

Mary Ellen Beach and John W. Marshall ^{1/}, U.S. Bureau of the Census
 Mary Ellen Beach, Rm. 3285/3, Field Division, Washington, D.C. 20233

1. INTRODUCTION

Taking the decennial census of population and housing requires an enormous expenditure of money and human resources. Different levels of resources and varying strategies for publicity and field activities are needed in different areas of the country because of our socio-economic and cultural diversity. Within the Census Bureau, the concept "hard-to-enumerate" (or HTE) refers to areas that require extraordinary efforts (and therefore resources) for accurate enumeration. During the testing cycle for the 1990 Census, we have begun to use the wealth of demographic and administrative information we possess in order to identify and plan for those areas and, ultimately, to improve our decisions about how to allocate resources.

Data that identify and describe hard-to-enumerate areas have many potential uses in census planning. On the large scale, HTE information can help us stratify our data collection offices so that the most appropriate field methods, procedures, and schedules are used in each area. Our efforts to design a successful community awareness program to improve cooperation with the Census could be more appropriately focused, and consequently more productive, if we used Census data to "target" specific areas or groups that are expected to need more intense outreach. On the small scale, we can vary the allocation of staff and resources within the collection offices by planning for special procedures (such as team enumeration in high crime areas) or administrative arrangements (such as the manager/supervisor/worker ratio or the compensation package that will be offered). Our Census Community Awareness Program will be more effective if strategies and staff resources are varied within the specific target areas. Our ability to predict mail response rates at low levels of geography will allow us to "target" the mailing of reminder cards. In this way, we can use the reminder cards to increase mail response specifically in those areas where the potential benefit is the greatest and avoid the expense of a "global" mailing.

2. CONCEPT

When we first began to think about why an area is hard-to-enumerate, it became apparent that the concept was complex. We identified three aspects that characterize areas that are HTE. They are difficulty in gaining respondent cooperation, difficulty in getting access to the respondent, and difficulty in hiring and maintaining the workforce we need to carry out our field work. These three factors may be present individually in an area, or they may act together to make our job extremely difficult. They contribute to our problems with differential undercount.

The first aspect of HTE, difficulty in gaining respondent cooperation, refers to lack

of respondent motivation to mail back the questionnaire, to complete all questionnaire items, to give accurate answers and to respond during follow-up operations. It is manifested in a relatively low mail return rate, a high edit failure rate, and refusals to cooperate with enumerators conducting nonresponse follow-up.

The second aspect of HTE, difficulty in getting access to the respondent, refers to various types of barriers that prevent or reduce our contact with the respondent. It includes physical barriers such as impassable roads and locked gates in rural areas, and apartment security measures that are common in urban areas. It also includes invisible barriers such as lack of English language proficiency and a high proportion of one-person households where the availability of the respondent may be extremely limited.

The third aspect of HTE, difficulty in hiring and maintaining the field workforce, complicates our ability to deal with the other two aspects. In those very areas where we tend to have the most problems in conducting enumeration and where we need the largest field staffs, we frequently also face the greatest challenge in attracting a large enough pool of qualified applicants, in training them to carry out their tasks, and in retaining them long enough to complete our work.

3. DATA SET

A Census Bureau Task Force on Hard-to-Enumerate Areas, consisting of staff from both research and operations functions, assembled a "shopping list" of information that should be studied to determine its usefulness in describing the three aspects of hard-to-enumerate areas. Census variables are available for a myriad of socioeconomic characteristics at several levels of geography: county, city, and tract. The list includes data on minority population size and concentration, income, educational attainment, housing, occupation, language spoken at home, unemployment, household size and composition, and other demographic variables. In addition, we have 1980 Census mail response rates and some "operational" or administrative data related to the conduct of the Census, such as the relative cost and timing of operations, our experience with "last resorts," that is, where minimum information was all that could be obtained, pass/fail rates for the test that was used to screen applicants for enumerator jobs, and information provided by field staff on local conditions, particularly related to access and workforce problems. These latter variables, in general, exist only for 1980 Census collection offices, which varied widely in geographic make-up.

The main drawback to the use of 1980 Census information is its age. These data will not accurately represent areas that have changed greatly during the decade. Although the test

censuses conducted in 1985 and 1986 were very limited in geographic scope, we had hoped to test the planned methodology for identifying hard-to-enumerate areas and to get some indication of how 1985 or 1986 data for the test sites differed from 1980 data. However, this study has not benefited from the test censuses due to the lack of tabulated sample data. In addition, special conditions existing during a test census could bias the operational information collected about the test, so any conclusions drawn may not be applicable to the Census environment.

