

Donna Brogan and Carolyn Miles, Emory University

General Survey Sample Design

An area probability sample of 2112 housing units in Georgia was planned in 1980 and fielded in 1981 in order to estimate the percentage of Georgia adults who have uncontrolled hypertension (diastolic blood pressure ≥ 105). The primary sampling units (p.s.u.'s), 153 counties, were classified into 38 strata on the basis of geographical location, urban/rural, and percentage of county population which was black by using 1978 county population projections obtained from the Georgia Department of Planning & Budget. One p.s.u. per stratum was selected via p.p.s. sampling, where "size" was the estimated 1978 adult population for the county. Each of six remaining counties in Georgia was sampled in a second household survey using the same interview instrument. The data from these 6 counties will be combined with the 38 other p.s.u.'s to form statewide estimates.

For each selected p.s.u., within county sampling was planned so as to give each housing unit in that county the same probability of being selected. Every adult in a selected H.U. was to be selected for the sample. In most area probability samples of H.U.'s, census data and maps are used to divide the county into geographic areas with a specified number of H.U.'s. Selections of particular census tracts (C.T.'s), enumeration districts (E.D.'s) or blocks then are made using p.p.s. sampling; selected areas are counted and listed by field persons, and addresses (H.U.'s) are chosen to be in the sample.

This survey faced the major problem of the most recent census data being 10 years old, with virtually no local availability of systematic data on number of H.U.'s for geographic subdivisions below the county level. Thus, we developed the following procedure to update the 1970 census data, where necessary, for the 44 counties (p.s.u.'s) selected for the sample.

Updating Procedures

First we determined whether each selected county had increased significantly in population (15% or more) since 1970. The two data sources used to determine population increase were the 1977 Current Population Survey (CPS), published in 1979, and 1978 estimates of county population. Based on this data, 10 of the 44 counties did not change significantly. No population or H.U. data were available for most counties past 1977-8 to 1980. Thus, we used 1970 census data on H.U.'s to select geographic areas for listing in these 10 counties - E.D.'s in 5 counties and blocks in 5 counties.

Five of the seven counties in the Atlanta SMSA were sample p.s.u.'s and had shown remarkable increase in population and housing since 1970. We used data from the Atlanta Regional Commission for these 5 counties to select 31 C.T.'s with pps sampling, where "size" was 1979 estimated number of H.U.'s per C.T. Ten of the selected C.T.'s had not changed significantly in H.U.'s since 1970; thus 1970 census H.U. data was used

to do pps sampling down to the block level for these C.T.'s. For the 21 remaining C.T.'s, we waited for 1980 provisional census data in order to select E.D.'s within the C.T. because updating the 1970 census for these selected C.T.'s would have been a costly field operation. For example, some C.T.'s had increased up to five-fold in H.U.'s from 1970 to 1979. Provisional 1980 census figures for selected counties could not be obtained from the Atlanta Regional Office of the Census Bureau. However, in all cases, sometimes after much effort, we obtained a copy of the provisional census figures and the maps from the appropriate county official.

For selected counties outside the Atlanta SMSA which had increased significantly in population since 1970, we were not able to find data similar to that available from the Atlanta Regional Commission. Some regional planning commissions had estimates of H.U.'s as recent as 1978, but the estimates were only for the entire county and/or for cities/towns therein. Since much of Georgia is rural, these data were not very useful in subdividing the county into smaller geographic areas with corresponding estimates of H.U.'s. Personal visits to utility companies, tax offices, and other local government agencies in general produced no immediately useable data for within county sampling. Sometimes, though, we were able to combine locally gathered information with other available information to estimate number of H.U.'s for county geographic subdivisions.

In two counties we obtained from the tax office the number of H.U.'s per tax district. In one county we used this information to estimate the number of H.U.'s per E.D., followed by pps sampling of the E.D.'s, followed by pps sampling of tax districts within selected E.D.'s. In the other county we chose tax districts directly using pps sampling.

In two counties we obtained estimated H.U.'s per "traffic zone" from the Georgia Department of Transportation. In one county we used the 1978 traffic zone data to first select C.T.'s with pps sampling. For selected C.T.'s which had not shown significant growth in H.U.'s since 1970, we used 1970 block data to do further pps sampling. For selected C.T.'s which had grown significantly since 1970, we selected traffic zones with pps sampling; traffic zones are larger than blocks but smaller than C.T.'s, ranging from about 50 H.U.'s to 900 H.U.'s. For the second county we sampled traffic zones directly using pps.

