as was mentioned earlier in this section. We have for this case

$$\mathrm{Var}\left(nr/(n-1) \right) = \frac{n^2}{(n-1)^2} \cdot \frac{1}{n^2} \left(n + \frac{n^2}{n^2} - \frac{n}{n} - \frac{n}{n} \right) = \frac{1}{n-1}.$$

This result under the assumption that the x_i's and y_j's are independent continuous variates is given as an exercise on page 396 of the book by Kendall and Stuart (5).

3. Potential uses of the generalized correlation coefficient in social research are easy to suggest. This section will describe several uses of the statistic and present two examples: one a reanalysis of data drawn from a classic sociological investigation and another using unpublished data.

In introducing the generalized correlation coefficient, we have alluded to its applications to the analysis of a sociometric matrix. A number of problems in social research seem formally identical to the problem of assessing the similarity or dissimilarity of two kinds of persons linked by friendship ties. In studies of formal organizations, for instance, one might be interested in the similarity in output between linked members of two strata within the bureaucracy. Substantively, links in this problem might be defined as ones of friendship, the flow of work, the pattern of consulting, or participation in the same chain of command.

Consider, for example, the well known Roethlisberger and Dickson investigations of the bank wiring room (7). This study investigated one work group involved in the assembly of telephone equipment. Within the work group were wiremen, who performed one kind of operation and solderers, who performed another. Among these men was a complex net of social relationships: some being friendly, some playing games together at lunch hour, some arguing about whether the window should be open or closed, some generally antagonistic to one another. A signal interpretation of this investigation was that the level and the quality of output of the men in the bank wiring room was related to their position in this net of social relationships. This interpretation, however, was intuitively derived from inspection of the relationships and the production scores -- a method not well suited to working out the complexities of what kinds of links are associated with what kinds of measures.

In a recent paper the contiguity ratio has been used to re-analyze some of this data with rather interesting results(9). There it seemed that sanctioning of deviant production levels operated less through refusal to interact (i.e., play games together at lunch), and more through the expression of sentiment (i.e., the expression of antagonism). That investigation, however, considered only the relationships between the wiremen and ignored the solderers. Using the generalized correlation coefficient, however, it is possible to investigate whether, for instance, wiremen and solderers playing games together have similar output. We can do this in spite of the fact that the mean levels of output and the variance of output for the two groups are different.

We undertook such a re-analysis, primarily as illustrative of the uses of the technique. A fairly full set of connections between solderers and wiremen exists when connection is defined as playing games together. Output quantities are not available for solderers but measures of the quality of work are available for solderers and wiremen. The generalized correlation for quality of output between solderers and wiremen linked by playing games together was found to be .39. When links are defined as who gets into arguments about windows, another fairly full set of connections, the generalized correlation is lower, .26.

It may very well be that the sample sizes in these cases and the number of connections are too small to presume that the generalized correlation is normally distributed under the null that $r = 0$. We have worked out the variances for these two examples according to formula (2.7) anyway and find that by the normal test the correlation for playing games can be regarded as significant at the one per cent level while the correlation for arguments cannot be regarded as significant even at the five per cent level.

This finding, tenuous though it may be, seems of some heuristic value. Within the group of wiremen it had been previously suggested that participation in games was not related to output. Between wiremen and solderers we have some indication of a relationship. The possibility that patterns of interaction may be differently related to output within and between occupational groups provides an interesting addition to Homans's discussion of the relationship between interaction and the variable he describes as "activity" (4).

In the preceding example the coefficient has been used in its most general form. A special case seems of enough importance to deserve separate comment. This is the case in which members of one sample are connected to several members of the second sample but members of the second sample are connected to only one member of the first. This condition obtains in many hierarchical situations. It obtains in the relationship between aggregates and their parts. It also obtains in the relationship between families over a generation.

This last observation suggests the potential applications of the statistic to the study of inter-generational social mobility. The generalized correlation coefficient provides the possibility of measuring the similarity between indicators of the socio-economic status of fathers and that of all their sons. In providing this measure, the coefficient controls differences in the mean and variance of socio-economic status between the two groups--differences which one would attribute to structural changes in the society rather than to the degree of openness of the society. A generalized correlation computed between socio-economic status of a sample of fathers and their sons would accomplish this standardization more accurately than many present techniques by using a more accurate estimate of the mean and variance of the variable in each generation.

Consider the following example. A recent study investigated retirement and pre-retirement problems of a non-random sample of older white couples living in the Piedmont region of North Carolina.[1] This study collected data on the educational level and the present or last occupation of men in the sample. The present occupational and educational levels of each of the sons in the labor force were also collected. Using a recently devised index of socio-economic status for occupations (6) and years of education as variables and combining both generalized correlations between fathers and sons and Pearsonian correlations between variables within generations, it is possible to to provide an interesting description of the mobility process within this sample. Table 1 provides these data. In that table .79 is the Pearsonian correlation between the level of education and the index of occupational status for fathers and .72 the Pearsonian correlation between the variables for sons. (These values are rather higher than similar correlations found

---

1. The study was financed through a grant made by the Ford Foundation for "Socio-Economic Studies of Aging." The data were collected between March 1960 and March 1961.

in other samples -- probably a result of over-representation of the tails of the occupational and educational distribution in the sample.) Other values in the table are generalized correlations.

The findings are interesting in that their pattern supports the model of inter-generational mobility suggested recently by Duncan and Hodge (2). In both generations the association between education and occupation is high while the intergenerational association between fathers' and sons' occupation, although significant, seems lower. It is also interesting that all generalized correlations are of about the same order of magnitude, with the association between fathers' occupation and sons' education being the highest and the association between fathers' education and sons' occupation being the lowest. This finding, perhaps, supports the notion that the major factor in the inheritance of status is related to the father's ability to purchase an education for his son.

One would be disinclined to push the analysis of these data farther because of the unsatisfactory nature of the sample. These findings, however, seem to indicate that further investigation of the use of the generalized correlation in the study of intergenerational mobility may be fruitful.

In summary, then, we feel that the generalized correlation coefficient is a statistic which should be a useful tool in the sociologist's repertoire -- one which deserves both empirical use and mathematical elaboration.

Table 1

Pearsonian and Generalized Correlations Among Level of Education and Occupational Socio-Economic Status of Fathers and All Their Sons[a]

	Fathers'		Sons'	
	Educa.	Occup.	Educa.	Occup.
Fathers'				
Education	--	.79	.55	.53
Occupation		--	.57	.56
Sons'				
Education			--	.72
Occupation				--

[a] All correlations are significantly different from zero at the .01 level.

REFERENCES

1. Dixon, Wilfred J., "Further Contributions to the Problem of Serial Correlation," Ann. Math. Stat., Vol. 15 (1944), pp. 119-144.
2. Duncan, Otis Dudley and Hodge, Robert W., "Education and Occupational Mobility: A Regression Analysis," Am. Jo. Sociology, Vol LXVIII (1963), pp. 629-644.
3. Geary, R. C., "The Contiguity Ratio and Statistical Mapping," Incorporated Statistician, Vol. 5 (1954), pp. 115-145.
4. Homan, George C., The Human Group. New York: Harcourt, Brace and Co., 1950, pp. 118-119.
5. Kendall, M. G. and Stuart, Alan, The Advanced Theory of Statistics Vol. 1, Three Volume Edition. London: Charles Griffin and Company, Ltd., 1958.
6. Reiss, Albert J., Jr., with collaborators, Occupations and Social Status. New York: The Free Press, 1961, Appendix B.
7. Roethlisberger, F. L., and Dickson, William J., Management and the Worker. Cambridge: Harvard University Press, 1946, pp. 379-548.
8. vonNeumann, J., "Distribution of the Ratio of the Mean Square Successive Difference to the Variance," Ann. Math. Stat., Vol. 12 (1941), pp. 367-395.
9. Winsborough, H. H., Quarantelli, E. L., and Yutzy, Daniel, "The Similarity of Connected Observations," Am. Sociological Rev., (forthcoming).

SOME SMALL SAMPLE RESULTS FOR THE VARIANCE OF A RATIO

Elizabeth Lauh and Wm. H. Williams
Bell Telephone Laboratories, Incorporated

1. Introduction

Ratio estimators are used to increase the reliability of the sample estimates. For each sampled unit an auxiliary variate, x, is observed in addition to the variate of interest, y, so that for each unit a pair (y,x) is obtained. From a random sample of n pairs (y_i,x_i) $i = 1,2,\ldots n$ a common survey problem is the estimation of population mean, $\bar{Y}$, subject to the assumption that the population mean $\bar{X}$ is known exactly. Ratio estimators are designed to utilize this information and the most common ratio estimator of $\bar{Y}$ is $\tilde{y} = (\bar{y}/\bar{x})\bar{X}$, which is generally biased. A study of certain statistical properties of the ratio-of-sample-means estimator of $\bar{Y}$ is the main interest of this paper.

