

A STATEWIDE SAMPLE SURVEY TO ESTIMATE THE PREVALENCE OF UNCONTROLLED HYPERTENSION

Donna R. Brogan, Emory University

The Department of Human Resources (DHR) of the state of Georgia is under contract to NHLBI to demonstrate the impact of statewide coordination of hypertension control activities upon hypertension control success and upon reduction of mortality from hypertension related diseases. As part of this demonstration a household sample survey of the state of Georgia is underway in 1980 and will be repeated in 1985 in order to estimate the percentage of the adult Georgia population who have uncontrolled hypertension. This percentage, or prevalence rate, is expected to decrease between 1980 and 1985 as a result of the statewide coordination activities.

For the purpose of this sample survey, uncontrolled hypertension is defined as a diastolic blood pressure ≥ 105 mm. Hg and its prevalence was estimated to be 3.8% in 1980. (Hanes-Health and Nutrition Examination Survey-data and data from the DHR hypertension screening program were used to derive this estimate). Using $\alpha = .05$ and $\beta = .10$, a sample size of 1000 housing units (H.U.'s) was determined necessary in 1980 and in 1985 in order to detect a decrease of 50% or more in the prevalence rate from 1980 to 1985. (Note: blood pressures are measured on all adults within a selected H.U.) This sample size estimate was determined by first assuming a simple random sample of H.U.'s and then inflating that sample size estimate by an estimate of the design effect or deff ($\hat{\delta} = 0.2$ and an average cluster size of 2.1 adults).

Counties were used as the primary sampling units (p.s.u.'s). There are 159 counties in Georgia, ranging from 1,300 to 419,000 in estimated 1978 adult population. We stratified the counties on three variables to try to reduce the variance of estimates and to allow comparisons between strata of interest to us. We used health district as our geographic region stratification variable; each county belongs to one, and only one, health district. The health district containing the Atlanta metropolitan area was subdivided into two regions because of its large population, resulting in a total of eleven health districts. The second stratification variable was urban/rural; a county was defined as rural if it contained no city, town, or place with a total population of 17,000 or more. All counties not rural were defined as urban. In some cases urban counties were further classified as an SMSA (Standard Metropolitan Statistical Area) county or a nonSMSA county. Thirdly, we categorized each county by percent nonwhite population as follows: $\leq 40\%$, 41%-60%, 61% or more. Using these three stratification variables, the 159 counties were categorized into 38 strata. All combinations of possible values of the three stratification variables were not realized, so the stratification variables are not completely crossed. Rather, the structure can be viewed as each health district containing various sub-strata. Eight of the 38 substrata contained only one county.

Six counties in Georgia (Macon, Peach, Taylor, Crawford, Twiggs, Wilkinson) were excluded from

the possibility of being selected as a sample p.s.u. for this statewide survey because they are included in a similar household survey being planned under an NHLBI grant to Fort Valley State College. These six counties have an estimated 1978 adult population of 42,300 and comprise 8% of the total adult population of the two health districts in which these counties are located. The six counties comprise just under 1% of the total Georgia adult population. Fort Valley grant personnel will share the data from their household sample survey with the statewide survey personnel so that statewide and health district estimates can be constructed. In effect, this plan is equivalent to allocating these counties to two additional substrata, these six counties falling in 2 of the 11 health districts for the statewide survey. The sampling rate is 80 times greater in these two additional substrata than in the other 38 strata combined.

The allocation of the 1000 sample H.U.'s to the 38 substrata in the statewide survey was done in two stages. First of all, each of the eleven health districts was allocated the same number of H.U.'s, even though the eleven health districts ranged in estimated 1978 adult population from 153,000 to 747,000. This was done because it was considered of prime importance to make estimates for different geographic areas within the state. Within a particular health district, allocation of the sample size to the substrata was done proportional to the estimated number of adults in each substratum.

The allocation of the total sample size for the six Fort Valley counties differed somewhat. Four of the counties are "experimental" counties and will receive an intervention. The other two counties are "control" counties (no intervention). The four experimental counties were allocated 480 sample H.U.'s, as were the two control counties. Within the four experimental counties and the two control counties, allocation of sample H.U.'s to each county was done proportional to estimated adult population.

