



University of Pennsylvania
ScholarlyCommons

Departmental Papers (Sociology)

Penn Sociology

1979

What Happened When the Census was Redone: An Analysis of the Recount of 1870 in Philadelphia

Frank F. Furstenberg

University of Pennsylvania, fff@ssc.upenn.edu

Douglas Strong

Albert G. Crawford

Follow this and additional works at: http://repository.upenn.edu/sociology_papers

 Part of the [Demography, Population, and Ecology Commons](#), and the [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#)

Recommended Citation

Furstenberg, Frank, Douglas Strong, and Albert Crawford. 1979. "What Happened When the Census was Redone: An Analysis of the Recount of 1870 in Philadelphia." *Sociology and Social Research* 63 (3): 475-505.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/sociology_papers/23

For more information, please contact repository@pobox.upenn.edu.

What Happened When the Census was Redone: An Analysis of the Recount of 1870 in Philadelphia

Disciplines

Demography, Population, and Ecology | Quantitative, Qualitative, Comparative, and Historical Methodologies | Sociology

WHAT HAPPENED WHEN THE CENSUS WAS RE-
DONE:
AN ANALYSIS OF THE RECOUNT
OF 1870 IN PHILADELPHIA*

Frank F. Furstenberg, Jr.
Douglas Strong
Albert G. Crawford
University of Pennsylvania

The authors wish to thank the research associates of the PSHP, in particular Ted Hershberg, Jeff Seaman, and Henry Williams, for their collaboration and sensitive reading of several drafts of this essay. In addition, we are grateful to the editors, Barbara Laslett and Jim Smith, and an anonymous reviewer for their helpful comments. Finally, we would like to acknowledge the following coding assistants for their careful attention to detail in the compiling of the data: Cynthia Brown, Caryn Fine, Shirley Guerra, Donna Robinson, Chris Walbash.

INTRODUCTION

In their exuberance at discovering untapped records of past behavior, historical researchers have not always been as critical of quantitative data sources as they might be. While professionally predisposed to cast a jaundiced eye toward contemporary opinion surveys, historians have sometimes been willing to admit as evidence statistical accounts that would not satisfy the methodological standards of even the most casual survey researcher today. In particular, while other archival materials have been subjected to external validation, federal, state, and local census materials have largely remained exempt from critical scrutiny (Sharpless and Shortridge, 1975). It has been more or less assumed that the quality of these records was ensured by adherence to careful procedures, as is the case when the census is con-

*This research was supported by the Philadelphia Social History Project, Theodore Hershberg, Director. The PSHP gratefully acknowledges the funding it receives from the following federal agencies: the Center for Studies of Metropolitan Problems, National Institute of Mental Health (MH 16621); the Sociology Program, Division of Social Sciences, National Science Foundation (SOC 76-20069); the Division of Research Grants, National Endowment for the Humanities (RO 32485-78-1612); and the Center for Population Research, National Institute for Child Health and Human Development (RO 1 HD 12413).

ducted today. In point of fact, however, even modern census techniques are highly imperfect, suffering from the inherent weaknesses of face-to-face interviews (Webb, et al., 1966; Phillips, 1972). Virtually all of the problems of the contemporary social survey existed in an exaggerated form in the census records collected in the nineteenth century (Heer, 1968).

The preface to the Ninth Census of the United States, written by Superintendent Francis A. Walker (1870), provides a sobering introduction to a host of problems. While Walker staunchly defended the integrity and conscientiousness of the census enumerators and the rigor of their procedures, he openly acknowledged that "the statistics of the census are not of uniform value" (Ninth Census, Vol. 1: 1vii). In his compendium of sources of error, Walker was most attentive to the problems of underenumeration occasioned by the lengthy period of data collection, what Walker termed "the essential viciousness of a protracted enumeration." He also critically commented on the accuracy of selected information such as occupational title, the incidence of marriage, age, and literacy, owing to vague question wording or to reticence on the part of the respondent. The most disquieting procedural defect was mentioned almost in passing by Walker in his plea for adopting self-administered schedules:

Under the present system the assistant marshal calls upon families in the course of his rounds at hours when the head of the family is generally absent from home. Some, at least, of the inquiries of the census, especially those relating to industry and wealth, are such as but few women are expected to be prepared to answer. The probabilities are, moreover, that in at least one case in six or seven the wife is also absent. The duty of answering inquiries, therefore, devolves upon servants and children, who are naturally incapable or unprepared to give full and correct answers (xxvii).

Contemporary survey researchers have produced compelling evidence that household members, even heads and spouses, are not always able to supply accurate information on other family members. Naturally, the level of error varies with the type of information sought and the relationship between the informant and other household members. It is difficult to believe that family members were better equipped to supply personal information in the nineteenth century than they are today, especially when families were larger and households often contained boarders, lodgers, and servants.

In the most careful review of the historical literature dealing with

census error, Sharpless and Shortridge (1975) identify only a small number of existing methodological assessments. While their survey is largely confined to the problem of underenumeration, their conclusions point to the need for a more general re-examination of the accuracy of census materials:

Clearly, a concerted effort must be made by historians to analyze the manuscript census systematically in order to determine both the dimensions as well as degrees of bias due to underenumeration . . . We need to measure the limitations of our data with exactitude and more rigorously to control for their effects, or we will not be able to discern a fact of interest from the artifact of the data collection process (434).

A vivid example of the potential dangers of ignoring measurement and sampling error can be seen in the literature on social and geographical mobility in the nineteenth century. A number of researchers have relied on census data for estimating inter- and intragenerational change in socio-economic status. If census data are unreliable, this means mobility figures are lumping together error and real change, resulting in a serious miscalculation of the true amount of social movement in past times. Moreover, if unreliability systematically varies among segments of the population, then contrasts between subgroups become highly questionable. (For a more extensive discussion of this problem, see the *Historical Methods Newsletter*, 1976).

