ANALYSIS OF INNER CITY CENSUS COVERAGE USING LOCAL HOSPITAL ADMINISTRATIVE RECORDS

Margaret S. Boone and David C. Whitford

INTRODUCTION

This report summarizes findings from a study during the past year of 1980 census coverage in samples of inner city Black residents. Samples were drawn from the administrative records of a large northeastern city's only public hospital. The use of medical records to develop rosters of individuals for census coverage improvement research has not been previously tried, although similar sources such as drivers' licenses or Internal Revenue Service records have been used. The goal of this research project was to understand better the types of individuals who may be more likely to be missed by 1980 census mail-back and field procedures. The study focused on inner Black samples because evaluation of census coverage of the population and demographic analysis suggest that the undercount rate in this population segment (or in sub-groups of it) is higher than in the general population. The purpose of this and many other research projects sponsored by the Census Bureau is to learn more about how to provide the best coverage possible in all future decennial censuses of the American population.

History and Development of This Census Survey Project. This research is a follow-up to a National Science Foundation sponsored 1979-80 health study on the correlates of low birthweight deliveries among Black mothers admitted to a public, inner city hospital. The purpose of the original study was to focus closely on a group whose reproductive health was strongly affected by poverty and the social, medical, and psychological ills that accompany it. The goal was to determine some specific correlates of very low birthweight, and thus identify factors which could alert health care providers to "high risk" pregnancies. Results of the health study suggested that women in all samples were poor relative to other urban population segments, and that the sample of women having very low weight infants was especially "disadvantaged," suffering not only from low income, but from the effects of poor general health, substance abuse, and inadequate nutrition. A follow-up census coverage project was proposed in order to further investgate samples of individuals who were known to be "disadvantaged" -- that is, evidencing more than simply a low level of income but also social and health disabilities that accompany a low socioeconomic level. The project attempted to match these disadvantaged sample persons to the 1980 census forms using hospital administrative records. Interviews--which were attempted with all sample persons--resulted in still other addresses and household rosters which aided the clerical match process.

The Utility of Inner City Hospital Administrative Records. Medical records of individuals who use an inner city hospital were especially useful in tracing sample persons because:

First, the hospital is used much like a "family doctor," and individuals came routinely for both inpatient and outpatient care over the space of several years surrounding Census Day 1980.

Second, there are many forms in medical records from which to note, compare, and transcribe addresses and telephones with great accuracy.

Third, addresses and telephones for relatives are available from the notations of "responsible persons" listed upon admissions to the hospital.

Cautions in Using Hospital Administrative Records. While census coverage research stands to gain much from the use of a large health care institution to obtain names and addresses, certain cautions should be noted:

First, no address listed in the medical record is identified specifically as the address where the individual was residing on April 1, 1980. The use of administrative records must make this assumption a priori.

Second, this research strategy cannot assume to cover those who do not use a public hospital either because they are "too poor" (for example, vagrants), or "too rich" (those who use private facilities).

Third, the logistics of large-scale surveys using records from many, widely dispersed hospitals would be prohibitive unless special and unusual measures were available. In the present study a lengthy, thorough, and appropriate procedure for permission to use medical records was required.

RESEARCH DESIGN AND METHODS

Specific Research Questions and Strategies. With these advantages and disadvantages in mind, a research design was developed to answer the following questions:

Question #1. What is the rate of census coverage of the original Black female sample persons, and is this rate higher, lower, or relatively equal to rates for other American groups? Furthermore, are hospital administrative records useful in census coverage improvement research in identifying individual who are missed? To answer these questions, addresses were transcribed for the sample persons and any known relatives, and searched in the records of the 1980 census. A final set of "match codes" assigned a status of "match." That is, the sample person was matched to the 1980 census records, non-match, unknown, or unable to classify for each sample person.

Question #2. How accurately do sample persons or their proxies recall detailed household rosters three and a half years after Census Day 1980? The accurate recall of co-resident individuals is an important issue, because all present and future census coverage research depends upon it. This strategy also allowed an analysis of the completeness of within-household coverage. To complete this analysis, all sample persons were given field or telephone interviews to determine detailed household rosters of all individuals co-resident with the sample persons on April 1, 1980. The rosters were compared in a painstaking, person-by-person routine, and coded either: On the census roster only; on the interview roster only; or on both.

Question #3. How can the individuals who were not covered by the census be distinguished from those who were covered? The comparison of matched and not matched individuals was based on data from the original health study as well as demographic data collected in this research project.

Random and Selective Aspects of the Samples. Samples were drawn from the population of women delivering infants in 1977-78, and were selected according to the birthweight of the infant. With-

in birthweight groupings, random, stratified samples of women were drawn for comparative analysis. However, the samples used in this survey are by no means random with respect to urban Blacks. Sample persons were preselected according to the folfowing factors: (1) sex--all sample persons are females in their reproductive years; (2) hospital --all women delivered at the only public, acute care facility in the city; and (3) race and residence--all were identified from hospital records as Blacks who resided within the boundaries of the city. Because of the non-random aspects of the samples, great care must be taken in drawing conclusions. Results cannot profile the American Black population as a whole, nor all of the urban and/or rural poor. However, they can be useful in improving our knowledge of a population segment which has been hypothesized to be difficult to enumerate in a national census.