The sampling variance of $\tilde{y}$ depends upon the sampling variance of $\bar{y}/\bar{x}$ for which there is no known exact expression. An approximation to this sampling variance, obtained by the use of a Taylor expansion, is well-known and can be found in most textbooks. A sample variance is usually formed by substituting sample expressions for the population quantities which appear in the sampling variance approximation. For large samples, the relationship between this sampling variance approximation and the sample variance thus obtained has been examined analytically and a relatively clear picture of the situation obtained. For small samples, the relationship is not so clear; nor is the relationship between the sampling variance approximation and the exact sampling variance. Unfortunately, these questions do not lend themselves to simple and productive analytic study and so this paper presents some Monte Carlo results for the sample sizes $n = 2(1)9$.

2. The Ratio Estimator Approximations

2.1 The Bias

Approximations to both the bias and sampling variance of $\bar{y}/\bar{x}$ can be found by the use of Taylor expansions. If the negative exponent of the identity

$$\frac{\bar{y}}{\bar{x}} = \frac{\bar{Y}}{\bar{X}}\left(1 + \frac{d\bar{y}}{\bar{Y}}\right)\left(1 + \frac{d\bar{x}}{\bar{X}}\right)^{-1}, \qquad (1)$$

where

$$d\bar{y} = \bar{y}-\bar{Y} \quad \text{and} \quad d\bar{x} = \bar{x}-\bar{X},$$

is expanded as

$$\left(1 + \frac{d\bar{x}}{\bar{X}}\right)^{-1} = \sum_{k=0}^{\infty}(-1)^k\left(\frac{d\bar{x}}{\bar{X}}\right)^k \qquad (2)$$

then a first approximation to the bias is

$$\text{Bias}_1\left(\frac{\bar{y}}{\bar{x}}\right) = \frac{\bar{Y}}{\bar{X}}\,(C\bar{x}\bar{x}-C\bar{y}\bar{x}), \qquad (3)$$

where

$C\bar{y}\bar{x}$ denotes $\text{Cov}(\bar{y},\bar{x})/\bar{Y}\bar{X}$

and

$C\bar{x}\bar{x}$ denotes $V(\bar{x})/\bar{X}^2$.

The expansion of Equation (2) is valid for $\left|\frac{d\bar{x}}{\bar{X}}\right| < 1$ and terms up to the second order have been retained to obtain Equation (3).

2.2 The Variance

The variance of $\bar{y}/\bar{x}$ is, by definition,

$$V(\bar{y}/\bar{x}) = E(\bar{y}/\bar{x})^2 - [E(\bar{y}/\bar{x})]^2, \qquad (4)$$

and it is convenient to consider the two terms on the right-hand side separately. Squaring both sides of Equation (1) and using the expansion,

$$\left(1 + \frac{d\bar{x}}{\bar{X}}\right)^{-2} = \sum_{k=0}^{\infty}(-1)^k(k+1)\left(\frac{d\bar{x}}{\bar{X}}\right)^k, \qquad (5)$$

the second order approximation to $E(\bar{y}/\bar{x})^2$ is obtained as,

$$E\left(\frac{\bar{y}}{\bar{x}}\right)^2 = \frac{\bar{Y}^2}{\bar{X}^2}\,(1 + 3C\bar{x}\bar{x} - 4C\bar{y}\bar{x} + C\bar{y}\bar{y}). \qquad (6)$$

Next, squaring $E\left(\frac{\bar{y}}{\bar{x}}\right)$, expanding, and dropping third and higher order terms gives

$$[E(\bar{y}/\bar{x})]^2 = \frac{\bar{Y}^2}{\bar{X}^2}\,(1 + 2C\bar{x}\bar{x} - 2C\bar{y}\bar{x}), \qquad (7)$$

so that subtraction from Equation (6) yields a first approximation to $V(\bar{y}/\bar{x})$ as

$$V_1(\bar{y}/\bar{x}) = \frac{\bar{Y}^2}{\bar{X}^2}\,(C\bar{y}\bar{y} + C\bar{x}\bar{x} - 2C\bar{y}\bar{x}). \qquad (8)$$

The large sample variance of $\tilde{y}$ can then be written as,

$$V_1(\tilde{y}) = \frac{1}{n} \bar{Y}^2(Cyy + Cxx - 2Cyx). \qquad (9)$$

These large sample approximations to the bias and variance of $\tilde{y}$ are those usually found in the literature, [1,3], and seem to hold for large n, see for example, Cochran, [1, p. 114].

An approximation to the mean square error can be found by the same procedure, see Kish [3] and Sukhatme [5].

The approximation V_1 can be written as,

$$V_1(\tilde{y}) = \frac{1}{n} V(y_i - Qx_i), \qquad (10)$$

where

$$Q = \bar{Y}/\bar{X}.$$

This suggests the sample estimate,

$$v_1 = \frac{1}{n(n-1)} \sum_{i=1}^{n} (y_i - qx_i)^2, \qquad (11)$$

with

$$q = \bar{y}/\bar{x}.$$

This sample estimator $v_1(\tilde{y})$ has a bias of order $O(n^{-1})$, [1, p. 119].

3. Small Sample Estimation

When the sample size is large, the approximations described in Section 2 have good accuracy, and it is not necessary to bring in the higher moments by inclusion of more terms in the expansions. With small samples however, the approximations do not always have a known accuracy and it is natural to consider improving the estimates by including some higher order terms.

By expanding Equation (2) up to fourth order terms, a second approximation to the bias of $\tilde{y}$ is obtained as,

$$\text{Bias}_2(\tilde{y}) = \text{Bias}_1(\tilde{y}) + \bar{Y}\left[-\frac{\mu_{03}}{\bar{X}^3} + \frac{\mu_{04}}{\bar{X}^4} + \frac{\mu_{12}}{\bar{Y}\bar{X}^2} - \frac{\mu_{13}}{\bar{Y}\bar{X}^3}\right] \qquad (12)$$

where $\mu_{rs} = E(\bar{y}-\bar{Y})^r(\bar{x}-\bar{X})^s$.

A second approximation to the variance can be found by expanding Equation (5) up to fourth order terms and again treating each term of Equation (4) separately. This gives the following equation as a second approximation to $V(\tilde{y})$.

$$V_2(\tilde{y}) = V_1(\tilde{y}) + \bar{Y}^2\left[-\frac{2\mu_{03}}{\bar{X}^3} + \frac{3\mu_{04}}{\bar{X}^4} - \frac{\mu_{02}^2}{\bar{X}^4} + \frac{4\mu_{12}}{\bar{Y}\bar{X}^2} - \frac{6\mu_{13}}{\bar{Y}\bar{X}^3} + \frac{2\mu_{02}\mu_{11}}{\bar{Y}\bar{X}^3} - \frac{2\mu_{21}}{\bar{Y}^2\bar{X}} + \frac{3\mu_{22}}{\bar{Y}^2\bar{X}^2} - \frac{\mu_{11}^2}{\bar{Y}^2\bar{X}^2}\right] \qquad (13)$$

In the Monte Carlo results presented in Section 4, y is generated as a linear function of x and so the consideration of some of the results when y is a linear function of x, is of interest. In this case, y = A+Bx, Equation (12) becomes

$$\text{Bias}_2(\tilde{y}) = \frac{A}{\bar{X}}\left(\frac{\mu_2}{\bar{X}} - \frac{\mu_3}{\bar{X}^2} + \frac{\mu_4}{\bar{X}^3}\right), \qquad (14)$$

where the moments of $\bar{x}$ about $\bar{X}$ are denoted with a single subscript, i.e. $\mu_{rs} = B^r\mu_{r+s}$. In terms of the moments of x_i about the mean,

$$\text{Bias}_2(\tilde{y}) = \frac{1}{n}\left[\frac{A\mu_2(x)}{\bar{X}^2}\right] + \frac{1}{n^2}\frac{A}{\bar{X}}\left[\frac{-\mu_3(x)}{\bar{X}^2} + \frac{\mu_4(x)}{n\bar{X}^3} + \frac{3(n-1)\mu_2^2(x)}{n\bar{X}^3}\right], \qquad (15)$$

where on the first term on the right is $\text{Bias}_1(\tilde{y})$ and the second term is $O(n^{-2})$.

In this case of a linear relationship between y and x, the difference between the two approximations to the variance is,

$$V_{2-1}(\tilde{y}) = V_2(\tilde{y}) - V_1(\tilde{y})$$
$$= \frac{A^2}{\bar{X}^2}\left(-\frac{2\mu_3}{\bar{X}} + \frac{3\mu_4}{\bar{X}^2} - \frac{\mu_2^2}{\bar{X}^2}\right), \qquad (16)$$

which in terms of the moments of x_i about the mean is

$$V_{2-1}(\tilde{y}) = \frac{1}{n^2}\frac{A^2}{\bar{X}^2}\left(-\frac{2\mu_3(x)}{\bar{X}} + \frac{3\mu_4(x)}{n\bar{X}^2} + \left(\frac{8n-9}{n}\right)\frac{\mu_2^2(x)}{n}\right) \qquad (17)$$

This difference is positive in symmetric populations and is also $O(n^{-2})$.

