HEALTH STATISTICS FROM RECORD SOURCES AND HOUSEHOLD INTERVIEWS COMPARED

By: O. K. Sagen, Ruth E. Dunham, and Walt R. Simmons
U. S. National Health Survey

1. INTRODUCTION

Analysis of the differences between ideal knowledge of a quantity and the product of a process for measuring that quantity can take many forms. Indeed the possible hazards of a statistical measurement are almost without bound. One early listing is offered in J. Edward Demings text on sampling.[1] Numerous other lists exist.

Statisticians may find it convenient to break discrepancies into four major components:

(a) Conceptual differences between the universes of the idealized quantity and of the defined measurement;
(b) bias of the measurement, i.e., the difference between expected value of the designed measurement process and the "pure" or "true" defined measurement;
(c) the average squared deviation of a designed measurement from the expected value of such measurements, i.e., the variance of the process; and
(d) the mistakes of measurement, i.e., the non-random errors of execution.

Within each of these major components, it is possible to identify sub-components, and also possible to identify areas of hazard which overlap more than one of the foregoing major components.

It is sufficient for present purposes to note that origins of error are found not just in the commonly recognized unrepresentative character of many "samples" and in inaccuracies of counting, but in every step of the statistical process, from initial birth of the notion that a measurement needs to be made, to the ultimate use which is made of the statistic produced.

General Techniques for Evaluating Data

Again a long list of approaches might be identified. Four general methods are recalled here.

(a) Global Comparisons of Two Sets of Statistics.—Certainly one of the oldest and most commonly utilized techniques is the comparison of end results from two separate procedures which in some sense are thought to be measures of the same phenomenon. The two procedures may both be sample surveys, as for instance is the case when trend of unemployment is measured both by the Monthly Report on the Labor Force, and by a 1-percent sample of claimants for unemployment insurance. Or the comparison might be between two presumptively complete enumerations: e.g., the total value of real estate in an area as indicated on the one hand by tax assessments and on the other by insurance evaluations. Or the comparison may be between results from a sample procedure and those from a census. One continuing, formalized version of this type is engaged in by the Bureau of Labor Statistics which compares annually its current estimates of employment from a sample of business establishments with totals from social security records on which taxes are paid. Adjustments of the former are made to the latter.[2] Each reader can provide other illustrations.

Except as it may be supported by subsequent further exploration, this type of evaluation is indeed but a comparison of two sources; it does not determine in itself which if either is "correct." Nor does it usually identify components of the discrepancy. It is a device for quantifying the global difference between two figures.

(b) Replication.—The traditional procedure of the physical scientist for validation of findings is repetition of the experiment. Thus a discovery by one physicist is checked by a second or third or more researchers who attempt to duplicate as nearly as possible the original procedure to test whether the original findings are reproduced. The same general process is of course common in laboratory and clinical trials in other fields.

There are many variations of this central notion. Much of the purpose of the entire technique of Design of Experiments, popularized by R. A. Fisher and employed especially throughout the biological sciences is to discover whether an observed number is truly meaningful; i.e., whether it is a valid finding. In one manner or another, the process seeks through statistical controls over selected factors to discover whether repeated observations on another factor do produce an acceptably stable characteristic.

In the social sciences, four further variations may serve as reminders of other aspects of repetitive checks. One technique, used by a number of experimenters, including especially those at the Survey Research Center of the University of Michigan and at the U. S. Census Bureau, is the procedure of reinterview of respondents in a household survey. While these reinterviews differ in purpose from one survey to another, the prime purpose usually is to attempt a repetition of the original interview by use of a more skillful interviewer and thereby to discover if the original interviewer was successful in carrying out his original assignment.

Another technique also employed widely in household sampling, is comparison of results from two or more sets of questions, each set being given to a different random sample of the same population by a fixed panel of interviewers under a fixed operational procedure. Here one seeks to discover whether the formulation of the question produces a stable result.

A third variant seeks to discover interviewer-variability through random assignment of interviewers to random subsamples of the population, and comparison of the results among interviewers. A description of some aspects of this matter is given in these Proceedings in the paper by Hansen, Hurwitz, and Bershad.