There are other data sets that are very helpful in updating the information that we have from the Census or in adding to it. For example, we have used city and county-level data from recent editions of the County and City Data Book. Demographic information collected in our current surveys can be extended by synthetic estimation to represent parts of the country not included in the primary sampling units (PSU). Current survey response rates and field staff turnover rates, both available at the PSU level, add to our knowledge of the motivational and workforce aspects of field conditions that make an area hard-to-enumerate.

4. METHODOLOGY

Our methodological approach has been evolving over the past year and is being refined as we work with a specific application--the targeting of areas for the Census Community Awareness Program--that will be described in the next section. In general, our task has been first to employ multiple linear regression to identify those variables that contribute the most to the variability in the criterion variable, for example, mail response rates or costs for field work.

Cities' (or whatever geographic entity is being studied) values on the variables are compared to the national mean and are assigned a "score" depending on their distance from the mean as measured by the standard deviation. Cities that rank high (or low, depending on the direction of the variable associated with the condition of hard-to-enumerate) on most of the relevant variables are grouped as "HTE," while others are grouped as "Possible HTE," or "Not HTE."

Classification analysis is used to verify the accuracy of the groupings. Based on the variables found to best differentiate among cities, the original groupings of "HTE," "Possible HTE," and "Not HTE" are tested against the predicted groupings produced by the model. This analysis shows not only the variables that are most effective in identifying hard-to-enumerate areas, but also shows probabilities that allow one to assess the strength of a particular city's inclusion in a group relative to others.

By accumulating the variables that are associated with the condition of HTE, it is possible not only to know that an area fits that general description, but also why. This allows local managers to react to the designation when considering the current status

of the area. They can then advise us about whether the HTE designation is accurate or not due to changes during the decade or special local conditions. The designation of HTE areas is a negotiated process, with the statistical analysis forming the basis and the local knowledge and experience of our managers refining what the numbers revealed.

5. FOCUSING COMMUNITY AWARENESS EFFORTS: A PILOT APPLICATION

Of the three aspects of hard-to-enumerate areas (or groups) described earlier, the motivational aspect is pivotal to increased public awareness of, and participation in, the 1990 Decennial Census. Census Bureau plans include an array of programs to address the need to promote Census awareness. Nationwide publicity may not penetrate all segments of the population evenly, however, nor can we be certain that its messages will be equally effective among those it does reach. Accordingly, the Census Community Awareness Program (CCAP) will provide intensive, personal outreach to communities where resistance to these general messages may be most intense.

To implement CCAP components, we will again look to a decentralized staff of Census Community Awareness Specialists (CCASs). This staff numbered almost 200 in 1980, and will be larger in 1990. Among other activities, CCASs will identify and work through grass roots community organizations, local opinion makers, minority radio and other media, schools, religious organizations, and other contacts to foster communication and awareness. From what specific communities should we seek to recruit this staff? Just where should their activities be concentrated? In short, given finite CCAP resources, how can we apply them so as to achieve the most favorable return on their efforts?

Our primary goal for the 1990 CCAP is to reduce still further the differential undercount of certain minority groups. In focusing CCAP resources, then, we must look first to areas encompassing significant minority populations. Census data for minority populations are readily available, and in fact served as the basis for the delineation of so-called "target areas" for the 1980 CCAS forerunners. Population data alone, however, proved an imperfect tool for the CCAP field manager to use in allocating scarce resources on a day-to-day basis. These choices were often made based on community knowledge and other experiential factors.

We will continue to rely heavily on the accumulated field and community experience of senior CCAP staff in focusing community awareness efforts in 1990. At the same time, we believe that we can inform this process with a more useful statistical frame of reference --and one which differentiates more effectively among geographic areas--than simple population counts or density. Our early attempts to do so (described below) provide a pilot application for the broader HTE project.

Our initial data set comprises the 957 cities with populations of 25,000 or more listed in the County and City Data Book, 1983.

Taking 20 socioeconomic variables 2/, we compute the population standard deviation, and rank each city according to its differentiation from the mean on each variable. Using factor analysis, we determine whether high or low values of a variable are associated with the hard-to-enumerate criterion. We next assess each variable for each city, assign a score of 1 for each variable that exceeds the threshold, and sum the scores to group the cities as either "HTE", "Possible", or "Not-HTE." Those cities scoring the highest, we hypothesize, necessarily include HTE areas where we should focus intensive outreach efforts.

