A NEW SAMPLING PLAN FOR THE CAROLINA HEALTH SURVEY

Jan-Jan Lam, University of South Carolina
Martin C. Weinrich, University of South Carolina

INTRODUCTION

The High Blood Pressure Control Program of the South Carolina Department of Health and Environmental Control (DHEC) is designed to coordinate existing programs and to encourage new initiatives to control high blood pressure in South Carolina. The Carolina Health Survey is an important part of the evaluation component of the High Blood Pressure Control Program. Both the Control Program and the Survey are funded by the National Heart, Lung and Blood Institute. Similar control programs and surveys are in progress in six other states.

The Survey has two principal goals: first, to provide feedback for hypertension detection and control programs in South Carolina, using data gathered in the Baseline Survey. Second, to help evaluate DHEC's coordination and intervention efforts. The specific objectives of the Survey are to gather information on blood pressure levels in South Carolina adults, as well as information on the prevalence, awareness, surveillance, treatment and control of hypertension.

In the sections that follow, we will first describe the sample design for the Baseline Survey. Then we will describe the goals for our follow-up survey and the sampling design that we plan to use. Finally, we will comment on the differences.

SAMPLING PLAN OF THE BASELINE SURVEY

The Baseline Survey was carried out from late 1978 to early 1980. It was designed as a series of cross-sectional surveys, that is, independent replicates, each consisting of a sample of approximatley 900 adults, and corresponding to about six weeks' worth of field work. Each replicate was drawn as a multi-stage cluster sample, and field work was carried out for several periods at once.

The Primary Sampling Units (PSU's) of the Baseline Survey were called "Sampling Districts". They were essentially 1970 Census Enumeration Districts (ED's), of which there were about 3000 in the state. These were combined or split as necessary, in order to insure that the variations in population were not too great. A simple random sample of these PSU's was drawn, and the selected PSU's were mapped in the field and divided into clusters of approximately fifty housing units apiece. Clusters were then selected at random from the list formed by mapping primaries, and the clusters selected at second stage were then mapped in detail, housing unit by housing unit. A systematic sample of eight housing units was then chosen from each of the clusters, and all adult residents of these housing units were included in the sample.

Two unusual features of this design are perhaps worth noting here: first, PPS sampling was not used at any state of sample selection. Second, there was no stratification of the primary sampling units.

GOALS AND GENERAL CONSIDERATIONS FOR THE FOLLOW-UP SURVEY

In designing the follow-up survey, we took into account the following goals: first, to estimate the prevalence of hypertension, both state-wide and for selected areas of the state, such as health service areas. Second, to compare blood pressures and the prevalence of hypertension in different demographic subgroups, such as black vs. white, male vs. female, and so on. Third, to compare the results of our follow-up survey with those of the baseline survey in order to estimate the magnitude of any change that may be occurring. These comparisons will be made on a state-wide basis as well as for specific areas and subgroups of the population.

We considered several ideas before deciding on a sampling plan. One idea that was discarded was to re-visit respondents from the baseline survey. We rejected this idea, in part because we would not have been able to separate the effect of being interviewed and having your blood pressure taken from the effects of other influences on blood pressure. Even had there been large differences, we would not have been able to rule out our survey itself as their cause. Another idea that was considered initially was to re-use the PSU's from the baseline survey. This might have reduced the standard errors of the estimates in comparison to those from an independent sample drawn using the same sampling plan as the baseline survey. But we were confident that we could obtain a more efficient design by starting from scratch.

SAMPLING PLAN FOR THE FOLLOW-UP SURVEY

Based on the above goals and considerations, we have made several significant changes in our sampling plan, which should result in major improvements. The PSU's will be stratified, using both geographic and demographic variables. Both maps and counts of population and housing units from the 1980 Census will be used to define boundaries and to obtain measures of size for the primary and secondary sampling units. Use of systematic sampling with probability proportional to size in a geographically ordered list will insure that a good geographic spread is obtained, and that the sample is "self-weighting". Variance estimates will be obtained using balanced repeated replicates, rather than a few independent replicates as in the baseline survey.

The 1982 sample will contain about 3000 households based on the 1980 Census data, these will contain about 5000 adults 18 years or older. The sample households will be chosen in three stages, as outlined below.