The fourth variation is one in which differences between respondents are noted by utilizing common interviewers, questions and procedures, but addressing the inquiry to different types of respondents.[3]

(c) Record-Check Studies.—This title, though coming to be widely used, is a misnomer for the class of validations which are treated here under this label. The reason for the title is made clear with two illustrations: (1) In the study described in the body of the present paper a comparison is made for each respondent between the answer given by the respondent in an interview survey and the written record of the fact about which the respondent was questioned. (2) Age as reported in the Population Census was checked against the record of birth registrations for a sample of cases in the 1950 Census.[4] The idea behind this class of validation is that there exists some different superior type of measure to the one being utilized in the main survey, and that this superior measure, through a case-by-case comparison, will provide an estimate of the bias of the main survey measure, and in addition may indicate the cause of error in the latter. "Record check" was termed a misnomer because in some instances of applying the same general process, the main survey measure is itself a written record, while the presumed superior standard may be a different sort of measure.[5] Normally, it would be supposed that the "superior" measure was unavailable on a comprehensive basis—for cost or other reasons—or it would be used as a substitute for the main survey measurement.

(d) Internal Consistency.—In some types of investigations, data from the survey itself afford an opportunity for an internal consistency check. For example, in the BLS Consumer Expenditure Studies it is necessary that revenue of all types approximately balance with outlays, if the statistics are to be valid. In housing statistics, the census of dwelling units at one point in time augmented by new construction, and adjusted for conversions and demolitions, should equal approximately the census at a later date, definitional changes being taken into consideration.

Commentary

Thus, a considerable and growing number of methods are available for assessing gross error and the many components of statistical error. Progress on a very difficult front is being made. But the authors state their belief that much needs to be done, both in undertaking and reporting specific evaluation projects, and in synthesizing error analysis.

Most schemes known to us suffer from either one or both the difficulties of unsatisfactory theoretical formulation and operational execution.

The theoretical difficulties are of many types, with perhaps two being especially prominent: (1) the task of establishing conceptual identity between a measurement and the phenomenon measured, and (2) the building of models which disentangle confounding components of errors in a tractable manner.

Operational difficulties also can be severe. They vary with the type of validation. As examples, note three illustrations:

(1) In a large scale survey the administrative burden of randomizing interviewers over geography is heavy, and the process is costly.

(2) In a record-check study, it is not always easy to match individual cases; further, it may not be clear that the "superior standard" is in fact the better measure.

(3) In replication procedures which involve human respondents and frequently human interviewers, can one be assured that conditioning of respondent or other agent has not taken place?

2. MEASUREMENT PRINCIPLES IN A HEALTH SURVEY

Consider the matter of how many persons in the United States are sick. There is first the idealized concept of this quantity, which in some sense is the number of persons who are unwell or unhealthy. Then there is a defined measurement which may be applied to each person. The measurement process may consist of obtaining answers on interviews to a specified set of questions, including perhaps such inquiries as : (1) were you sick at any time last week or the week before? (2) Last week or the week before did you have any accidents or injuries, or suffer from the effects of a previous injury? (3) Last week or the week before did you take any medicine or treatment for any condition, or see a doctor concerning any condition, or cut down on your usual activities because of some health condition? A positive answer to any of the questions may be used to classify the person as ill.

There clearly are possible differences between this defined measurement and the idealized concept of illness. For example, the person might have glaucoma and be completely unaware of it.

Focus attention next on a survey designed to produce the intended measurement. This survey, let us say, is on an area-type, household multi-stage stratified probability sample design, in which the answers to questions are obtained through personal interview. This survey process and its associated estimation procedure has an expected value which may differ from both the idealized concept and the defined measurement, for a variety of reasons including elements from sampling, training, respondent conditioning, the classification of replies, non-response, the formal estimation process, and many others. These same sources of potential error are also responsible for the variance of the measurement process from one application to another, as different samples are drawn, different interviewers selected, different temporal factors introduced. Finally, a gross error in execution can contribute to differences among the possible measurements: such phenomena as an interviewer who cheats; a storm which prevents interviewing in an entire sector of the nation; a faulty circuit in the electronic computer which edits the data.

To carry the illustration further, turn again to the area of uncertainty which arises from the interviewer. One needs to be careful in labeling interviewer deviations. Quite different interpretations are possible depending on what is accepted as the universe of interviewer-recordable replies. There is, for example, a within-interviewer variance which reflects variation in treatment of the same question with different respondents or at different times by a single interviewer. That same interviewer may have a bias which is defined as the difference between his expected product and that of some group of interviewers of which he is a member. But this same difference becomes a part of the between-interviewer variance when viewed at another level of analysis. In turn, the expected value of the product of all interviewers in the groups combined may be a bias or a component of variance depending on the universe designated as frame of reference.