It should be noted that both the bias and the variance are direct functions of the y intercept, A, so that if $A = 0$, $\tilde{y}$ is exactly unbiased and $V_1(\tilde{y}) = V_2(\tilde{y}) = 0$. This is true for any sample size. If y is a nonlinear function of x, say quadratic, $y_i = A + Bx_i + Cx_i^2$, then the bias and variance of $\tilde{y}$ are affected not only by the y-intercept A, but also by the nonlinear coefficient, C. Then, even if $A = 0$, neither the bias nor the variance vanishes.

A higher order approximation to the mean square error can be obtained by combining the identity,

$$\left(\frac{\bar{y}}{\bar{x}} - \frac{\bar{Y}}{\bar{X}}\right)^2 = \left(\frac{d\bar{y}-Qd\bar{x}}{\bar{X}}\right)^2\left(1 + \frac{d\bar{x}}{\bar{X}}\right)^{-2},$$

with the expansion of Equation (5) and including terms up to the fourth order. This gives,

$$MSE_2(\tilde{y}) = \bar{X}^2\ E(\bar{y}/\bar{x}-Q)^2$$

$$= \bar{Y}^2\left(\frac{\mu_{20}}{\bar{Y}^2} - \frac{2\mu_{11}}{\bar{Y}\bar{X}} + \frac{\mu_{02}}{\bar{X}^2} - \frac{2\mu_{03}}{\bar{X}^3} + \frac{3\mu_{04}}{\bar{X}^4} + \frac{4\mu_{12}}{\bar{Y}\bar{X}^2} - \frac{6\mu_{13}}{\bar{Y}\bar{X}^3} - \frac{2\mu_{21}}{\bar{Y}^2\bar{X}} + \frac{3\mu_{22}}{\bar{Y}^2\bar{X}^2}\right), \qquad (18)$$

which is an expression that appears in Kish [3] and Sukhatme [5]. The difference between $V_2(\tilde{y})$ and $MSE_2(\tilde{y})$ is

$$V_2(\tilde{y}) - MSE_2(\tilde{y}) = -\frac{\bar{Y}^2}{\bar{X}^2}\left(\frac{\mu_{02}}{\bar{X}} - \frac{\mu_{11}}{\bar{Y}}\right)^2 \qquad (19)$$

which is the square of the estimate of bias obtained in Equation (3), and, since $\mu_{11} = \rho\ \sigma_{\bar{x}}\sigma_{\bar{y}}$,

$$V_2(\tilde{y}) - MSE_2(\tilde{y}) = -\bar{Y}^2 C_{\bar{x}}^2(C_{\bar{x}} - \rho C_{\bar{y}})^2 \qquad (20)$$

which is always negative and becomes negligible for large n.

Finally, when y and x are bivariate normal, $MSE_2(\tilde{y})$ reduces to,

$$V_1(\tilde{y})\cdot(1+3C\bar{x}\bar{x}) + 6\ \mathrm{Bias}_1^2(\tilde{y}). \qquad (21)$$

The sample estimate $v_2(\tilde{y})$ of $V_2(\tilde{y})$ is obtained by estimating each term of Equation (13) from the sample. Similarly, an estimate $mse(\tilde{y})$ of $MSE(\tilde{y})$ can be found from Equation (18), or, if y and x are bivariate normal, from Equation (21).

Recently, estimators which reduce bias have been constructed by splitting up samples, see Quenouille [4]. Tukey [6] later discussed the split-sample procedure from the point of view of ease in variance computations. The method consists of constructing estimates $q_{(i)}$, $i = 1,2,\ldots n$, based on all but the ith observation and then forming

$$q^*_{(i)} = nq - \overline{n-1}\ q_{(i)}, \qquad (22)$$

where q is the estimator based on all the observations. The estimator

$$q^* = \sum_{i=1}^{n} q^*_{(i)}/n, \qquad (23)$$

then has a sample variance given approximately by,

$$v(q^*) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(q^*_{(i)}-q^*)^2. \qquad (24)$$

Thus in the case of the ratio-of-means estimator

$$q_{(i)} = \frac{\sum_{j=1}^{n} y_j - y_i}{\sum_{j=1}^{n} x_j - x_i}, \qquad (25)$$

$q = \bar{y}/\bar{x}$, and an estimator of $\bar{Y}$ is given by $\tilde{y}^* = q^*\bar{X}$, with an approximate sample variance given by,

$$v(\tilde{y}^*) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(q^*_{(i)}-q^*)^2. \qquad (26)$$

4. Monte Carlo Study

To examine the problem of variance estimation for small sample sizes, random samples of size n = 2(1)9 were drawn from a population of fifty thousand pairs (y,x). The ratio $\bar{y}/\bar{x}$ was used as an estimator of $\bar{Y}/\bar{X}$ but the results can be applied directly to $\tilde{y}$.

It is a very large undertaking to draw all possible samples for a given sample size so instead one thousand samples were drawn from the population and $q = \bar{y}/\bar{x}$ computed for each. The variance of the one thousand q's should be a close approximation to the true sampling variance. The split-sample estimator q* was found as well as q, the four variance estimators $v_1(q)$, $v_2(q)$, $mse_2(q)$, and $v(q^*)$ were also calculated for each sample.

A population of fifty thousand pairs, (y_i,x_i), was constructed with x_i N(10,4), and y_i defined as $5(x_i+e_i)$ where e_i is a random error distributed as standard normal independently of x_i, y is therefore N(50,125), and the correlation between y and x is .89.

Normal probability plots were made of the one thousand sample estimates of q and q*. The first three figures show the plots of q for sample sizes 3, 6 and 9. The figure numbers in all the plots correspond to sample sizes. These plots show that even for small sample sizes the ratio $\bar{y}/\bar{x}$ is essentially normal when the original y and x populations are normal. The indications of the probability plots of q* were the same as those for q.

Table 1 gives the mean and variance of the sampling distributions of q and q* for the different sample sizes. The expected values are essentially the same, but the precision of q seems to increase over that of q* as n decreases.

TABLE 1

Mean and Variance of 1000 Samples of q and q* for Small Sample Size n

	q		q*	
n	Mean	Variance	Mean	Variance
2	4.9975	.1235	5.0000	.1481
3	4.9889	.0896	4.9887	.0915
4	4.9971	.0623	4.9967	.0622
5	4.9928	.0497	4.9930	.0507
6	5.0055	.0421	5.0053	.0433
7	4.9932	.0367	4.9933	.0366
8	5.0041	.0300	5.0043	.0323
9	5.0026	.0275	5.0026	.0288

Histograms to summarize the distributions were made for each of the four estimators of variance. The next three plots show the histograms of the first approximation $v_1(q)$ for sample sizes 3, 5 and 9. The histograms for $v_2(q)$, $mse_2(q)$ and $v(q^*)$ were similar. As the sample size increases, the distribution of the variance becomes less skewed. This behavior is similar to that of the gamma distribution for increasing values of the shape parameter.

A plotting technique similar to normal plotting has been designed for the gamma distribution [7]. This technique can be used to determine whether a random sample of observations come from a gamma distribution; if the distribution is gamma, the points will yield a straight line configuration.

Gamma plots were made of $v_1(q)$ for each sample size, and the last three plots for n = 3, 6 and 9, show that even for small sample sizes it is not unreasonable to assume that the distribution of the estimate of variance of q is chi-square.

The mean and variance of the sampling distributions of the four estimators of V(q) are given in Tables 2a and 2b.

TABLE 2a

Mean of 1000 Samples of the Estimates of V(q) for Small Sample Size: x Distribution Normal

n	$v_1(q)$	$v_2(q)$	mse(q)	$v(q^*)$	$V(\bar{y}/\bar{x})$
2	.1285	.1394	.1550	.1481	.1235
3	.0853	.0892	.0927	.0915	.0896
4	.0607	.0620	.0636	.0622	.0623
5	.0490	.0498	.0509	.0507	.0497
6	.0424	.0428	.0436	.0433	.0421
7	.0354	.0356	.0362	.0366	.0367
8	.0318	.0320	.0324	.0323	.0300
9	.0283	.0284	.0287	.0288	.0275

TABLE 2b

Variance of 1000 Samples of the Estimates of V(q) for Small Sample Size: x Distribution Normal

n	$v_1(q)$	$v_2(q)$	mse(q)	$v(q^*)$
2	.0317	.0392	.0472	.0488
3	.0072	.0080	.0088	.0096
4	.0025	.0026	.0028	.0027
5	.0014	.0014	.0015	.0017
6	.0007	.0007	.0007	.0008
7	.0004	.0004	.0004	.0005
8	.0003	.0003	.0003	.0003
9	.0002	.0002	.0002	.0002

Some improvement over $v_1(q)$ might be achieved by using $v_2(q)$ or $v(q^*)$. The mean of the one thousand $v_2(q)$'s is closer than $v_1(q)$ to the true variance for n = 3,4,5, and 7; the mean of the $v(q^*)$ is closer for n = 3,4 and 7. However, the precision of $v_1(q)$ is never less than the precision of any other estimator, and for n = 2,3 and 4, the precision of $v_1(q)$ is actually greater.