At the first stage of sampling, Primary Sampling Units (PSU's) will be chosen using population counts from the 1980 U.S. Census. These counts will be obtained from the PL-94 171 reapportionment tape, where they are given by Block Group (BG) and Enumeration District (ED). Block statistics are given for all cities with populations of ten thousand or more; in these areas, Block Groups will be used as PSU's. In areas where block statistics are not available, Enumer-

ation Districts will be chosen as PSU's. In contrast to 1970, Block Groups and Enumeration Districts are mutually exclusive in the 1980 Census. This allows a rough stratification of the PSU's into urban and rural PSU's. These PSU's will be chosen without replacement with probability proportional to size (PPS) using systematic (fixed-interval) selection from lists that are sorted as outlined later. Our measure of size for PSU's is total population, since counts of housing units at BG and ED levels will not be available at the time the sample of PSU's is drawn.

Stratification by race-mix and geography will be accomplished by sorting the BG and ED lists as follows. First, the PSU's will be sorted by Health System Area (HSA). There are five HSA's in South Carolina, one of which lies mainly in Georgia and contains only one county in South Carolina. That county has been absorbed into the Three Rivers HSA, which leaves four HSA's for stratification purposes. Variation in their populations is not large enough to justify a departure from EPSEM sampling. The four HSA's and two types of PSU's form eight strata with firm boundaries. Within each of these eight strata, the PSU's will be sorted as follows. The primary sort will be into three categories based on the racial mix, that is the percent of blacks in each PSU. Each of these racially homogeneous sets of PSU's will be sorted by county, tract (where applicable), and block group or ED number. In the untracted counties of the state, Census County Divisions (CCD's) will be used in place of tracts for sorting.

Estimates of the intraclass correlation coefficients and data on travel costs from the baseline survey have been used to estimate the optimal PSU size, that is 20 adults interviewed per PSU. This is equivalent to 14 housing units after adjustment for non-response. The appropriate number of PSU's is therefore 216. Since the population in the ED areas and the blocked areas is about the same, half of these, or 108 PSU's will be allocated to the blocked areas. In order to save travel cost, we doubled the optimal PSU size in the rural areas. Therefore, only 54 PSU's will be selected from among the ED's. Linking of small PSU's to other PSU's will be performed where necessary in order to meet minimum size requirements. With the sorting order mentioned above, systematic PPS sampling of PSU's should yield small standard errors and a good cross-section of the PSU's statewide.

The PSU's choosen for the first-stage sample will then be divided into chunks for the second-stage sample. The ED's will be mapped and their housing units will be counted in the field; for the blocked areas, we plan to use Census housing units counts from the MARF (Multiple Area Reference File) tape. These housing unit counts will then be used for chunking. The minimum PSU size will be large enough so that we can always form four or more chunks within each Primary Sampling Unit. The minimum size for these chunks is seven housing units (HU's) for block groups and fourteen HU's for ED's.

The second-stage sample will be drawn by choosing two chunks from the four or more in each PSU; selection will be PPS, and the measure of size will be the number of housing units (from Census data or field observation), rather than population count.

The chunks selected at the second stage will then be listed, housing unit by housing unit, for the third stage of sampling. A systematic sample of seven HU's per chunk in the blocked areas, and fourteen HU's per chunk in the ED's, will then be drawn, using the lists of housing units. During actual interviewing, the usual "half-open interval" procedure will be used, in order to insure that the 5% of housing units that our listers may miss do have the same chance of inclusion in the sample as any other HU's in the state. We are confident that the follow-up survey design can achieve a design effect close to unity for blood pressure and related variables. This contrasts with the baseline survey for which our estimated design effect was about 1.5 using independent replicates.

SUMMARY

The procedure outlined above should insure that every adult in the state has roughly the same probability of being included in the sample. It should give good precision for estimates, entirely comparable (at least for blood pressure measurements) to that obtained in a simple random sample of the same size. It will allow reasonably good precision for HSA-level estimates, and minimizes the chance that a "non-representative" sample of housing units is selected by "bad luck". We further expect to save travel cost -- especially in rural cases, since our new design doubles their chunk size. Finally, the new design uses only two of the four or more chunks in second stage sampling, so that much of the work of developing a sample frame for future surveys has already been accomplished at modest additional cost.