In dealing with these problems of measurement, we face not only a great variety of types of discrepancy but also many possible dimensions of analysis for evaluating them.

The U. S. National Health Survey is making a number of attempts to better understand and evaluate the data obtained from its household interview survey on health conditions. One of these is the study identified as the H.I.P. record check, which seeks to explore an area which, in essence, is the intersection of two fundamental theses. The first thesis is that the concept of illness inherent in a written medical record created in the course of rendering medical services differs from the concept of illness contained in a household respondent's replies to a specific line of questioning by a survey interviewer. Certainly the interview questionnaire on health conditions and the medical record would be expected to produce some different statements as to health or illness status of the individual concerned. In contrast, the second thesis is that the household interview report and the medical record have some elements in common and that the information from one source is correlated with that from the other source. At least, for certain classes of individuals and conditions, the two information sources ought to furnish measures of the prevalence of illness which vary together from class to class. In summary, interview and medical record information ought to be different but still similar to each other in important respects.

3. THE 1958 H.I.P. RECORD CHECK ON THE NHS HOUSEHOLD INTERVIEW

A stratified sample of families enrolled in the Health Insurance Plan of Greater New York (H.I.P.)[6] was interviewed during May and June, 1958 to compare the illness conditions and hospitalization experience reported on interview against the comparable information on the records of H.I.P. Since H.I.P. enrollees do not necessarily obtain all their medical care from H.I.P. even though they are entitled to full care, this study is largely confined to a one-way comparison, namely, how much of what appears on the H.I.P. record gets reported on interview. However, an upper limit on the extent of overreporting can be inferred from the data on interview-reported conditions which were not on record with H.I.P.

There was a conscious and strong attempt to duplicate the conditions present in a regular National Health Survey household interview so as to make the study representative of the ordinary interview process, an impossible task, to be sure. Interviewing was spread out over a two-month period, interviewers had the same instructions and the same questions for the items under study, and interviewing was conducted by the Bureau of the Census under the supervision of the regular Census Bureau supervisor for the New York City area. Furthermore, this supervisor conducted reinterviews on a sample of the interviewed households just as in the regular survey.

Despite all this, there are a number of specific differences between the interviewing for the record-check study and the interviewing in the regular survey. First, in our study the interviewers were given a surname and a specific address whereas in the regular survey the interviewers approach six dwelling units in a cluster of addresses without any surnames. This fact, together with others, makes it impossible to conceal from the interviewer that the study interviews are something special instead of routine. Second, certain questions were deleted from the regular questionnaire and others added for the study. For example, the study needed to obtain signed consent for examination of the family's medical records and it was also necessary to identify the physician who rendered service for

a medically attended condition; third, 12 interviewers were used in the study whereas only three interviewers normally worked on the regular survey in the New York City area. This circumstance made it necessary to recruit and train a small corps of new interviewers for the study. Also, the study sample was stratified to yield more reports of chronic illness than appear in the regular survey.

Even if the ordinary conditions of interviewing could be completely and faithfully reproduced for this type of study, there are several basic problems which again can at best be handled imperfectly. One of the difficult problems rests on how to set up a reliable correspondence between the terminology used by physicians for medical records and the descriptions used by the whole gamut of respondents to a household interview. For example, when the medical record shows "duodenal ulcer and hemorrhoids" while the respondent mentions only "stomach trouble" is there any, some, or no correspondence between the two?

When the medical diagnosis is "club foot" and the respondent reports "flat feet" how should we handle the report? This indicates a need for very broad categories of diagnosis for classifying the diagnostic statements on the medical record coupled with a scheme for grading the degree of correspondence. The man with the club foot at least reported the affected part of the body correctly and no one quarrels with the view that a person with an ulcer has stomach trouble. Accordingly, for this study three degrees of match were used to classify household interview reported conditions which had a possible counterpart in the medical record.

The problem of classifying responses is also complicated by the ground rules for interviewing.[7] The interviewer asks about (a) illness, accidents, injuries, other conditions and medical service in the past two weeks; (b) ailments and conditions that have been bothering one for a long time; (c) a check list of 26 conditions possibly present sometimes during the previous 12 months; and (d) a check list of 9 impairments. A chronic condition is one which is elicited by this procedure and which either appears on one of the check lists or had its onset more than 3 months before interview.