After exhausting the procedures above, there remained 25 counties where the population had increased significantly since 1970. The most recent data for these counties was the 1977 CPS which estimated total population for the county and for each city/town within the county. We assumed that that the percentage increase in H.U.'s since 1970 was equal to the percentage increase in population from 1970 to 1977; this yielded an estimate of the number of increased H.U.'s in the entire county and in each city/

town. However, we needed an updated estimate of H.U.'s for each E.D. in the county. For cities/towns which comprised only one E.D., we used the CPS data to obtain the updated estimate. For cities/towns of more than one E.D., we used the CPS data to estimate the total number of additional H.U.'s from 1970 to 1977, but we did not know to which E.D.'s to allocate the increased housing units. Likewise, we used the CPS data to estimate the additional H.U.'s throughout the unincorporated sections of the county, but again did not know to which specific E.D.'s to allocate the additional H.U.'s. At this point field staff were asked to determine, for particular groups of E.D.'s within a given county, the location and number of new H.U.'s in the county since 1970, especially large subdivisions. City/town officials, zoning officials, planning agencies, utility companies, and tax offices were approached to locate this information. (Note: published data on housing permits issued were not useful since (1) they do not go below the county level and (2) housing permits do not always end up as housing units.) The amount of information obtained ranged from excellent to nil, depending upon the willingness and knowledge of the approached county officials. Local information on additional H.U.'s was obtained for six counties and the E.D.'s were updated accordingly. Then E.D.'s were selected with pps sampling for these six counties.

There remained 19 counties where it was not possible to allocate all, some or any of the additional H.U.'s to particular E.D.'s based on local information. In eleven of these counties, the additional H.U.'s were allocated to the appropriate group of E.D.'s proportional to number of H.U.'s in the E.D. in 1970. E.D.'s then were selected, selected E.D.'s were listed and H.U.'s were chosen. For the remaining eight counties with significant growth, we waited for 1980 provisional census data before selecting E.D.'s for listing.

A summary of the various updating techniques is:

- (1) No updating from 1970 census because no significant change in total county population from 1970 to 1977-78 (10 counties).
- (2) Use of 1979 estimates by C.T. for Atlanta SMSA counties, followed by 1970 block data for chosen C.T.'s that had not changed significantly from 1970 and waiting for 1980 provisional data for the remaining chosen C.T.'s (5 counties).
- (3) Use of H.U. information for tax districts from county tax offices (2 counties).
- (4) Use of H.U. information for traffic zones from Ga. Dept. of Transportation (2 counties).
- (5) Use of 1977 CPS survey supplemented by local estimates of E.D.'s in which additional H.U.'s had been built since 1970 (6 counties).
- (6) Use of 1977 CPS survey with minimal or no local information to pinpoint E.D.'s where an increase in H.U.'s had occurred (11 counties).
- (7) Use of 1980 provisional census data (8 counties).

Evaluation of Updating Procedures

Since selection of E.D.'s and blocks for listing was done in the latter half of 1980, the 1980 census data on H.U.'s provides an evaluation of the accuracy of some of the various updating techniques. The only census data available to date in Georgia is final counts of population and housing units for counties and county subdivisions (cities/towns). This will not allow comparisons with 1980 census data on any of the 1970 block data, but will allow comparisons of single E.D.'s where a city/town comprises one E.D. or groups of E.D.'s where a city/town comprises more than one E.D.

Technique 1: In the 5 counties where 1970 block data were used to select blocks with pps sampling, a comparison can be made for the selected blocks between number of H.U.'s from 1970 census data and number of housing units actually listed by the field person. The average percentage discrepancy per block, $\left[\frac{\text{Listed H.U.} - 1970 \text{ H.U.}}{1970 \text{ H.U.}} \right]$ for these 5 counties was .31. The 1970 census underestimated H.U.'s much more often than overestimated them.

For the 5 counties where 1970 E.D. data were used to select E.D.'s with pps sampling, the 1970 census data were compared to the 1980 census data for each E.D. or groups of E.D.'s throughout each county. The average percentage increase in H.U.'s per E.D. ranged from .16 to .31 in these five counties.

For the 5 counties where blocks were selected with 1970 census data, the percent increase in HU's countywide from 1970 to 1980 was 13%, 18%, 29%, 36% and 54%. The latter county showed a 7% decrease in population from 1970 to 1977 according to CPS, showed a 1.3% increase in population from 1970 to 1980, yet showed a 54% increase in HU's from 1970 to 1980.

For the 5 counties where ED's were selected with 1970 census data, the percent increase in HU's countywide from 1970 to 1980 ranged from 17% to 29%.

Technique 2: The 5 Atlanta SMSA counties increased in HU's from 1970 to 1980 by 19% to 162%. For those 10 CT's where 1970 block data were used to select blocks, the results were similar to that described above for the 5 blocked counties of Technique 1. For the 21 CT's which had increased significantly in population from 1970 to 1979, we waited for 1980 provisional census figures to sample ED's from these CT's. The 1980 provisional census figures were very close to the number of listed HU's as well as close to the final 1980 census figures.