Thus, when y and x are normally distributed, $\bar{y}/\bar{x}$ and the first approximation $v_1(q)$, are good estimators of $\bar{Y}/\bar{X}$ and $V(\bar{y}/\bar{x})$ respectively. Any improvement in estimating the true variance by $v_2(q)$ or $v(q^*)$ will most likely not warrant the extra computational difficulty, in this case.

In order to examine the behavior of the variance of q for non-normal populations, a second study was made similar to the first. Here, x was chi-square with 2 degrees of freedom, and a constant added such that $E(x) = 12$, $V(x) = 4$; y was defined as before so that $E(y) = 60$, $V(y) = 125$. The correlation between y and x was again equal to .89. For this population also q was an unbiased estimator of the population ratio.

Tables 3a and 3b give the mean and variance of the four estimators of $V(q)$ for each sample size.

TABLE 3a

Mean of 1000 Samples of the Estimates of $V(q)$ for Small Sample Size n: x Distribution Exponential

n	$v_1(q)$	$v_2(q)$	mse(q)	$v(q^*)$	$V(\bar{y}/\bar{x})$
2	.1096	.1134	.1143	.0728	.0734
3	.0769	.0777	.0780	.0527	.0541
4	.0561	.0562	.0563	.0393	.0399
5	.0460	.0460	.0460	.0318	.0327
6	.0402	.0401	.0402	.0281	.0277
7	.0336	.0336	.0336	.0235	.0240
8	.0306	.0305	.0305	.0214	.0204
9	.0272	.0271	.0271	.0190	.0182

TABLE 3b

Variance of 1000 Samples of the Estimates of $V(q)$ for Small Sample Size n: x Distribution Exponential

n	$v_1(q)$	$v_2(q)$	mse(q)	$v(q^*)$
2	.0245	.0263	.0268	.0084
3	.0063	.0065	.0066	.0023
4	.0023	.0023	.0023	.0009
5	.0013	.0013	.0013	.0005
6	.0006	.0006	.0006	.0003
7	.0004	.0004	.0004	.00015
8	.0003	.0003	.0003	.0001
9	.0002	.0002	.0002	.00007

In this case the estimators $v_1(q)$, $v_2(q)$ and $mse_2(q)$ all lead to serious overestimates of the true variance. The means of these three estimators are about equal. The ratio of $v_1(q)$ to the true variance varies about 1.5 and the two estimators $v_2(q)$ and $mse_2(q)$ are actually worse than $v_1(q)$ for $n = 2,3,4$; for $n = 8$ and 9 it appears that they may begin to improve over $v_1(q)$. The precision of these three estimators is approximately the same. However, the mean of $v(q^*)$ is consistently near the true value of the variance and its precision is much greater than the other estimators.

The normal plots, gamma plots and histograms were similar in behavior to those of the first study.

The maximum observations of $v_1(q)$ and $v(q^*)$ are given in Table 4; the minimum values were approximately equal.

TABLE 4

Largest Observation From the Distributions of $v_1(q)$ and $v(q^*)$ for Small Sample Sizes: x Distribution Exponential

n	Maximum $v_1(q)$	Maximum $v(q^*)$
2	1.1249	.5282
3	.6689	.3385
4	.4256	.2000
5	.2358	.1254
6	.1772	.0922
7	.1284	.0697
8	.1128	.0670
9	.0912	.0526

Table 4 shows that the spread of $v_1(q)$ is nearly twice that of $v(q^*)$ but aside from this the distributions of the four estimators of the true variance are approximately the same, specifically, chi-square, even when the original distributions are badly skewed.

From the results of these two studies, it may be inferred that the bias of the estimator $v_1(q)$ is dependent upon the degree of skewness of the original y and x populations. Estimates of the true variance taken from higher order approximations lead only to slight improvements over the second order approximation $v_1(q)$, and in some cases the estimate is actually worse. The precision of $v(q^*)$ is nearly double that of $v_1(q)$ for exponential x distributions and the bias of $v(q^*)$ is smaller than that of $v_1(q)$. Thus it appears that the split-sample estimator q^* may be definitely preferable to q in some situations.

Computations in the Monte-Carlo study were done on the IBM 7090 computer at Bell Telephone Laboratories in Murray Hill. Plots were drawn by the Stromberg-Carlson 4020 microfilm printer using output from the 7090.

REFERENCES

[1] Cochran, W. G., Sampling Techniques, New York, John Wiley and Sons, Inc., 1953.

[2] Goodman, L. A. and Hartley, H. O., "The Precision of Unbiased Ratio-Type Estimators," Journal of the American Statistical Association, 53 (1958), 491-508.

[3] Kish, L., Namboodiri, N. J., and Pillai, R. K., "The Ratio Bias in Surveys," Journal of the American Statistical Association, 57 (1962), 863, 867.

[4] Quenouille, M. H., "Notes on Bias in Estimation," Biometrika, 43 (1956) 353-360.

[5] Sukhatme, P. V., Sample Theory of Surveys with Applications, Ames, The Iowa State College Press, 1954.

[6] Tukey, J. W., "Bias and Confidence in Not-Quite Large Samples," The Annals of Mathematical Statistics, 29 (1958), 614.

[7] Wilk, M. B., Gnanadesikan, R., and Huyett, Miss M. J., "Probability Plots for the Gamma Distribution," Technometrics, 4 (1962), 1-20.

