## George E. Hall, Bureau of the Census

In recent years there has been increased interest among social scientists in the use of longitudinal analysis to determine change and causality in various socio-economic phenomena. The longitudinal approach which takes either the form of a follow-up survey or panel survey 1/ is not new. The Census Bureau has been conducting follow-up studies for a number of years. These studies usually consist of a sample of persons whose characteristics were identified from information provided on one of the Bureau's regular panel studies such as the Current Population Survey.

It is only within recent years however, that the Bureau has become involved in longitudinal panel studies. Currently there is a survey of Medicare recipients in which the members of the panel are interviewed monthly for a period of 15 months. Primarily, the purpose of this survey is to build up estimates of medical care costs. A Retirement History Survey 2/ which will interview a panel of married couples and unmarried individuals, biennially for at least a ten-year period is now in its developmental stage. Both of these surveys are being conducted for the Social Security Administration.

The Bureau's most relevant recent experience has been gained through the National Longitudinal Surveys of labor force experience.

These surveys sponsored by the Manpower Administration of the U. S. Department of Labor, were initiated in 1966. The content and analysis are the joint responsibility of the Manpower Administration and The Center for Human Resources Research at The Ohio State University; the sample design, collection and processing of the data are the responsibility of the Bureau of the Census.

There are four separate longitudinal surveys, each with a focus on a different cohort; males 45 to 59 years of age, females 30 to 44 years, males 14 to 24 years and females 14 to 24 years. The design provides for six annual surveys of each group bounding a five-year period. The initial interviews for each cohort were staggered to permit adequate development of questionnaires, procedures, etc. Data collection for the first survey, Men 45-59 years of age, began in late May 1966; the second, Males 14-24 years, in October 1966; the third, Women 30-44 years, in the spring of 1967; and the last, Females 14-24 years, in January 1968.

The original design calling for an annual personal interview was modified on the basis of experience gained in the early surveys. For the older cohorts, it was decided to move from an annual to a biennial personal interview. A mail survey was substituted in the intervening year to provide current labor force status, and work experience in the preceding year. These changes were dictated by cost considerations and the knowledge that labor force changes for older adults would not be very rapid. Sample persons who do not respond to the mail inquiry are visited by an interviewer to complete the limited questionnaire. Because of the more rapid changes expected from the 14-24 groups, they will continue to be visited annually to obtain more current information.

The National Longitudinal Surveys utilize a multistage area probability sample located in 235 sample areas (PSU's), 3/ which represents the civilian noninstitutional population of the United States within each of the age-sex groups. One of the survey requirements was to provide reliable statistics for nonwhites in each of the cohorts. The sample was therefore stratified in such a way as to be able to increase by a factor of three the probability of selection of nonwhites. The sample was selected to provide a number of households that would yield approximately 5,500 sample persons in each of the initial surveys with approximately 1,833 nonwhites and 3,666 whites. When this requirement was examined in light of the expected number of persons in each age-sex-color group it was found that approximately 42,000 households would be required in order to find the requisite number of nonwhite males in the 45 to 59 age group.

The initial screening interview took place in March and April 1966. During this initial interview a complete household roster was completed in each occupied unit. The name of each household member, relationship to household head, marital status, date of birth, age, race and sex were obtained. These Record Cards were subjected to a clerical review and a "punch card" prepared for each 45 to 59 year old male. From these punch cards four lists of persons were constructed; white stratum, whites in the nonwhite stratum and nonwhites in the nonwhite stratum. Persons were designated for inclusion in the final sample by simple systematic selection within each list.

The original plan called for using the initial screening to select the sample for all sample groups (only three groups were contemplated at that time, females 14-24 were added later). On reflection it was decided to rescreen the sample in the fall of 1966 prior to the first interview of males 14-24. Males in the upper part of that age group are the most mobile group in the entire population and a seven-month delay between the initial screening and the first interview seemed to invite problems.

To increase efficiency, it was decided to stratify the sample for the rescreening by the presence or absence of a 14 to 24 year old male in the household. The probability is great that a household which contained a 14 to 24 year old in March will also have one in September. However we had to insure that the sample also represented persons who had moved into sample households in the intervening period, so that a sample of addresses which had no 14 to 24 year old males was also included in the screening operation. This phase of the screening began in early September 1966. Since a telephone number had been recorded for most households at the time of the initial interview, every attempt was made to complete the short screening interview by telephone. These households yielded a total of 5,789 males 14 to 24 years of age and 5,446 women 30 to 44.

The subsequent decision to add a sample of females 14 to 24 years of age again raised the question of rescreening since that survey would not go into the field until late January 1968, seventeen months after the beginning of the previous screening. However the cost of the screening operation itself and the cost of supplementing the original sampling frame with a sample of new construction made it obvious that a better allocation of resources would be to designate the individuals to be surveyed from the previously screened sample. Every effort would have to be expended to find each person selected from this source.

Response rates in both surveys of women tend to bear out the judgment not to rescreen. In the initial older woman's survey the response rate was 94.3 percent; in the younger woman's survey the rate was 94.5 percent. These compare with initial rates of 91.1 percent and 91.7 percent for older men and young men respectively.