Technique 3: In one county, tax district information was used, along with 1977 CPS, to update estimates of HU's per ED. This county increased 36% in HU's from 1970 to 1980. For each ED, comparison was made between estimated HU's per ED (or group of ED's) and 1980 HU's from census data. The average of the discrepancies in HU's per ED, $\left[\frac{1980 \text{ HU} - \text{estimated HU}}{\text{estimated HU}} \right]$, was .15. Thus, this technique worked well in this one county.

The other county increased 79% in HU's from 1970 to 1980. Tax districts were chosen with pps without choosing ED's first, so no comparison can be made between estimated HU's per tax

district and 1980 census data. In this county some selected tax districts had more HU's listed than estimated by the tax district office. In many cases this seemed to happen because the selected tax districts contained several trailers. All in all, selection by tax districts was much better than using 1970 census data for this county.

Technique 4: In the two counties where traffic zones were selected, the percent increase in HU's from 1970 to 1980 was 25% and 34%. No comparison can be made between 1980 census figures and traffic zones. Comparison of estimated HU's per traffic zone showed no major discrepancies with listed HU's per traffic zone. Thus, this technique worked fairly well.

Technique 5: In these six counties, the percent increase in HU's per county from 1970 to 1980 ranged from 39% to 56%. For each county, comparison was made for each ED or a group of ED's between 1980 HU's and estimated HU's using technique 5. The average of the percentage discrepancies in HU's per ED, $\left[\frac{1980 \text{ HU} - \text{estimated HU}}{\text{estimated HU}} \right]$, ranged from .15 to .25. among 5 counties with one county at .43. Although technique 5 underestimated the HU's per ED in almost all instances of discrepancy, the extent of underestimation was considerably less than it would have been had 1970 census data been used.

Technique 6: In these 11 counties the percent increase in HU's per county from 1970 to 1980 ranged from 19% to 50% for 10 counties, with one county actually losing 7% of its HU's from 1970 to 1980. (Technique 6 was used on this county because 1977 CPS data indicated that one of its cities had increased 23% in population from 1970 to 1977, although the 1980 census later indicated that the city lost 7% of its population from 1970 to 1980.) Over all 11 counties, the average of the percent increases in HU's from 1970 to 1980 was 30%.

For each county, comparison was made for each ED or a group of ED's between 1980 HU's and estimated HU's using Technique 6. The average of the percentage discrepancies in HU's per ED (or ED group), $\left[\frac{1980 \text{ HU} - \text{estimated HU}}{\text{estimated HU}} \right]$ ranged from .13 to .41, with a mean over the 11 counties of .25. Almost all instances of discrepancy were underestimations of the HU's per ED. However, the extent of discrepancy was less with Technique 6 than it would have been using only 1970 census data.

Technique 7: The eight counties for which we waited for 1980 provisional census data for selection of all ED's increased in HU's from 1970 to 1980 from a minimum of 29% in one county to a maximum of 124% in another county, or an aver-

age over the 8 counties of 57%. (Note that it was impossible to wait for 1980 provisional data for all 44 counties in the survey because we needed to begin interviewing in early 1981.) This method worked well in all instances as evidenced by the field count of HU's in selected ED's being very close to the 1980 provisional census figures for those ED's. Further, a comparison of the provisional 1980 census figures with the final 1980 census figures for these counties showed minor variations.

Conclusions:

(1) It was very time consuming to search for and collect the data to implement these updating procedures as well as to use several different within county sampling schemes for the 44 selected counties. Note, also, that the diversity of within county sampling schemes used here precludes the possibility of closed form variance formulas for point estimates. However, this procedure was judged preferable to using only 1970 census data to select ED's and blocks for listing.

(2) When using the 1977 CPS data to update the 1970 census, we assumed that the percent increase in population from 1970 to 1977 was equal to the percent increase in HU's from 1970 to 1977. This assumption was faulty. For the 44 counties in our survey, the percent increase in HU's from 1970 to 1980 was about 2.5 times larger than the percent increase in population from 1970 to 1980. Thus, in almost all cases, we underestimated the increase in HU's since 1970. One extreme county had a 1.3% increase in population and a 53.5% increase in HU's.

(3) Utilizing 1970 census data in counties or ED's or CT's where we estimated no significant increase in HU's turned out moderately well. We underestimated HU's in most cases, but not by serious degrees in most instances.

(4) Tax district information was useful to update HU's per ED if (a) the tax district boundaries could be mapped fairly well into ED boundaries and (b) if the tax district figures did not seriously underestimate the number of HU's (e.g. trailers, multi-family housing, etc.)

(5) The traffic zone data worked well in the 2 counties where it was used. However, such data are not available on a regular basis for all counties.

(6) Using the 1977 CPS to update ED's certainly was better than using the 1970 census only. Supplementing CPS with scouting in the county to locate major areas of new construction improved this method further. This method seems a reasonable method to use if updating is needed and no better information is available.