Sample Size and "Rank." Samples were drawn from the original hospital population of women with infants of various weights. Results from the health study suggested that women could be ranked according to their degree of disadvantaged status, from women with very low weight infants, to women with low weight infants, to women with normal weight infants (who were the least disadvantaged of all). This implied ranking was used both in the drawing of samples and in the data analysis. The two original health study samples of 105 women each (minus 1 individual who appeared twice in a single sample) were supplemented for the census coverage project with two additional samples of 120 each, making a grand total of 449 individuals in the study. The original stratification scheme based on month of delivery and sex of infant was maintained.

Data Collection and Analysis. Data were collected from several sources: the original health-related records from the first study; a new, month-long data collection phase at the inner city hospital; a two-month interview phase in which attempts were made to contact all sample persons and inquire as to location and identities of co-residents in 1980; and, a census match and data capture phase which lasted four months this past Spring. Data were keyed and transmitted to the main Census Bureau for analysis in early Summer 1984. This report constitutes therefore a timely but still preliminary analysis of these data.

The data analysis employs several methods, including: (1) bivariate analysis using chi-square tests and regression models, in which match status is analyzed according to various health, social, and demographic factors, (2) simple tabulation of household relationship types and analysis of their distribution and match status, and (3) analysis of distribution maps that picture the residential distribution of sample persons by census tract and various socio-economic and health indices.

RESULTS

Rate of Census Coverage Match procedures showed that slightly more than half of the sample persons --52 percent--in this study were matched to census records. This rate is lower than for groups in previous studies. The Non-match rate was slightly less than 25 percent, and the rate of "Unknown" and "Unable to Classify" cases together totalled more than 25 percent. In general, results suggest that these Black female sample persons were not well covered by 1980 census mail-back and field procedures. However, it should be remembered that the study took place three and a half years after Census Day 1980. Results show that the ability to match sample persons to the census records was strongly affected by the identity of respondent. If the sample person could be located for an interview, there was a greater likelihood of a match to census records. If the respondent was the sample person, she matched to the census 69 percent of the time, compared to 52 percent of the time for all sample persons.

Furthermore, there was no indication that the rate of coverage varied a great deal from one sample of women to another. The severely disadvantage women (from the first health study) were matched to census records at about the same rate as other women in other samples. While the coverage rate was indeed in the direction expected (with women having normal weight infants and fewer disabilities matched to the census at a higher rate than the severely disadvantaged women with low weight infants), the differences were not significant. Whatever the factors that reduced the coverage rate of these samples to approximately one-half, they appear to have--from the measures devised for this study--a uniform effect throughout the entire group of 449 sample persons.

Match Status by Social and Health Indices. A number of factors and indices were used to understand better the characteristics of those who were missed by the census. For the analysis, chi-square tests were used to compare these factors to Match and Non-match classifications of cases, omitting the Unknowns and Unable to Classify.

An Index of Disadvantaged-ness was constructed from variables that were significantly related to very low birthweight in the first study. These variables were examined by simple summation and with no weighing. Results for the total Index and for each of its component parts -- smoking, alcoholism, residence in a particular area of the city, migrant status, lack of prenatal care, and and hypertension history--show that Match and Non-match cases do not vary significant with respect to any of these variables. Results suggest that while all of these variables significantly distinguished women as high health risks, the factors are not useful in distinguihing those who may be more likely to be missed by the census.

Similarly, the distribution of scores on an Index of the Stability of Social Support--constructed on the basis of the constancy with which sample persons listed a "responsible person" on successive admissions to the hospitaltal--was not significantly different for individuals who were matched and not matched to the 1980 census.

However, the spatial distribution of social support scores in the urban landscape shows a grouping similar to the one for cases with high scores on the Disadvantaged Index, and, for women with very low birthweight deliveries, on whose characteristics the index was con-

structed. Furthermore, the distribution of all census non-match cases shows a similar concentration of cases in a particular area of the city. These distributions together suggest the existence of a general area where disadvantageous health and social factors coincide, and perhaps indicate an area more difficult to enumerate.

Within these samples of inner city Black residents, socioeconomic status may be too "simple" an explanation of either health status or coverage by the census. Results from both the original health study and this census coverage project show that some variables which are routinely found to be correlated to health and social status in the general population do not "work well" among individuals with many severe disabilities. For example, both education and marital status--two widely used correlates of low birth-weight in the general population -- failed to distinguish women with problem pregnancies in this hospital service population. Education both of the mother and of the father (two variables from the original study) also failed to distinguish Match and Non-match cases.

However, marital status did distinguish Match and Non-match cases. The rate of census coverage was significantly different (= .05) for those sample persons who had ever been married vs. those who had never been married, with ever-married individuals covered better by the 1980 census. This finding is especially interesting in light of the fact that 75 percent of the women in all of the health study samples were not married at the time of delivery. The data collected during this census coverage project showed that most women had not married three to four years later at the time of the 1980 census. Therefore, in this inner city group where the rate of marriage is relatively low, marriage does serve as a good indicator of likelihood of being enumerated.