The consequences of noninterviews in a longitudinal survey are considerably more serious than in a cross-sectional survey, because of the problem of compounding over time. Therefore, a prime consideration in this series of surveys has been the minimization of noninterviews.

A number of strategies have been adopted to achieve this end. First, it was decided to allow considerably more time in the field than in the usual Census survey. For example in the longitudinal surveys the interviewing typically is permitted to extend for a period of two months although every effort is made to complete the bulk of the work in the first four weeks. This contrasts with the seven to ten days needed to complete the monthly Current Population Survey. Although this leads to inefficiency and higher cost, the value of the completed interview overshadows the negative aspects. For example, in the survey of Females 14 to 24 interviewers were allowed to hold questionnaires from the late January start, until early April when some college students, who perhaps could not have been interviewed otherwise, would be home for spring vacation.

In all of the surveys persons who refuse to be interviewed are sent explanatory letters from a regional office in an attempt to solicit cooperation. The "Refusal" is then visited by a regular regional office supervisor who tries to obtain the interview. Only after such measures fail is a sample person designated as confirmed refusal.

When the final returns are received in Washington, the noninterview rates for each PSU are examined and the actual questionnaires containing a history of the interview attempts are examined for any PSU which seemed to have an unusually high noninterview rate. If on the basis of this examination it is felt that there is a reasonable chance to reduce the rate, the cases were returned to the field for another try. For example in the first survey of Young Men 98 questionnaires were returned to the field. This procedure has not seemed necessary in the more recent surveys.

Traditionally it is felt that one of the most serious interview problems in a longitudinal survey is the problem of finding persons who have moved. During the first detailed interview each respondent was asked for the names, addresses and phone numbers of two persons who would know where he was if he moved. For subsequent surveys this "anchor" provided a reliable means of locating movers.

Names of family and friends were not sought at the time of the screening, so that persons who moved between the screening phase and the initial interview had to be traced by other means. In general, the interviewers used great inventiveness in finding the current address of persons included in the survey. In addition to talking with neighbors and the post office, interviewers checked school records, utility company records, the police and even domestic relations courts when they had information from neighbors or relatives pointing to marital discord.

The 1967 follow-up of males 45 to 59 years of age provided the first opportunity to examine the Bureau's ability to locate people after some passage of time. Of the 5,030 cases assigned in that survey, 97 cases or 1.9 percent were not interviewed because they could not be located. Refusals on the other hand (2.1 percent) were somewhat more of a problem than the interviewers inability to locate the sample person. Only 0.4 percent were not interviewed for other reasons, such as being away, in institutions, etc. for the entire interview period. In addition to the noninterviews 60 respondents died between the first and second interview.

Because of their higher mobility, the young men's survey was thought to provide more of a test of the interviewers ability to find respondents. However there was a smaller percentage of young males lost for this reason, out of 5,234 assigned only 90 or 1.4 percent were not located. The refusal rate was also lower, down to 1.2 percent. Twenty-two respondents from this survey died between the first and second interview.

It appears that when the field operation of a longitudinal survey is reasonably administered the cost is not markedly higher than a retrospective survey of similar magnitude and complexity. Of course this can be misleading. In a retrospective survey all information about the various points in time will be collected at one time, in a longitudinal survey the same information would be collected through a series of observations. Thus for the same information the cost could be higher by as much as a factor equal to the number of observations.

Although it was found that the field costs per observation were not higher than we would expect for a similar non-longitudinal survey, the same statement cannot be made for processing costs. A number of factors tend to increase processing and programming costs. The most obvious one relates to inconsistancies which inevitably appear when data from one observation are combined with that taken from a different observation. It is also unreasonable to establish allocation routines to account for missing data, yet nonresponses to individual questions, like complete noninterviews, have a more serious impact on a longitudinal series because of the cumulative affect. Obviously information allocated at the time of one observation would probably result in an inconsistancy when compared with data from a different observation. Thus considerably more care must be lavished on the processing of longitudinal questionnaires.

A number of problems still remain to be solved in these surveys. Perhaps the largest single problem remaining involves efficient methods of handling the vast amounts of data which are becoming available for each respondent. One of the methodological problems which remains partially unresolved is the question of including persons in subsequent surveys, who were not eligible by definition for inclusion in the initial survey, for example young men who were in the Armed Forces in October 1966. The sample, while representative of the civilian noninstitutional population at the time of selection, is limited in its ability to describe the same cohort at the conclusion of the survey series. There is no difficulty with the mechanics of selecting samples of persons who were ineligible at the initial period but are eligible now. There is however a considerable cost involved and some difficulty in merging data from these persons into the general analytical framework.

1/ Bruce K. Eckland "Retrieving Mobile Cases in Longitudinal Surveys" Public Opinion Quarterly Vol. 32 No. 1 Spring 1968, pp. 52.

2/ See "A Retirement History Study" by Lola Irelan and Joseph Steinberg elsewhere in these proceedings.

3/ The basic sample design is similar to that used for the Current Population Survey which is described in Technical Paper No. 7 "The Current Population Survey, a Report on Methodology" U. S. Bureau of the Census, Washington, D. C., 1963.