It is easy to assume that the medical record is the perfect source of reliable information on the medically treated conditions which were encountered by the population under study. However, as in any record system we must face the possibility of errors in the criterion document. There are the errors of omission arising from the failure to record pertinent information on the medical record. More serious for this type of study are the errors committed by mis-recording a diagnosis, recording a tentative diagnosis, and mistakenly assigning one patient's ills to another.

The criterion record used in this study is the H.I.P. "Med-10" report, which is not a medical record in the usual sense. The Med-10 is a line entry report for each patient seen by the doctor during a single day; it is designed for statistical and administrative purposes only. For each visit, the doctor identifies the patient, the place at which he was seen (office, home, or hospital), the nature of the service rendered (preventive, obstetric, surgical, and other) and an abbreviated diagnosis.[6] Various studies in H.I.P. have shown that the Med-10 is a reliable record and, for statistical purposes, is both more reliable and convenient than the complete clinical record.

In order to more fully investigate factors which might be connected with interview reporting of medical information and to check up on the validity of the Med-10 record as a criterion record, a physician interview phase of the study was designed. Herein, the H.I.P. physicians who rendered care to the subjects of the household interview were questioned about the case of each such patient. Because from 7 to 8 months necessarily elapsed between the household interview with the enrollee and the interview with his physician it was not feasible to again go back to the household respondent and question him further about discrepancies. Also, H.I.P. policy considerations prevented a check back to the respondent after having compared his interview response with the Med-10 record.

Full analysis of the study is far from completed. Furthermore, it should be borne in mind that a single study of this type cannot be definitive, and can at best be indicative. Nevertheless, the broad outlines of some preliminary findings can be described at this time.

4. THE INITIAL HOUSEHOLD INTERVIEWS AND RECORD CHECK

The persons for study were drawn from the subscribers and their dependents enrolled in H.I.P. on June 30, 1957, a total of 513,000 enrollees. Interviews were conducted on about 5,000 of these enrollees in 1,413 households. Of these households 959 were in the "chronic" stratum (i.e., at least one member had service for a chronic disease after June 30, 1957) and 454 households were in the "non-chronic" stratum. In the population represented by this sample, an estimated 44 percent had at least one chronic condition on a Med-10 record after June 30, 1957.

Table 1 summarizes the principal components of the interview which were checked against the Med-10 records. It is noteworthy, although not surprising, that the experiences which are least difficult to describe, viz., hospitalizations and recent acute illness or injury, are far more fully reported than are the chronic conditions.

The percent of conditions reported varies considerably with the kind of illness. Among the acute conditions, respiratory ailments are 73 percent reported while acute eye and ear conditions are reported only 40 percent of the time. Among

Table 1. Correspondence between experience on record and household interview reports by type of respondent for 6,609 persons (weighted frequencies)

Class of experience	All respondents			Self-respondents[a]			Proxy respondents[a]		
	Med-10[b] record	HHI[c] report	Percent	Med-10[b] record	HHI[c] report	Percent	Med-10[b] record	HHI[c] report	Percent
Non-chronic conditions	201	127	63	58	35	60	143	92	64
All chronic conditions	4648	1481	32	2222	791	36	2426	690	28
Class 1--Check list, unqualified	1872	826	44	878	419	48	994	407	41
Class 2--Check list, qualified	1231	340	28	605	216	36	626	124	20
Class 3--Non-check list	1545	315	20	739	156	21	806	159	20
Hospitalization episodes	350	306	87	205	180	88	145	126	87
Days hospitalized									
All episodes	2634	2397	91	1533	1438	94	1101	959	87
Unreported episodes	258	0	0	133	0	0	125	0	0
Reported episodes	2376	2397	101	1400	1438	103	976	959	98

[a]When the respondent and the subject of the interview are the same person, the case is called self-respondent; if the respondent is a responsible adult member of the subject's family, the case is called proxy-respondent.

[b]Refers to data from Med-10 except for "days hospitalized." Hospital stay was obtained directly from the hospital record because the Med-10 only records the days on which the doctor saw his patient in the hospital. Only stays of overnight or longer were counted.

[c]HHI = Reported in Household Interview.

the chronic conditions asthma and hay fever are 76 percent reported while just one-third of the cancer is reported and only 26 percent of mental illness is reported. In general, however, the more medical service is given for a chronic condition the better it is reported. All chronic conditions receiving only one service were 20 percent reported while those with ten or more services were 80 percent reported. None of the cross classifications lead to any surprising inferences but at the same time there is no simple classification scheme which lends itself to a convenient definition of items that are best reported on interview.