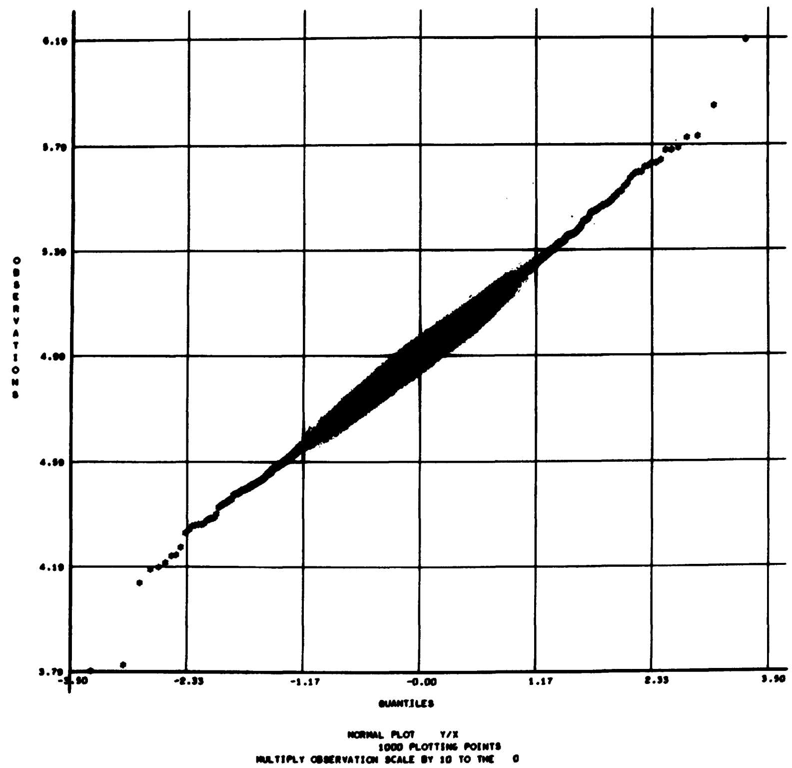


Figure 1

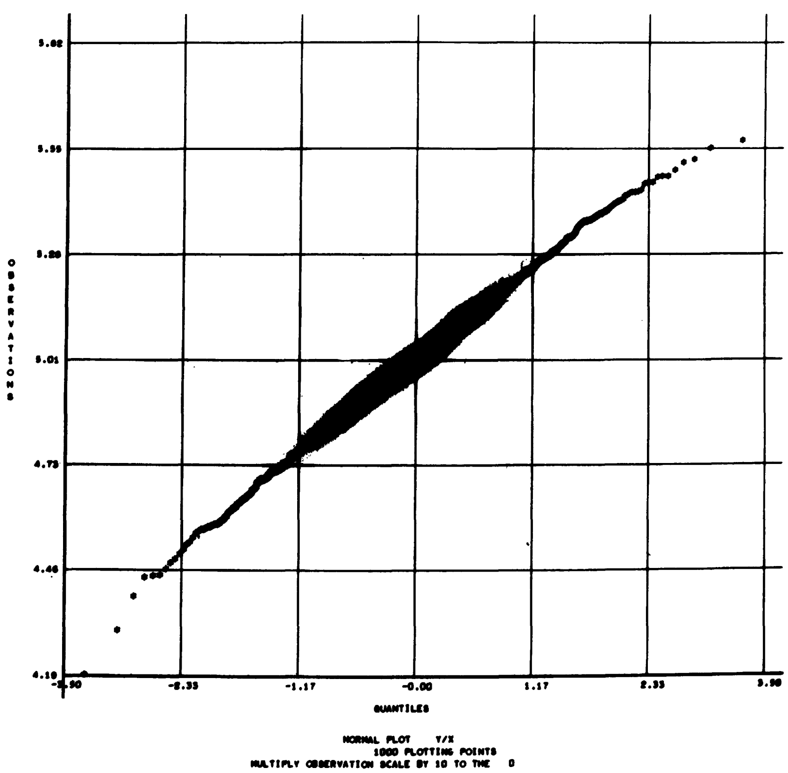


Figure 2

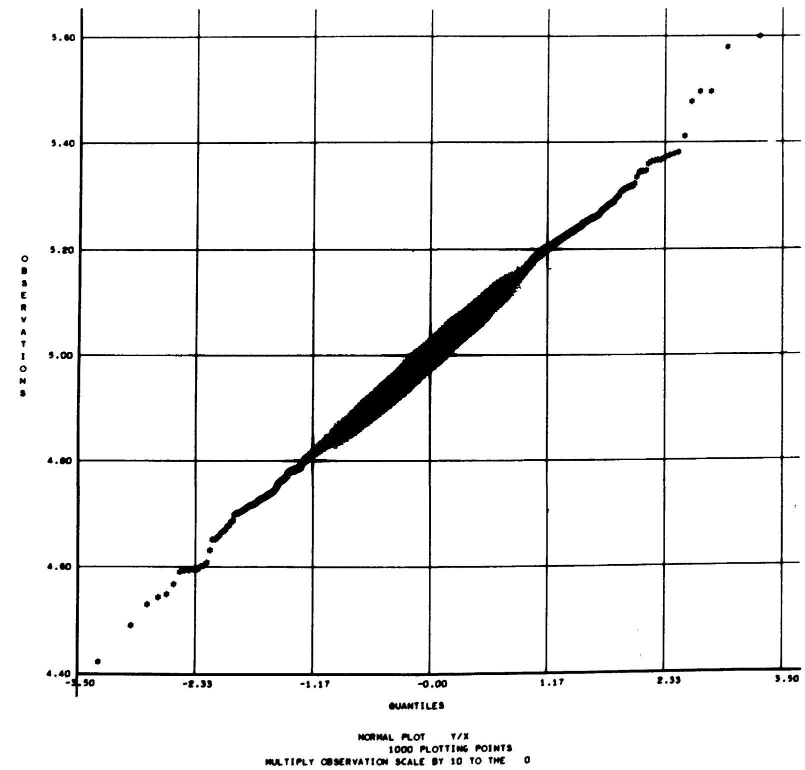


Figure 3

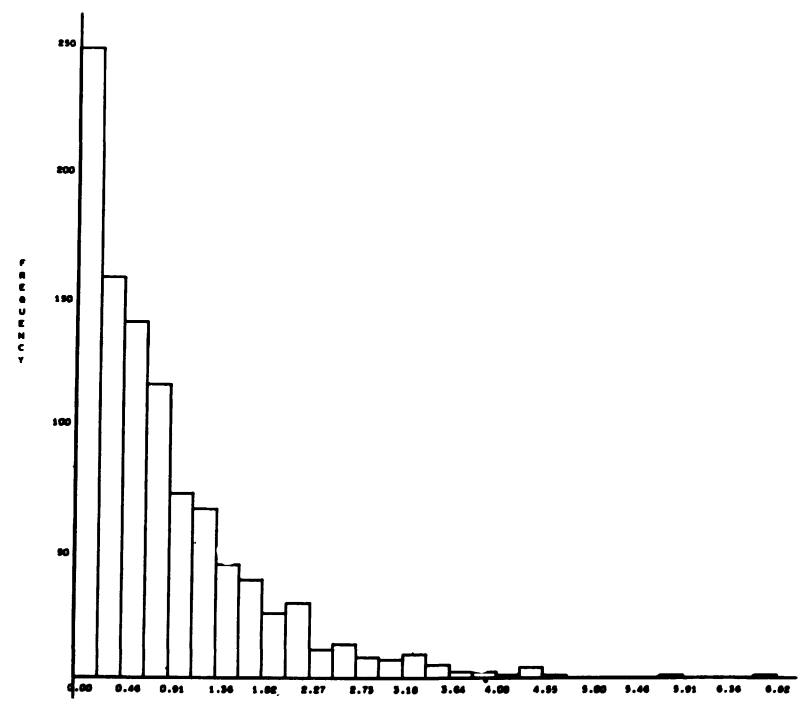


Figure 4

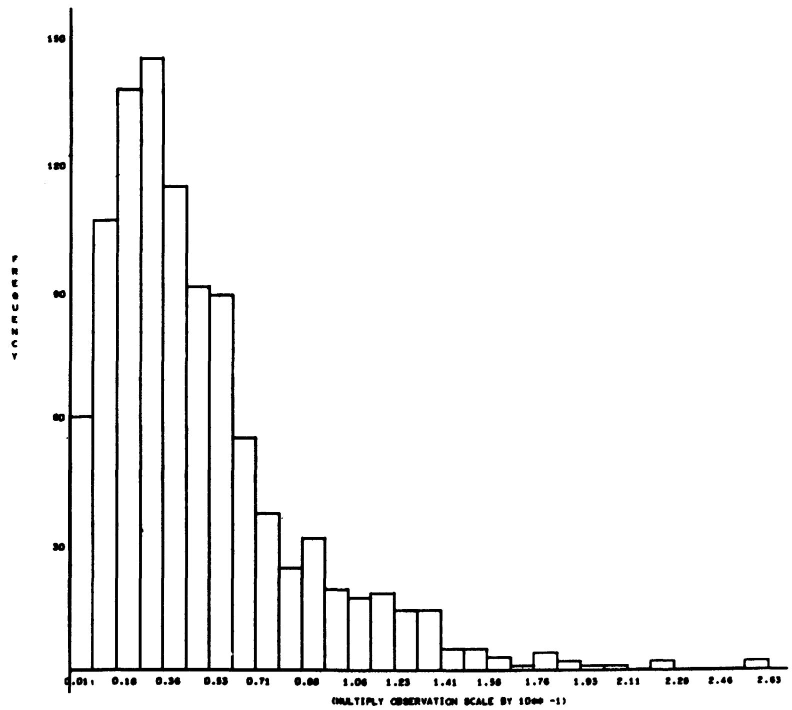


Figure 5

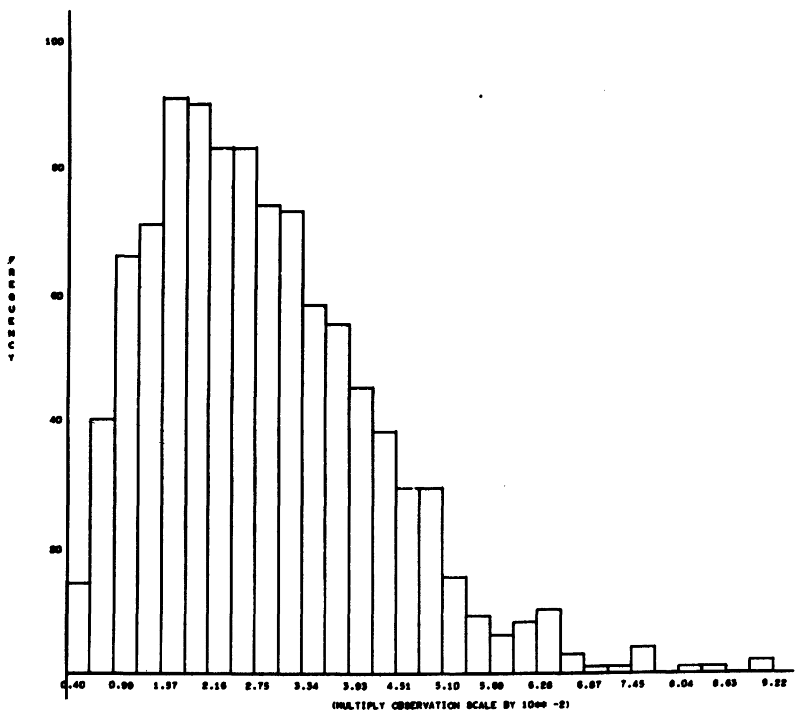


Figure 6

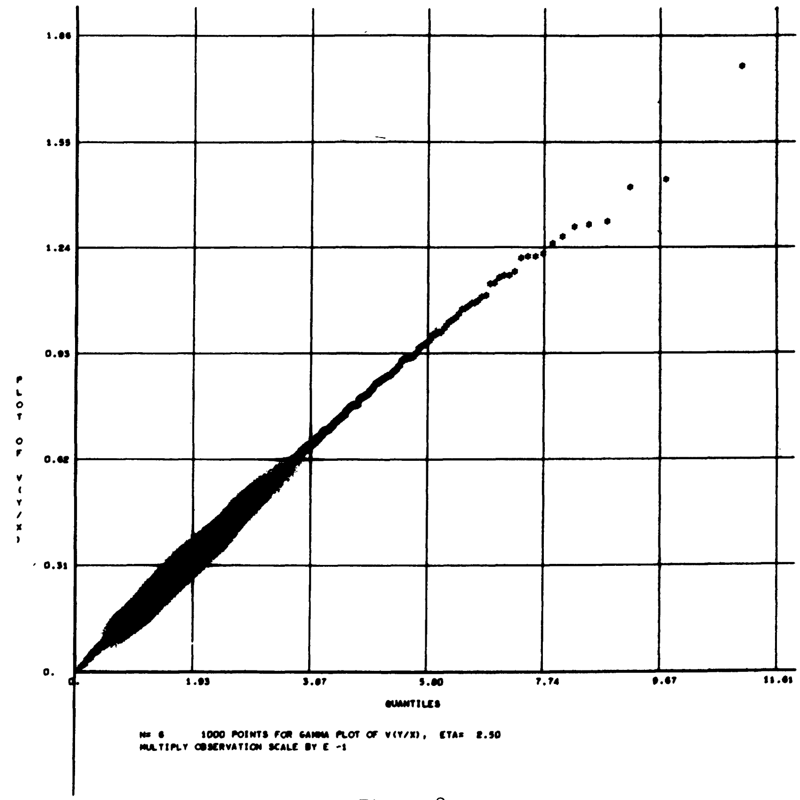


Figure 8

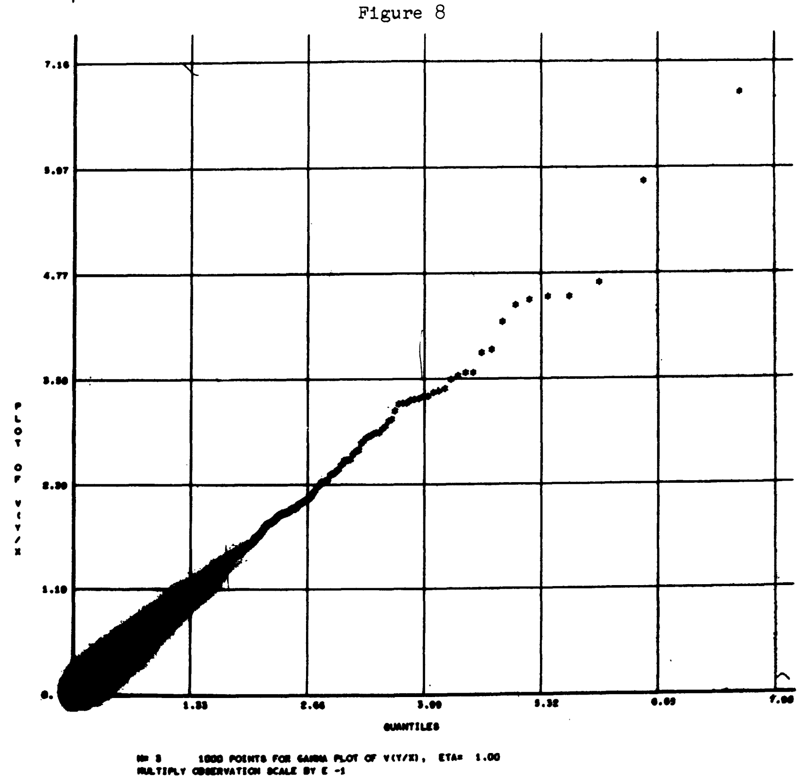


Figure 7

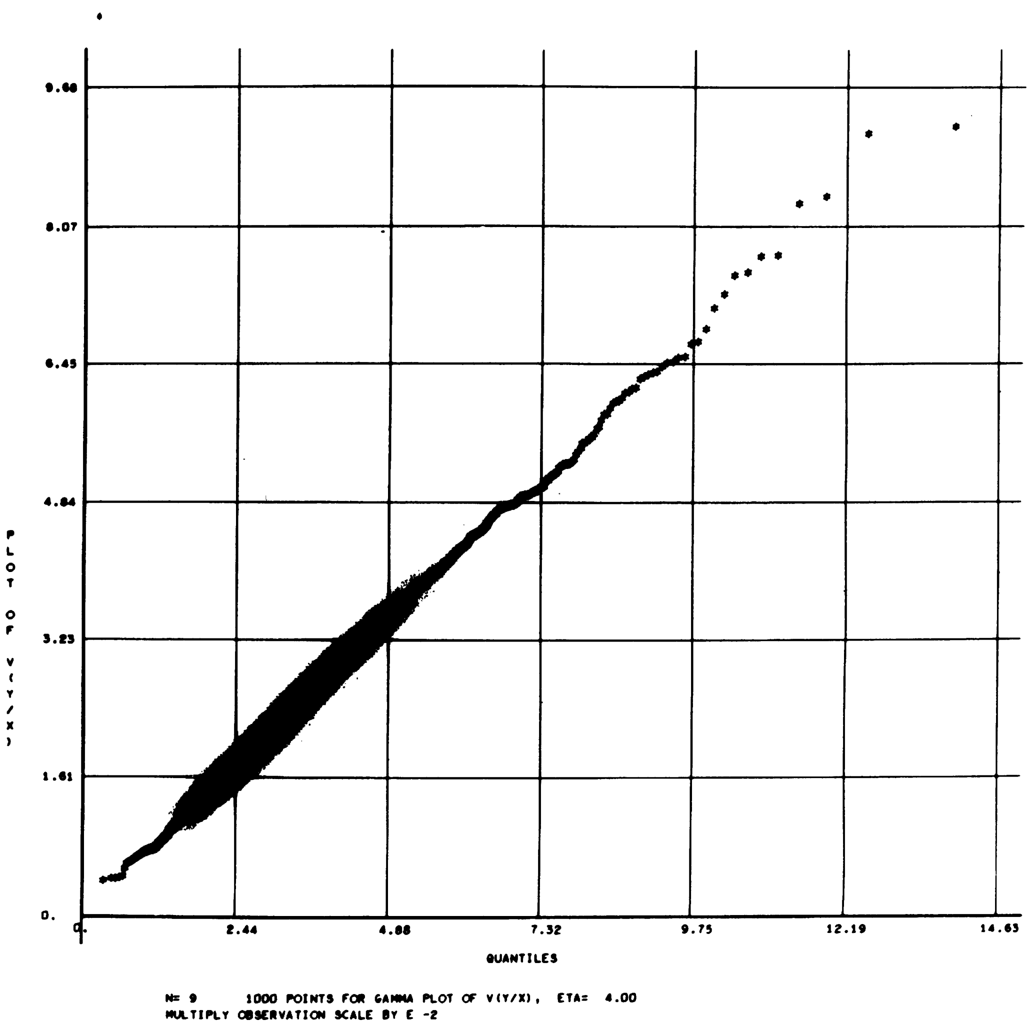

9.68
8.07
6.45
4.84
3.23
1.61
0.
PLOT OF V(Y/X)
0.
2.44
4.88
7.32
9.75
12.19
14.63
QUANTILES
N= 9 1000 POINTS FOR GAMMA PLOT OF V(Y/X), ETA= 4.00
MULTIPLY OBSERVATION SCALE BY E -2

Figure 9

IDEA ORIENTATION AND INGRATIATION IN THE INTERVIEW: A DYNAMIC MODEL OF RESPONSE BIAS*

Kurt W. Back, Duke University
Kenneth J. Gergen, Harvard University

The question of the validity of survey interviews continues to occupy an important place in social research methodology. Although a great proportion of data in social science derives from interviews, and inferences from these data are the bases of many accepted facts and theories, the legitimacy of the technique itself is still being questioned. Even if we leave out of account naive acceptance and blanket attacks, we still find contrasting models of what happens in the interview. On the one side we have the models of survey research, which think that the respondent is motivated for information-giving, and on the other of the clinically oriented ones which consider each answer to an interview an intense interpersonal manipulation. Running through these discussions is the realization that two different processes occur during the interview, only one of which produces the information which the researcher is after.

In this paper we shall start with the occurrence of both kinds of processes in the interview, and develop a model of interpersonal interaction in the interview, taking both into account. We shall then derive the conditions which further and hinder either of these processes and derive specific measures which can be built into an interview to measure the relative strength of each process. Finally, we shall illustrate through the secondary analysis of a completed survey how this kind of analysis of an interview process can proceed.