At this time, we are attempting to refine this first-cut analysis in several ways. First, we utilized stepwise discriminant analysis to identify those variables among the 20 which performed best as predictors in that they contribute most to the differentiation among cities in the data set. Nine variables emerged from this procedure: minority population (Black; American Indian, Eskimo and Aleut; Asian and Pacific Islander; Spanish origin), unemployment, median household income, household income (\$10-19,000), families with no workers, and children below poverty level. Based solely on these nine variables, we again grouped the cities, used classification analysis to verify the preliminary HTE designations, and found them to be the same in 81% of the cases. Further study indicates that adding variables to the analysis does not improve overall classification, but may affect an individual city's migration from "HTE" to another group (or the reverse).

Second, we plan to address some of the obvious problems with a data set which entirely ignores rural areas, and the concentrations of minority population--for example, Native Americans--to be found there. In this connection, we will analyze county data to identify the most important variables, eliminate those counties containing HTE cities, and group the remaining counties as HTE, Possible or Not-HTE. This undertaking may highlight groups of nonurban counties which might also warrant the attention of CCAP personnel and will inform us about how applicable a model fit on large cities is to rural areas.

Lastly, we cannot overlook the experiential factors. In this connection, we are asking field managers and CCASs to bring their years of community knowledge to a review of the preliminary HTE classifications. Their comments will help repair some of the problems noted previously, and also no doubt surface other information with which we can further refine the HTE analysis. Hopefully, through successive iterations of this analysis-comment cycle, we will derive a statistical model which increasingly approximates operational "reality" as perceived by those CCASs who carry out this critical awareness function in 1990.

Because we have not yet concluded this process, it is appropriate to illustrate our methodology by means of a hypothetical example.

EXAMPLE

CITY	ORIGINAL GROUP	RESULTS OF CLASSIFICATION ANALYSIS				
		HTE- CLASS-1	HTE- CLASS-2	PROB 1	PROB 2	PROB 3
A	1	2	1	.0973	.8067	.0960
B	1	1	2	.8562	.1438	.0000
C	3	2	3	.0456	.5891	.3653
D	3	3	2	.0298	.3793	.5909
E	2	2	1	.1486	.8510	.0004

NOTE: 1=HTE, 2=Possible HTE, 3=Not HTE

Our original grouping of the five cities shown above was determined by the scores the cities accumulated using intervals around the national mean to identify relatively large or small values. City A's values, for example, were significantly different from the mean on 17 of the 20 variables while City C accumulated only 7 HTE variables, hence their original groupings of 1 (HTE) and 3 (Not HTE), respectively. The classification analysis showed that City A has the highest probability of being correctly classified in Group 2 (possible HTE) rather than 1 while City C should also be in Group 2 rather than 3. Our Field managers will review these classification results, supplementing the probabilities with personal experience to help them make decisions about the final group placement of individual cities.

So far we have described a process useful primarily at a headquarters or "macro" level. Identifying potential HTE areas, together with some idea of their relative difficulty, is essential in planning the nationwide CCAP program and allocating resources among the Census Bureau's regional offices. To support field or "micro" level decision-making, however, the CCAP coordinator needs to know what areas within, say, Chicago should be given priority. We will attempt to meet this need later in 1987 by applying a refined HTE analysis to data at the Census tract level for at least the larger urban areas. With this analysis in hand, the CCAP coordinator will have a sound point of departure in targeting program resources and determining the appropriate mix of program activities to implement in that "target area."

This pilot application, then, will have tangible short-term products for at least one aspect of 1990 operations planning. It will also be invaluable as a constant in future evaluations of the CCAP: by regressing 1990 mail return rates, recruiting experience, and other variables against the model, we will be able to determine its value as a predictor and--to some extent--our success in focusing effort in the right places.

6. FUTURE DIRECTIONS

6.1 Improving the Data

Our problems in obtaining relevant operational data for criterion measures of

hard-to-enumerate areas have highlighted the need for the Census Bureau to collect better and more detailed administrative data (such as costs at a low level of geography) on 1990 operations. This need is recognized in other areas of the Census Bureau concerned with budget planning and with developing the management information systems we will require for 1990. As a result, systems and procedures have been established to collect information that we need for more effective planning of the year 2000 Census, and the HTE project will benefit directly from this in the next decade.