The reinterview phase of the study did not contribute enough to better reporting to be taken into consideration here. While some additional conditions were elicited on reinterview, there was not enough to appreciably affect the completeness of reporting when that information was added to the information from original interviews.

5. THE PHYSICIAN INTERVIEW

The individuals about whose illnesses the physicians were interviewed were a sample of 341 adults, drawn randomly from the two strata of people in the record-check study and who met the following criteria: (1) were self-respondents in the household interview; (2) gave written permission during the household interview for review of medical records; and (3) received service from H.I.P., according to Med-10 data, during the study year for one or more "queriable" conditions.

These conditions, termed queriable, were chosen for purposes of clarity of results and ease of interviewing. They were limited to certain conditions defined in the National Health Survey as chronic conditions or as impairments. Excluded were acute conditions (such as pneumonia), certain minor chronic conditions (such as fungus infection of the feet) for which an interview would be pointless or awkward for the participants and produce little useful information, and diagnoses indicated as tentative on the Med-10.

Personal interview of the attending physician was chosen as the means of obtaining the information sought. This choice was made on the basis that personal interview (1) would result in greater response than would other methods, e.g., mail queries; and (2) would be likely to result in better information of the kind needed. This latter point was related closely to the choice of interviewer personnel.

Physicians were decided on as interviewers because it was believed that they could be expected to gain entree to the attending physicians

readily, could establish a good relationship with them quickly, and could obtain the needed information in the shortest time. Six Public Health Service medical officers were specifically trained for this purpose and conducted the physician interviews.

The interview used a 4-page questionnaire, part of which was quite rigidly structured to provide categorical answers and part of which allowed free responses. Each questionnaire dealt with one patient and provided space for comments by the physician on three conditions. In some instances, a physician was interviewed regarding several patients during one appointment, and for a number of patients more than one physician had to be interviewed for services rendered the same person. The physician interviews took place during 3½ weeks in January and February 1959. Two hundred and eighty-four H.I.P. physicians in 30 of the 32 H.I.P. medical groups were seen.

6. RESULTS FROM THE PHYSICIAN INTERVIEW

There were 341 persons whose chronic illnesses and impairments were investigated in the physician interview. About nine percent of the chronic conditions recorded on the Med-10 represented only tentative diagnoses and another four percent of the diagnoses had been ruled out and the patient so informed before interview. Although the tentative and ruled out diagnoses were not considered appropriate as items to check against interview reports, nonetheless 30 percent of these diagnoses did get reported on interview.

In 97 percent of the cases with a definite diagnosis, the patient had been informed of the diagnosis in some fashion, accounting for 452 conditions. If the medical terminology had been given the patient, 63 percent of the conditions were reported in the household interview, whereas if the

Table 2. Influence of complaints, medication, interference, and distress, as determined in physician interview on reporting of chronic conditions by class of condition (weighted frequencies)

	Total conditions		Class 1 Checklist conditions unqualified[a]		Class 2 Checklist conditions with qualifications[a]		Class 3 conditions not on checklist	
	Number of conditions	Percent reported	Number of conditions	Percent reported	Number of conditions	Percent reported	Number of conditions	Percent reported
Total conditions	452	53	275	62	63	56	114	32
Conditions with:								
Complaints or symptoms	359	58	213	66	56	55	90	40
Medication or regimen	356	59	219	68	53	57	84	37
Interference in usual way of life	186	69	108	77	34	68	44	52
Physical or emotional distress	296	62	172	70	49	78	75	47
Conditions without:								
Complaints or symptoms	91	35	60	47	7	57	24	0
Medication or regimen	96	31	56	36	10	50	30	17
Interference in usual way of life	264	42	166	51	28	43	70	19
Physical or emotional distress	151	37	100	48	14	50	37	3

[a] A check-list condition is termed unqualified, if it is listed simply and without limiting restrictions; e.g. "Heart Trouble", a condition listed with any limiting description is called qualified; e.g. CHRONIC bronchitis.

