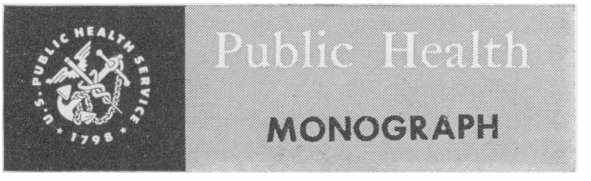
# *Sampling Plan for a Small Household Survey*

Whether older persons would enter the labor force, if they could, is a question with many complications. In a study conducted in Hagerstown, Md., the Division of Public Health Methods of the Public Health Service attempted to develop an interview to determine the degree of interest in working among older people and the extent to which poor health might be an obstacle to their employment.

This monograph deals with the sampling aspects of the study and presents in more than the usual detail the decisions that had to be made and the procedures finally followed in sampling for a household survey. The methods used in estimating sampling error from the survey results are also set forth. Subsequent papers will cover the evaluation of the interview schedule, analysis of scales of availability for work in the Hagerstown trial, and the illness among older persons in the city as revealed by the interview.

The requirements of the study which chiefly determined the sample design were: (*a*) the need for a probability sample of all persons 45 years of age and over, exclusive of those in resident institutions, living within the city limits of Hagerstown; (*b*) the need for a sample of sufficient size to permit the analysis that was contemplated; and (*c*) the undesirability of conducting more than one interview in any one household.

When we employ probability sampling we can say in advance how much sampling error we are willing to tolerate and then select the sample so that statistical conclusions about the characteristics of the whole population will not have a sampling error greater than the tolerable limits, except very rarely. Probability sampling requires that the probability of including any individual in the sample be known. In order to accomplish this, it is necessary to have some sort of list of all the individuals, or of units including all of the individuals, in the population that is to be sampled. In Hagerstown, this list was supplied by the city directory and a city map.




## No. 40

The accompanying summary covers the principal characteristics of the sampling plan for a small household survey described in Public Health Monograph No. 40, published concurrently with this issue of Public Health Reports. The author is a biostatistician with the Division of Public Health Methods, Public Health Service.

Readers wishing the data in full may purchase copies of the monograph from the Superintendent of Documents, Government Printing Office, Washington 25, D. C. A limited number of free copies are available to official agencies and others directly concerned on specific request to the Public Inquiries Branch of the Public Health Service. Copies will be found also in the libraries of professional schools and of the major universities and in selected public libraries.

• • •

Woolsey, Theodore D.: Sampling methods for a small household survey. Public Health Monograph No. 40 (Public Health Service Publication No. 480). 16 pages. Illustrated. U. S. Government Printing Office, Washington, D. C., 1956. Price 20 cents.

The bulk of the sample was obtained by selecting what is known as a systematic sample of addresses from the directory.

Since directories are out of date as soon as they are issued, it was known that the sampling from the directory would have to be supplemented to allow a chance for households living at addresses not listed in the directory to come into the sample. This was managed by selecting a sample of blocks from a listing of all city blocks made up from a map of the city. Each block falling into the sample was canvassed, and all addresses were carefully checked against the directory. Any address in one of these sample blocks that was not listed in the directory was included in the sample. Finally, one area of the city where there had been a substantial amount of new construction was treated separately. All addresses in this area were listed and a sample was drawn.

Thus, the probability sample was carried out by dividing the addresses in the city into three strata: directory addresses, unlisted addresses scattered throughout the city, and unlisted addresses concentrated in new construction areas. An account is given in the monograph of the sampling in each of these three strata. Some unexpected obstacles were encountered, and the method of overcoming these is described.

The objective of obtaining a sample of sufficient size to support the desired analysis was achieved partly by guesswork and partly by providing a safeguard in case the guesswork was poor.