For both surveys a chunk or a segment was defined as 24 H.U.'s. In the statewide survey, the sampling rate within chunks was chosen to be 1/4. The selection of a somewhat larger chunk size than usual for the statewide survey was determined by (1) not wanting to do more intense sampling within chunks than 1/4 and (2) wanting to have at least 6 H.U.'s to interview in a selected chunk. Since the sample size of the Fort Valley survey is about the same as the statewide survey, but for a much smaller geographic region, the sampling rate within chunks for these six counties was chosen to be 1/2. Note that this arrangement gives the opportunity to do some empirical investigations of the effect of within chunk sampling rate on point estimates of means and ratios and point estimates of intracluster correlation coefficients.

A further difference between the Fort Valley survey and the statewide survey is the opportunity to plan for replicate samples in the Fort Valley survey. Since this survey is within a much smaller geographic area, enough chunks were

allocated to each of the six counties to allow the sample to be selected as four independent replicates. This will allow a quick computation of estimated variances and will also allow some empirical investigation of different techniques for estimating variances and intracluster correlation coefficients.

Because of the defined chunk size as 24 and the within chunk sampling rate as 1/4 for the statewide survey, allocated sample size to each selected county needed to be a multiple of 6 H.U.'s. For Fort Valley, allocated sample size needed to be a multiple of 48 H.U.'s due to planning four replicates using chunks of 24 H.U.'s sampled at a 1/2 rate. Hence, the allocated sample sizes determined previously were adjusted to meet this constraint. In the statewide survey, the sample size was always adjusted upward for rural counties and for counties with a percentage nonwhite population of 40% or more. This was done to give a slight oversampling of nonwhite and/or rural residents. With this procedure we anticipate 27% of our sample subjects to be nonwhite and 41% of our sample subjects to be rural.

Since both surveys are being fielded 10 years after the 1970 census but before official 1980 census results are available, it is risky to rely only upon 1970 census data for the within county sampling. Thus, we obtained recent information on each sample county and judged whether or not it was necessary to update the 1970 within county census data. One good source was the 1977 estimates of county population, by place (CPS). Another good source was local area planning and development commissions. The Atlanta Regional Commission provided 1979 estimates of number of H.U.'s by census tract for the counties comprising the Atlanta metropolitan area. Georgia Department of Transportation furnished some useful data from traffic zone studies. The 1970 within county census data were used for those counties showing less than a 15% change from 1970 to 1978 or 1979 in total population and/or number of H.U.'s. For counties showing more change than this, we updated the 1970 estimates of H.U.'s or population using a variety of techniques, depending upon available information.

In tracted counties the sample chunk(s) was (were) allocated to census tracts using systematic pps sampling, where size is number of H.U.'s. Within a selected census tract, the sample chunk

(s) was(were) allocated to a block group, or to a block where appropriate, again using systematic pps sampling. At this point the selected geographic area was visited by a field worker and a list of addresses was compiled for the area. From this list the statistical office defined chunks, selected one or more chunks by simple random sampling, and selected a random start for the within chunk systematic sample.

In nontracted counties, the sample chunk(s) was(were) allocated to enumeration districts (E.D.'s) using systematic pps sampling. If the selected E.D. was small enough, the entire E.D. was listed as described above. If the selected E.D. was too large to list, a further geographic breakdown into sub E.D.'s was done by using locally available data or field inspection. Following this, the sample chunk(s) was(were) allocated to sub E.D.'s by systematic pps sampling. The selected sub E.D.'s then were listed as described above and chunks were chosen.

A difficulty in planning both surveys is the nonavailability of 1980 census data at this point. Some 1980 preliminary census estimates have been used in counties with large growth since 1970. However, its imminent availability will allow an empirical investigation of the accuracy of our update procedures. Further, since both surveys are being fielded six to eight months after the 1980 census, the 1980 census data can be used effectively for poststratification purposes in the analysis phase.

After the statewide survey was designed, additional funding was received from NHLBI to increase the total sample size from 1000 to 2000 H.U.'s. This will allow us to make more precise comparisons of various health districts or various subpopulations. No additional p.s.u.'s were selected. For a blocked census tract already selected within a sample county, an additional block or block group will be selected with pps, and one chunk will be selected from that block or block group. For an E.D. already chosen and completely listed, an additional chunk will be randomly selected. For a larger E.D. already chosen and broken down into sub-E.D.'s for listing purposes, an additional sub-E.D. will be chosen with pps. All of these selections will be done as though a systematic pps sample of 2 chunks were being selected from each E.D. or census tract.