Table 3. Completeness of reporting chronic conditions among cases investigated in the physician interview according to how many times medical service was received in the 12 months preceding interview (weighted frequencies)

	All cases		Cases with some botheration[a]	
	Number	Percent reported	Number	Percent reported
Total conditions	452	53	156	72
Medical service given:				
Once	137	36	31	65
2-4 times	164	46	44	55
5-9 times	73	64	37	76
10 or more times	78	87	44	93

[a]Complaints or symptoms, medication or special regimen, interference with usual way of life, and physical or emotional distress.

condition had been described in some other way only 46 percent were reported.

An important objective of the physician interview aspect of the study was to see if certain characteristics of a chronic illness might have a bearing on how well the condition was reported. The physicians accordingly were asked if the condition:

(1) caused the patient any complaints or symptoms, or
(2) required any medication or special regimen, or
(3) might have interfered with the patient's usual way of life, or
(4) might have caused the patient physical or emotional distress.

The findings are summarized in Table 2. Each of these four characteristics which might accompany a chronic condition produced about the same pattern of response for the three classes of conditions, a situation which does not obtain in the absence of these characteristics.

A little different axis of trouble associated with a chronic condition is the volume of medical services required by the condition. Table 3 demonstrates that the most troublesome conditions get reported best.

The physician interview study could not help greatly in interpreting factors in the reporting of specific diagnoses because of the small frequencies For example, it would not be correct to say that the study demonstrated more nearly complete reporting for diabetes than for hemorrhoids, although the data do suggest this and it is not an unreasonable finding. However, when we confine attention to the cases confirmed by the physician as definite chronic disease, some items of interest were very well reported. For example, asthma and hay fever were 88 percent reported, sinusitis 73 percent, diabetes 93 percent, ulcers 80 percent, heart disease 70 percent, and arthritis and rheumatism 61 percent.

7. DISCUSSION

The overriding conclusion to be drawn from the preliminary results of the H.I.P. record-check study is that the prevalence of chronic illness obtained from household interviews must essentially rest on an operational definition determined by the interview procedure itself.

That is to say, the prevalence of heart disease as estimated from the survey is not its prevalence in the clinical sense of one disease but the frequency of positive answers to a specific line of questioning, carried out in the survey context.

Comparison of the household interview data with H.I.P. records shows definite and substantial underreporting of chronic disease for which medical service was rendered. However, the evidence is just as clear that substantial correlation exists between the two sources and that the correlation is greatest for the conditions which combine a large volume of medical service with considerable travail and inconvenience to the patient. The complexity and multiplicity of factors affecting correspondence between the two sources makes it a formidable task to describe their precise relationship. Nevertheless, preliminary results justify the hope of satisfactorily describing classes of conditions for which correspondence is high, others for which it is lower, and so on. A table, which arrays conditions along an axis graded by degree of importance or severity, showing percent

of correspondence, definitely establishes the basis for this hope.

Percent correspondence between household interview and medical record for definitely diagnosed chronic conditions of which the patient had been informed by his physician

Characteristics	Percent correspondence
1. No symptoms, distress, medication or interference with usual activity	15 percent
2. Symptoms only	19 percent
3. Symptoms and distress only	41 percent
4. Symptoms, distress, and medication only	52 percent
5. Symptoms, distress, medication, and interference	72 percent
6. All four characteristics together with 10 or more medical services for the condition	93 percent

While this paper does not take up the question of overreporting, viz., the reporting of conditions in the household interview which are not reflected in medical records, the full study does show that such possible overreporting fails by far to compensate for the underreporting just discussed. It is fairly evident that any diagostic information which an individual is able to report about his illness is dependent on his having had medical care for the ailment and to a large extent on his being given the medical terminology for the condition by the physician. Otherwise, all the individual can do is to report complaints which may or may not be significantly symptomatic of a diagnosable disease.

Several investigators have discounted the household interview survey technique for its inability to measure the prevalence of chronic disease and they advocate surveys using clinical examinations as yielding more information and a truer picture.[8-12] At the same time this has tended to obscure the fact that the data obtainable from clinical examination also rest on an operational definition of its own which is inherent in the type of examination conducted, the length of time over which the individual is observed, and the point in time at which the examinations are concluded.

The situation with respect to acute illness and hospitalization is somewhat different. Both types of experience are much more clear cut. Limitation of the time reference to the two weeks preceding interview for acute conditions not only holds down memory loss but makes the items inquired about much more specific and understandable to the respondent. A similar time restriction for chronic disease would tend to ignore the main feature of much chronic illness, i.e., that episodes of acute distress come and go and that once the individual has adjusted to his chronic difficulty he tends to consider the situation normal for him.