The Two-Game Model

The interview may be defined as a conversation with a purpose. Both parts of this definition are important. The interview consists of verbal communication between two persons. However, this communication is neither spontaneous nor solely determined by the interaction itself. From the sponsor's or interviewer's point of view, the conversation may have been well planned and a systematic record is the ultimate aim of the interview. In keeping with this formal setting, the roles of the two participants are kept distinct. The interviewer asks a standard set of questions in a standard manner, and attempts to maintain some type of record or to make some judgment on the basis of the information received. His satisfaction will by and large depend on the adequacy of this information. The respondent is thus regarded simply as the source of information; his satisfaction is not usually taken into account except at perhaps the opening or closing stages of the interview.

The question can be asked, however, as to what sorts of satisfactions a typical respondent may be attempting to obtain while being interrogated. At least two seem to stand out. On the one hand, the respondent may feel some degree of satisfaction from presenting his actual views, whatever they happen to be at that moment. In general, our culture is one in which personal integrity seems to be held at a premium and honesty thus provides its own rewards. On the other hand, our culture seems to be one in which affiliation is becoming increasingly important. As Riesmann has pointed out, we are becoming increasingly sensitive to the cues of approval and disapproval offered to us by others. Thus, at least some degree of satisfaction may be gained by the respondent if he can succeed in gaining the interviewer's positive regard.

We can thus construct a model of the interview by visualizing it as two games which the respondent is attempting to play simultaneously. The respondent's answers are his moves in both games. In the first game his pay-offs are determined by the degree to which he can express his own views. This game can be called the <u>information-giving</u> game. In the second, his pay-offs are derived by the impression which he feels he is making on the interviewer. The better the impression, the greater the pay-offs. This game we can call the <u>ingratiation</u> game. The weight or the relative importance of each game may be different for different respondents and different interviewing techniques. Of course, the ideal interviewer will maximize the importance of the information-giving game, encouraging the respondent to believe that maximum pay-offs can be achieved by giving complete information. This, however, is an unusual situation for any person to find himself in. Respondents are often likely to be sensitized to the interviewer's personal preferences or the implicit preferences of the sponsor of the interview.

Conditions of the Two Games

Let us now discuss some of the variables in the construction of the interview and the characteristics of respondents which may make the information-giving game more or less important.

<u>Interview characteristics</u>. Apart from the skill of the interviewer, there are two ways in which personal preferences or expectations may be shown in the course of the interview. One is simply from the cumulative content of the interview. The concentration of questions on one topic, for example, usually gives the respondent sufficient cues to determine what the real intent or purpose of the interview is. Once the core content of the questions has been established, the respondent is in a better position to play the ingratiation game. Another feature is the provision of response categories by the interviewer. In contrast to a free answer or open-end question, a closed-end question provides the respondent with cues regarding the possible range of opinion which the interviewer expects from normal respondents. It has been shown that the

mere provision of a range of possible answers is an important factor in attitude formation. This may make it easier for the respondent to follow the pattern of the interviewer's preferences or expectations as well.

We are proposing here two measures of relative weight of the ingratiation game: change of response after the direction of the interview has become clear, and change when possible response categories are provided.

Respondent characteristics. Cumulative concentration on one topic and question form may make it easier for a subject to accommodate himself to the interviewer's expectations. But what kinds of respondents may be more or less willing to play the ingratiation game? In the common interviewing situation where the interviewer has no power and no control over the respondent, we can hypothesize that the respondent will put more weight on the ingratiation game if important parts of his self-image are connected with the topic under discussion. Thus a psychiatrist may feel that he has to have definite views on mental health problems or a Catholic on contraception, especially if they are approached as a psychiatrist or as a Catholic. By contrast, the information game will have the highest pay-off for individuals who are less able and less motivated to consider personal relationships with the interviewer. This characterization fits older people because of emotional detachment, members of alienated minority groups and even men in comparison to women. Persons with these characteristics will be more likely to play the information game, although there might be other difficulties in obtaining the information from them, such as inadequate communication.

Method

The data to be presented here derive from secondary analysis of a study conducted in 1950. This study was conducted by one of the top research organizations, and therefore we may be confident that gross response errors are eliminated as much as possible. Close to four thousand interviews were collected in a random sample in one city in the Southeastern United States. The general topic of the interview was attitude toward mental health problems.

The structure of the interview made it possible to apply our measures of strength of ingratiation. In accordance with standard practice the first questions were quite general, asking about problems of the town and attitudes toward human nature. Then came a series of open-ended questions presenting instances of some unusual behavior and asking for the cause and possible cures of this kind of behavior. The situations went through the life cycle starting with an undisciplined three year old, a delinquent fifteen year old, a moody twenty year old girl, and ending with an eccentric seventy-five year old. Then the second question, the one about the delinquent, was repeated, but as a closed-end question with six alternative solutions. After this list, another situation was presented, about an extremely jealous woman, and open-end answers were sought. Immediately after the open-end question, the question was repeated and alternative answers were presented.

It can be seen that this interview provides an ideal example of our model. The concentration of 4000 interviews in one city makes it possible to have significant secondary effects of ingratiation, even though the information game itself predominates. The first open-end question of the delinquent boy is presented at a point where the general trend of the interview, a study of mental health attitudes, is still obscure. Answers to it represent reactions before the ingratiation game can be played, and may unleash spontaneous views on punishment of a delinquent. The closed-end question on the same topic is given after the trend must have been clear to most people. Change shows the combined effect of this knowledge and of question form. The pair of questions on the jealous woman given after the series shows the effect of question form only.

In both pairs of questions, the open-end answers could be roughly matched with the fixed categories and ordered along a dimension of relevance to the mental health content. For the delinquent boy the categorization was according to punitivity--reform school vs. psychiatric or social treatment; for the jealous woman a comparable dimension was the use of professional help--mental health oriented vs. legal. An intermediate category in both questions was ambiguous answers, answers which did not use either of the indicated alternatives or which used both. For each of the four questions we have three classes along the mental health dimension and a possible shift from minus two to plus two points. Of course, the scores of the answers to the delinquent boy and jealous woman questions cannot be compared with each other.

Ingratiation is indicated by a change toward the mental health orientation. In the delinquent boy questions, this can be due to the joint effect of increased familiarity with the tendency of the interview and the change in question form. In the jealous woman question, hardly any additional information on the purpose of the interview is given, and the change in this pair of questions can be considered to be a function of the form of the question only.

Results

Extent of shift. Table 1 shows that the shift from the open-end to the closed end question was indeed considerable on both issues. In agreement with the hypothesis that the first set of questions measured the joint effect of learning the purpose of the interview and of question form, and the second of question form only, the shift in the first set is stronger and also the pattern is different. In the situation of the boy there was a strong shift toward the non-punitive mental health alternative, which was chosen by four-fifths of all the respondents in the closed-end questions. In the second set, the jealous woman, the main change results from a

Table 1

Distributions of the Four Questions

	Delinquent Boy		Jealous Woman	
	Open End	Closed End	Open End	Closed End
Mental Health Treatment	22%	79%	33%	57%
Other	45	5	58	16
Punitive Treatment	33	16	9	27
n	2969		2966	

Note: For this table all coded data in the original cards were used. In the following tables, only those answers where a definite change or constancy could be established were used, resulting in a reduced number.

Source for this and following tables: Roper Comm. No. 43, 1950.

Table 2

Characteristics of Changers

Respondents	Change toward mental health	No change	Change away from mental health	N=100%
	Delinquent Boy			
Sex: Male	58%	33%	9%	1316
Female	65	30	5	1495
Race: White	63	31	6	2433
Negro	57	32	11	388
Age: 18-44	65	30	5	1682
45-64	62	30	7	819
65+	52	37	11	33
Education:				
Grade school	61	30	9	1280
High school	69	27	4	1244
College	43	53	4	284
Economic Level:				
A & B	52	40	8	203
C	64	30	6	1575
D	63	29	8	909
	Jealous Woman			
Sex: Male	29	30	40	729
Female	28	33	40	815
Race: White	29	32	39	1311
Negro	29	27	44	237
Age: 18-44	24	32	43	950
45-64	37	29	35	429
65+	28	37	35	172
Education:				
Grade school	36	29	35	703
High school	23	33	44	711
College	17	40	43	135
Economic Level:				
A & B	24	40	36	99
C	28	32	40	843
D	31	28	40	537

decrease in the intermediate reactions, dividing about evenly between the punitive and supportive professionals. The main function of the alternatives seems to have been the availability of these professionals, while the general pro-mental health attitudes were already realized in the open-end question. The difference in the two sets of questions is also shown in the amount of turnover; 43.7 per cent of the respondents stayed consistent in their attitude toward the boy, while this number increased to 54.8 per cent with the woman.

Respondent characteristics. These changes in answer show that to a certain extent a great proportion of the respondents played the ingratiation game. There are, however, differences among population groups. To indicate the absolute size of these differences we show first the percentages of shifting respondents classified according to selected population characteristics (Table 2). For the first pair of questions, the most strikingly different group is that made up of people 65 and over; 10 per cent fewer of the old respondents change toward a mental health orientation than the next younger age group. These data are made somewhat doubtful by the small number of the older group. The age trend is consistent, however, and statistically significant. Smaller differences, of the order of six percentage points, are due to sex and race. Men and Negroes tend to change their answers less. For the second set of questions, only education seems to be important, but this varies to a marked degree. Almost twice as many respondents with grade school education as with college education shift toward the advocacy of a mental health professional.