Finally, the likelihood of being enumerated is also reflected in the likelihood of the sample person's being found in this follow-up survey. There was a significant difference in match status depending on whether the respondent in the interview was the sample person, herself, or an acceptable proxy. (Within specific guideline, proxy interviews were freely taken because of the assumed high level of knowledge about friends and relatives in an inner city Black community.)

Inner City Household Composition. A detailed summary of census household rosters and household rosters taken during an interview three and a half years later gave some indication of the relationships of coresident individuals on April 1, 1980, and which people were best covered in the within-household rosters on census records. Chi-square tests were performed to determine whether there were significant differences between types of relatives and/or acquaintances with respect to the respondent's ability to recall them. Types of relationships were compared according to whether they appeared on both the census and interview rosters, or appeared on only one roster (in which case there was an assumed "error" in memory). Results quite unexpectedly showed there was no significant difference in error rate for remembering coresident male relatives vs. female relatives. However, there were significant differences (= .01) in the error rates for remembering older- vs. same- vs. younger generation household members. Members of the younger generation were recalled significantly less well. Finally, other chi-square tests indicated that members of the sample person's immediate family were recalled significantly better than more more distant relatives and/or non-family members.

DISCUSSION

Results from this study suggest that: (1) Using inner city hospital records of multiple addresses, telephone numbers, and relatives, it is possible to determine the match status of three-quarters of the sample of disadvantaged, inner city Blacks. While medical records are the source of detailed, varied, and longitudinal information, there were two problems associated with this specific study. First, the permission and data collection phases for this type of research were lengthy, involved, and not now feasible on a widespread scale. Second, interviews which can lead to more complete matching to the 1980 census records were dependent upon memory several years after Census Day.

(2) Of all disadvantaged Black sample persons in this study, approximately one-half were matched to census records; one-quarter were not matched; and match status for one-quarter could not be determined. This match rate is much lower than for groups in most other census evaluation research.

(3) The best demographic indicator for the identity of sample persons who may be more often missed by the census in this type of community is marital status. Inner city Black females in their reproductive years who have never been married are more likely to be missed. It should be noted that this particular result implies nothing about the coverage of Black males.

(4) On the other hand, the best demographic indicators for persons who are not reported correctly within a household (either forgotten in the interview or on the census) are generation and distance in relation to the main family unit. Persons in the youngest household generation and persons who are not immediate relatives are more likely to be misremembered.

(5) Unexpectedly, because of previously estimated undercount rates of 10-15 percent for adult Black males, there was no evidence that inner city Black male household members are remembered with less accuracy than females. Therefore, in evaluating reasons for high undercount rates for inner city Black males, there is reason to support explanations other than faulty memory. Two other general types of explanations for high undercount rates among Black males should be further considered. These explanations are: (a) different cultural definitions of the concepts of "household," "residence," and "family;" (b) desire to hide the identity of some household members because of supplemental income program requirements.

(6) Most socioeconomic indicators used in this study cannot be used unambiguously to determine who may be more likely to be missed by the census. As in the original health study, socioeconomic status-as measured by education or a summary index of significant variables from the health study--

gives no simple key to who may be more likely to be missed by the census in this type of community.

(7) A map for non-matched cases does, however, mirror the geographic distribution of very low birthweight deliveries in the health study, and to a certain extent, scores for low social support. If one part of the hospital service area could be identified as "more difficult to enumerate," it would be this area. Because of the similarities in the distribution of a particularly severe inner city health problem (i.e., low birthweight and infant mortality), scores for low social support and non-match cases, it is suggested that part of any explanation of failure to enumerate lies in the nature of the population, or the nature of this population in interaction with census enumeration efforts.

(8) Finally, the quality of recall data in this type of census evaluation research is dependent upon locating and interviewing the sample person rather than a proxy. Recall of household members was better when the respondent was the sample person. Use of proxies, even in communities where "accountability is high," i.e., where people tend to know the details of their friends' and relatives' lives, should be weighed very carefully.

RECOMMENDATIONS

Recommendations for future census evaluation research and improvement in the taking of the decennial census grow out of the above conclusions.

First, because the younger generation was misremembered significantly more often, future research and efforts at enumeration (which include construction of question frames, probing technique and printed "examples") may benefit from focusing on children -- where they sleep, where they eat, and who takes care of them on a routine basis.

Second, while it would be difficult to construct an enumeration procedure which assured improved coverage among non-married persons, the indicator of "never married" could be useful in future census coverage research. It is an easily determined demographic charactertistic and while results of the original health study suggest that marital status is not related to low birthweight, it is related to likelihood of being enumerated. This finding again suggests that the focus of coverage improvement in the inner city should be on the young.

Third, in the the assignment and investigation of particular areas as "difficult to enumerate," it may be useful to examine the distribution of other social and health disabilities.

Fourth, in an interview conducted several years after Census Day for the purpose of matching census records, memory of household composition is poor; however, results can be significantly improved if interviews are strictly confined to sample persons.