In the current H.I.P. study it was found that known hospital episodes were 13 percent underreported in the household interviews while the aggregate days of overnight or longer hospital stays reported in household interviews were 8 percent less than on record at the hospital. In San Jose, California, Belloc[13] observed, respectively, 13 percent and 7 percent underreporting of the two items. In addition, preliminary results from another current NHS study* yield almost identical figures on underreporting. All three of these studies used a one year recall period for the hospitalization questions and confined themselves to hospital stays of overnight or longer. Belloc found that, among San Jose residents, hospital episodes were 11 percent overreported, thereby very nearly compensating for the underreporting. However, neither the H.I.P. nor SRC studies find overreporting to anywhere near this extent.

8. SUMMARY

A discussion of concepts and principles involved in using record sources as a standard of comparison for reports from survey interviews has been presented to evoke interest in further research and study of this aspect of social survey techniques. The problems of concept and design for a study of health interviews compared against medical records have been described and some preliminary results from the study have been presented. Even these rather fragmentary findings leave little room for doubt that in an interview survey of individuals with respect to their health, strictly operational definitions apply to the data on the prevalence of chronic conditions by diagnostic categories. Correspondence between data from household interviews and medical service records varies with the degree of travail and inconvenience to the patient and the volume of medical service entailed by the condition.

Comparisons with alternative sources of data in the same general area can illuminate the operational definitions but cannot replace them with more nearly absolute concepts.

Data on acute illness and on hospitalization present less severe problems and show closer correspondence between reports from household interview and medical record of service than is the case for chronic disease.

*A contract study by the Survey Research Center, University of Michigan utilizing a sample of discharges from 21 hospitals scattered over the United States. The households of the discharges were interviewed in 1959 by the Bureau of the Census according to the regular NHS specifications.

REFERENCES

1. Deming, William Edwards. Some Theory of Sampling, New York, John Wiley & Sons, Inc., 1950.

2. Simmons, Walt R. "Measurement of Industrial and Commercial Employment in the United States," The American Statistician, Vol. 10, No. 4, October 1956.

3. Nisselson, H. and Woolsey, T.D. "Some Problems of the Interview Design for the National Health Survey," Journal of the American Statistical Association, Vol. 54 (1959), pp. 69-86.

4. Marks, Eli S., Mauldin, W. Parker, and Nisselson, Harold. The Post-Enumeration Survey of the 1950 Census; a Care History in Survey Design, Journal of the American Statistical Association, Vol. 48, June 1953.

5. Simmons, Walt. R., "Concepts and Response in Statistics on Business Establishments." Study described at the annual meeting of the A.S.A. in Detroit, September 1956.

6. Division of Research and Statistics. HIP Statistical Report - 1958. Health Insurance Plan of Greater New York, 625 Madison Ave., New York, N.Y., November 1959.

7. U.S. National Health Survey. Concepts and Definitions in the Health Household-Interview Survey. Health Statistics. Series A-3. Public Health Service Publication No. 584-A3. Public Health Service, Department of Health, Education, and Welfare, Washington, D.C., September 1958.

8. Elinson, Jack, and Trussell, Ray E., "Some Factors Relating to Degree of Correspondence for Diagnostic Information as Obtained by Household Interview and Clinical Examinations," American Journal of Public Health, Vol. 47 (1957), No. 3, pp. 311-321.

9. Krueger, Dean E., "Measurement of Prevalence of Chronic Disease by Household Interviews and Clinical Evaluations," American Journal of Public Health, Vol. 47, No. 8 (1957), pp. 953-956.

10. Commission on Chronic Illness, Chronic Illness in the United States Vol. III: Chronic Illness in a Rural Area, The Hunterdon Study. Harvard University Press, Cambridge, 1959.

11. Commission on Chronic Illness, Chronic Illness in the United States Vol. IV: Chronic Illness in a Large City, The Baltimore Study. Harvard University Press, Cambridge, 1957.

12. Eichorn, R.L., and Morris, W.H.M., "Respondent Errors in Reporting Cardiac Conditions on Questionnaires," Proceedings of the Purdue Farm Cardiac Seminar, Purdue University Agricultural Experiment Station, Lafayette, 1959, pp. 46-48.

13. Belloc, Nedra B., "Validation of Morbidity Survey Data by Comparison with Hospital Records," Journal American Statistical Association. Vol. 49 (1954), pp. 832-846.