As background variables are interrelated and interact, we use multiple regression for further analysis and summary of the respondent characteristics. The procedure used was through a multiple regression program which selects the variables in order of their contribution to the multiple correlation, and stops at a predetermined cut-off point (Table 3).

Four characteristics turn out to be significantly related to the shift toward lower punitivity in the first, delinquent boy, question. In order of importance, they are: age (young), sex (female), race (white), and economic level (low). The first three characteristics are those which we have discussed before as leading to the importance of the ingratiation game. People of lower economic level may, of course, be most con-

Table 3

Rank Orders, Betas and Standard Error of Beta of Significant Respondent Characteristics for Change of Answer

Change Toward Mental Health

Delinquent Boy			Jealous Woman		
Variable	Beta	SE Beta	Variable	Beta	SE Beta
Age	-.084	.0201	Education	-.154	.0278
Sex	.054	.0201			
Race	-.042	.0204			
Economic Level	-.036	.0204			
Multiple R = .11, F = 6.56, p < .01			Multiple R = .15, F = 30.54, p < .01		

cerned with the treatment of the delinquent boy and be ambivalent about it. They are one of the most punitive groups to the open-ended question--only 16 per cent of the lowest economic group advocate the non-punitive treatment in the open-end question, as compared to 35 per cent of the upper group; but they are ready to change, given a chance.

We have postulated the first set of questions as a joint effect of learning the direction the interview is going to take and the provision of possible "mental health" answers in the closed-end questions as well. In the second set of questions, learning the rules of the ingratiation game should have been completed. Thus only lack of knowledge of the possible alternatives would make the difference between open- and closed-end questions. Indeed, we find that the only trait which relates to change in the "jealous woman" pair is education. The low-educated person, who has less readily available information on professional help, changes his answer when presented with a list of possibilities. The other traits which identified the persons wanting to learn the rules of the ingratiation game are not significantly involved in change of answer at this point.

Effect on relations of questions. The preceding analysis of the two games helps us in understanding the variables entering into the answers to the different questions themselves. We can do this by examining the relative importance contributed to the mental health aspects of the four population characteristics and in addition of one general attitude question which corresponds to an underlying dimension of the opinions expressed and so show the relative importance of idea orientation in the different questions. This was a four-step scale of faith in people, ranging from, "Most people are basically bad," to "Most people are basically good." As a measure of weight of the games we are using beta values (partial regression coefficients) and their standard errors and showing the total influence of these secondary effects--i.e., the interviewing errors--by the multiple correlation coefficients.

The first question, on the delinquent boy, in its open-ended form, can be taken as the best expression of information orientation, before the respondent has been exposed to a series of mental health questions and provided with possible alternatives. We see here that education, sex and economic level are the most important variables which affect the answer. The general underlying attitude, faith in people, is also an

Table 4

Rank Orders, Betas and F-Values of Significant Respondent Characteristics for Four Questions

Delinquent Boy: Open End			Delinquent Boy: Closed End			Jealous Woman: Open End			Jealous Woman: Closed End		
Variable	Beta	SE Beta	Variable	Beta	SE Beta	Variable	Beta	SE Beta	Variable	Beta	SE Beta
Education	.122	.0219	Sex	.144	.0194	Education	.176	.0263	Race	.055	.0229
Sex	.072	.0201	Age	-.143	.0206	Sex	.114	.0258	Education	-.050	.0229
Faith in People	.071	.0201	Race	-.101	.0198	Faith in People	.081	.0264			
Economic Level	.073	.0210	Education	.090	.0216	Race	-.063	.0258			
Race	-.052	.0200	Faith in People	.068	.0199						
Age	-.035	.0208	Economic Level	-.042	.0208						
Multiple R = .22, F = 20.89, p < .01			Multiple R = .28, F = 35.90, p < .01			Multiple R = .24, F = 21.53, p < .01			Multiple R = .07, F = 5.32, p < .05		

important variable. Better-educated women of upper economic levels are more likely to be not punitive in this situation. Race (being white) and age (being young) also lead to less punitivity but are less important.

By the time the same question was asked in the closed answer form, the shifts have occurred which we are attributing to the ingratiation game. For some people the change reinforced the previous direction: Women, younger people, and to a lesser degree whites, were less punitive to begin with and also likely to accept the point of view of the interview. We find now sex, age and race the most significant variables, while education which did not lead to the ingratiation game has slipped to a lower place, as has faith in people. Economic level, as we indicated previously, leads to conflicting responses and now is barely significant.

In the open-ended jealous woman question, sex, race, and faith in people stay important, as in the previous question. Education is of paramount importance here; as we have indicated, mainly because education is needed for knowledge about different professional help. Indeed, in the closed-end question the importance of education is reversed, lower-educated people turning more to the mental health profession, and race remaining the most important characteristic to predict the answer. Idea orientation, as represented by faith in people, has lost its influence. As we had surmised, this game loses in relative importance during the course of the interview.

Conclusion

Our model has made it possible for us to distinguish the information-giving and ingratiation processes in the results of a survey. Although the data were not collected for this purpose, we could gain an impression of the relative weight of both processes and of the effect which they had on specific relationships. With a complete experimental design woven into a questionnaire, the two games and their relative weight can be stated in a specific quantitative form and the conditions under which the respondent gains maximum pay-off can be tested. Thus it can be hoped that this model can point a way for a method to check on the interview process during the data collection.

FOOTNOTE

* This study was supported by the Office of Naval Research, Group Psychology Branch (Contract Nonr 1181(11), Project NR 177-740), a grant from the Ford Foundation on socio-economic studies of aging, and by a faculty grant from the Duke University Research Council. Some of the computations reported in this paper were carried out at the Duke University Computing Laboratory, which is supported in part by the National Science Foundation. The data were made available through the facilities of the Roper Public Opinion Research Center, Williamstown, Massachusetts. The authors thank Miss Judith Johnson and Mr. Stanley Morse for their assistance.

MINUTES OF ANNUAL MEETING OF THE SOCIAL STATISTICS SECTION

Cleveland, Ohio
September 6, 1963

1. The meeting was called to order by Leslie Kish, Acting Chairman of the Social Statistics Section. The agenda was read and an attendance list was circulated.

2. The Chairman reported that a draft of a resolution on Race and Color Statistics had been discussed by the ASA Council and referred to the Section for further discussion and revision before submission to the Board of Directors.

3. Con Taeuber presented a redraft of the resolution on Race and Color. It was moved and seconded that the Section recommend this resolution to the Board of Directors and Council. In discussion, questions were raised regarding a reference to prohibitions on the publication of statistics and on a specific reference to social security statistics.

4. Fred Stephan raised the question of what effect current changes in race relations may have on statistics and what changes in statistics might be needed. He suggested the appointment of a committee to study this question.

5. It was pointed out that the resolution wording implies that an individual can be required to divulge race. It was moved that the proposed resolution be amended to include a specific reference to social security statistics and that a clause be inserted to provide for publication of race and color data but only in the form of statistics. After discussion the amendment was adopted.

6. The motion to recommend the amended resolution to the Board of Directors was carried.

7. Stephan moved and Hauser seconded that the Section take steps to study the long range implications of the resolution on Race-Color Statistics.

8. The Acting Chairman reported on actions taken to express officially our grief at the death of our Section Chairman, Dr. Harold F. Dorn, on May 9, 1963. A letter of condolence was sent to Mrs. Dorn in the name of the Section by the Section Secretary. An obituary was prepared by Dr. Jerome Cornfield and appeared in the June 1963 issue of The American Statistician.

9. Chairman Kish announced that Jack Feldman would be the Program Chairman of the Social Statistics Section for the 1964 ASA meeting to be held in Chicago December 27-30, 1964. Suggestions for sessions at the 1964 meeting were:

Parker Mauldin - Educational Statistics

Anita Bahn - Question of a permanent "birth number" identification

Chairman Kish requested that any other suggestions be sent to Jack Feldman, National Opinion Research Center, 5720 South Woodlawn Avenue, Chicago 37, Illinois.

10. The Chairman expressed the gratitude of the Section to Parker Mauldin for his fine work as Program Chairman for the 1963 meeting and announced that Ed Goldfield would again edit the Proceedings for 1963. The Chairman urged that each program participant send a copy of his paper promptly to Ed Goldfield.

Respectfully submitted,

Eli S. Marks
Secretary

1963 Officers of the Social Statistics Section

Chairman-Elect:	Leslie Kish*
Vice-Chairmen:	W. Parker Mauldin (1962-63) Daniel O. Price (1963-64)
Secretary:	Eli S. Marks (1962-63)
Proceedings Editor:	Edwin D. Goldfield

*Serving as Chairman during 1963, following the death of Harold F. Dorn.