In addition, we plan to begin working with data we have from current surveys, particularly the Current Population Survey which has a state, rather than a national, sample. Demographic data from surveys will allow us to update some of the variables we are using from the Census, making them more current. We need to study the areas of the country where we have the most difficulty with survey response and with staff turnover to learn how these areas correspond to the areas our analysis shows are the most difficult in the Census. This is a confirmatory analysis in the sense that many of the problems associated with field work are common to both the Census and to sample surveys.

6.2 Modelling

The ultimate goal of this project is to build an optimal allocation model that will help us to plan more effectively for the staff and funding required to take future Censuses. If we know each area's rank on the HTE/Not-HTE continuum and the reasons for the rank, we can allocate our resources to best accomplish the task, whether it is hiring CCASs to work at the local level to increase awareness of and participation in the Census or deciding how many collection offices we need, where they should be located, and how they should function. At the "micro" level, we will be able to plan more realistic schedules for field operations, we may adjust our hiring strategies by area, and we may employ a more fine-tuned set of field operations.

While the principal payoff from HTE modelling is expected to be found in Decennial Census resource allocation decisions, we also anticipate some applications to current survey planning. For example, as part of a broader review of the Census Bureau's field structure, we are currently evaluating the costs and benefits of experimental "satellite offices" which situate field supervisors closer to data collection problems. Should this experience prove successful, HTE analysis may well help determine where such offices--or simply modified management practices--are appropriate.

7. SUMMARY

The project described here, the identification of areas that are hard-to-enumerate and the use of this information to plan the optimal allocation of our resources, will become an integral part of future Census planning. For the 1990 Census, we have developed a preliminary application of the methodology which will help us decide where our staff of Census Community Awareness Specialists should focus their efforts to improve public participation in the Census.

Future work will concentrate on examining data at the Census tract level, on improving the data set by updating some variables and adding other information from current surveys, on refining the methodology used to identify HTE areas, and on using this knowledge to build optimal allocation models for use in future Census and survey planning.

REFERENCES

- Beach, Mary Ellen (1986), "Status Report on Hard-to-Enumerate/Optimal Allocation of Resources Research," unpublished memorandum, Washington, D.C.: U.S. Bureau of the Census
- Bruce, Antonio (1987), "Identifying Hard-to-Enumerate Areas: Preliminary Research and Evaluation," unpublished memorandum, Washington, D.C.: U.S. Bureau of the Census
- County and City Data Book (1983), Washington, D.C.: U.S. Bureau of the Census
- Diffendal, Gregg J., Heckert, N. Alan, and Moore, Roger H. (1984), "Characterizing the 1980 Census District Offices: An Application of Exploratory Data Analysis," unpublished report, Washington, D.C.: U.S. Bureau of the Census
- Hill, William F. (1983) "Data Collection in New York Regional Office," unpublished memorandum to Lawrence T. Love, Washington, D.C.: U.S. Bureau of the Census
- Hill, William F. (1983), "Draft of Hard-to-Enumerate Areas Report," unpublished memorandum to Carmina Young, Washington, D.C.: U.S. Bureau of the Census
- Johnson, Robert (1983), "Change in Ethnic and Racial Residential Patterns in Central Los Angeles County Test Census Sites, 1970 - 1980," unpublished memorandum to Henry Woltman, Washington, D.C.: U.S. Bureau of the Census
- Kleinbaum, David G., and Kupper, Lawrence L. (1978), Applied Regression Analysis and Other Multivariate Methods, Boston, MA: Duxbury Press

This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

1/ The authors wish to acknowledge the contributions of Richard F. Blass, Al Mirabal, Sandra Lucas, Keith Holt, Patricia Fuellhart, George Dailey, Sara-Kathryn McDonald, and members of the Task Force on Hard-to-Enumerate Areas, U.S. Bureau of the Census.

2/ Black population, American Indian population, Asian/Pacific Islander population, population of Spanish origin, total educational attainment, persons 25 years old and over with high school education, persons 25 years old and over with college education, persons unemployed, families with no workers, median household income, median family income, family households, households with a female head, families below poverty level, families with a female head below poverty level, persons below poverty level, children under 18 below poverty level, household income < \$10,000, household income \$10 - 19,000, and number of crime incidents.