It had been decided, for reasons that will be presented later, that only one person 45 years of age or over could be interviewed in each household. The question was: In what proportion of the Hagerstown households would one or more persons who were past the 45th birthday be found? There were no available statistics on this point, but a guess was made that this proportion would be 50 percent. Since 500 completed interviews were thought to be a minimum number necessary for the analysis, it appeared that a sample of 1,000 households would be required to produce 500 households containing one or more persons 45 years of age or older. As a safeguard, however, it was decided that three separate samples of 500 households each should be selected. After the interviewing in each sample had been completed, it would be possible by hand tally methods to determine not only how many completed interviews were available but also how large some of the frequencies that were important for the analysis would be. If a larger sample appeared to be needed, it was merely necessary to proceed with the interviewing of an additional sample of 500 households.

The population of Hagerstown in the 1950 census was 36,260; the number of households, approximately 11,400. To make allowance for inability to interview some persons, a sampling ratio of 1 in 21 for each of the three samples was decided upon. This ratio could be expected to provide 540 or more households in each sample. As it turned out, 56 percent of the households contained one or more persons 45 years of age or older. Consequently, it was found possible to stop the interviewing after the completion of the second sample.

Many of the questions in the interview were of a type that could be answered properly only by the person whom the question concerned. In other words, a proxy respondent could not be accepted. It was believed, however, that, if more than one person 45 years of age or older were to be interviewed in any one household, the replies of the later respondents would be influenced by overhearing the earlier interviews. It was, therefore, a requirement of the study design that a single respondent 45 years of age or over be selected from among those in this age group in the household. Furthermore, in order to meet the probability sample requirement mentioned earlier, the single respondent must be selected in such a way that the probability of any given individual coming into the sample could be known. In other words, the interviewer must be allowed no freedom in the selection of a respondent from among those 45 years of age or over who were found in the household.

This was accomplished by providing for opening questions to be addressed to any adult who could answer for the household as a whole. The first of these questions (which followed a brief introduction) was designed to find out whether there was anyone in the household who had passed the 45th birthday. If there was not, the interview was terminated. If there was, a complete roster of the members of the household was

obtained as well as certain information about the demographic characteristics and labor force status of each person.

When the household contained only one member 45 years of age or older, all the questions on health and availability for work were addressed to this person. When there was more than one such member, the interviewer was provided with a sampling table which specified which member was to be interviewed in that household. Thus, in households with two persons in the 45-years-or-over class one of the two was selected at random, in households with three such persons one of the three was selected, and so on. In analyzing the results, it was necessary to restore the balance in this sampling by weighting each interview by the number of persons in the household who were in the age group of interest. This was done by reproducing the punchcard that was prepared from the interview, as many times as necessary.

The monograph also presents statistics showing the degree of success which the interviewers had in reaching the sample. Of the total of 1,114 households in which an initial interview was desired, the interviewers completed 1,093 (98 percent). Six hundred fifteen of the households in which an initial interview was completed were determined to have one or more persons 45 years of age or over. The main interview was sufficiently complete to be included in the study in 600 (98 percent) of these. The 600 households contained 1,030 persons 45 years of age or older, but, as has been stated earlier, only one such person in each household was interviewed.

The monograph completes the exposition of the sampling methods employed with an illustration of the method of estimating the sampling errors from the survey results. A table of standard errors for percentages is also presented.

The sampling methods used in the Hagerstown survey are believed to have a wide range of application where the investigative technique is the interview-survey of a population sample.

## Institutes in the Care of Premature Infants

A series of institutes for physicians and nurses in the care of premature infants is scheduled for the 1956–57 year at the New York Hospital-Cornell Medical Center. These institutes, which have been held each year since 1949–50, are sponsored by the New York State Department of Health and the United States Children's Bureau. They are designed for physicians and nurses in charge of hospital premature nurseries and special premature centers and for medical nursing directors and consultants in State and local premature programs.

Institutes are scheduled to begin September 17, 1956, November 5, 1956, and January 14, 1957. If the number of applicants is sufficient to warrant them, fourth and fifth institutes will be held beginning March 11, 1957, and early in May 1957.

Attendance at each institute is limited to 6 physician-nurse teams. The program for physicians lasts 2 weeks, and that for nurses, 4 weeks. Participants pay no tuition, and stipends are provided to help cover expenses during attendance. Early application is essential since plans are contingent on the number of applicants.

Additional information may be obtained by writing Box 143, Institute in the Care of Premature Infants, The New York Hospital, 525 East 68th Street, New York 21, N. Y.