Much of the reason for the absence of studies evaluating the accuracy of past censuses is due to the difficulty of developing techniques that effectively address the issues involved. One method of verifying census accuracy is to design techniques that measure the internal consistency of the census data itself. Sharpless and Shortridge, for example, review the studies on age heaping ("The tendency for ages to be reported so that the proportion of persons in some age groups is over-stated and the proportion in other understated") and find that error in reporting age varied widely by subgroups of the populations. He would be unwise, however, to judge the quality of the census based on this variable alone, because age may not have been accorded the same importance in the nineteenth century as it is today.

Another powerful method of checking census accuracy is to compare the census manuscripts on a case-by-case basis with other comparable population lists. Ideally, these lists should have been compiled at nearly the same time as the census in order to minimize the effects of actual change and to identify error more clearly. Sources such as city directories and vital statistics records provide an opportunity to check the accuracy of census enumerations, but the problems of linking

cases are monumental. Indeed the difficulty of locating a household in two population lists is almost severe enough to require that a common address be the primary linking criterion.

This paper reports the results of a systematic examination of a data set peculiarly well suited to avoid the typical problems of linkage. By a rather uncommon turn of events, the Ninth Decennial Census, the 1870 Census, was done twice in Philadelphia, providing an opportunity to investigate the congruence between the first and second enumerations — in effect a classic test and retest design. Our report begins with a description of the circumstances which led to the duplication of the census in Philadelphia and continues with a review of the procedures that we developed to perform an analysis of the two 1870 censuses. Results are then presented concerning the accuracy of the census data. In particular, we shall examine both the problem of underenumeration and the reliability of selected information on household composition. Finally, we illustrate briefly how the data can supply important substantive information about change in household composition over time.

The Philadelphia Census of 1870

Many cities anticipated that the 1870 census would confirm a pattern of spectacular growth over the decade of the 1860's. Their expectations, of course, were fueled by a desire for increased political representation in Washington as well as by intense local pride. As General Walker, the Superintendent of the Census, wrote in his introduction to the statistical compendium of the Ninth Census (Ninth Census: xx):

The internal changes of the United States for the last ten years have been so fierce and rapid as to put calculations at defiance. In the absence of definite information estimates as to the growth of cities and states soon became wild and extravagant . . . Claims that perhaps were first made in a spirit of banter soon are taken as serious, and in the event people become angry to find that not true which was originally asserted only to irritate a rival.

Philadelphia, apparently was no exception.

Based on projections from listings of registered voters, political leaders had calculated that the city would number nearly 900,000. They were staggered when informed that the census count produced a

total of only slightly more than 650,000. Without impugning the methods of the census marshal, the city council petitioned President Grant and Superintendent Walker for a recount, claiming that thousands of residents were vacationing in early June, the period of the first enumeration. While the President deliberated, the City Council authorized funding to conduct a selective census of one ward, which had been reported to have dropped in population. Just before the local survey was to be launched, President Grant authorized a recount; and, several days later, the second census was commenced.

Although the two censuses were conducted under the same auspices, there was a certain amount of procedural variation between the first and second enumerations. In order to speed up the recount, a much larger army of deputies was used in the November re-enumeration. It was hoped that the census might be completed in a single day, but, in fact, it took nearly a week, still roughly half the time of the June census. The newspapers urged full cooperation, reminding the citizens that in the recount, information would not be collected on occupation and property, items which apparently had created some resentment in the original census. In fact, the second enumeration was confined to a listing of each member in the household and a recording of age and sex, but omitting information which had been collected in the earlier census on occupation, wealth, place of birth, and the like.

The census takers were instructed to confine the enumeration to residents who had been living in Philadelphia on or before June 1st, but it is not clear how this determination was made. In any case, the census recount proved to be a disappointment to the city and a vindication of General Walker, who pronounced that the results proved conclusively "that the United States marshals and their assistants overcame, to a remarkable degree, the difficulties attending the census of such a population in the summer months." The census recount in Philadelphia revised the total population upward only by some 16,000 persons, an increase of approximately two and a half percent.

Although the overall figures would seem to bear out General Walker's claim that the recount demonstrated the accuracy of the census, a closer look at the Philadelphia results provides a less reassuring result. While the total counts in the two censuses were remarkably close, the correspondence at the ward level was far from perfect. Of the 28 wards, seven gained more than 5 percent in the recount, including three which increased by 18, 21, and 26 percent, respectively. Moreover, four of the 28 wards experienced significant drops in population during the five month interval. Less interested in the balance sheet than in the bottom line, the city leaders raised no further objections, and the results of the recount became official.

METHODS AND PROCEDURE

Many alternative procedures could have been followed to test the accuracy of the two census enumerations. Conceivably, we might have: 1) traced from the first to the second census or from the second back to the first; 2) selected our population simply at random or stratified according to certain criteria; and, finally, 3) examined either households or individuals as the unit of analysis. Our sampling strategy was to some extent dictated by convenience and resources as well as analytic considerations, but in weighing our options, we fortunately were able to rely on an existing bank of data assembled by the Philadelphia Social History Project (for a description of the PSHP data, see *Historical Methods Newsletter*, 1976).

Two main reasons led us to go from the first to the second enumeration rather than the reverse. First, tracing households back from the second enumeration would have greatly reduced our information on the features of the unlocated households because the census recount collected only rudimentary data on household composition. By contrast, the first census contained the full gamut of demographic and sociological information on the household. Second, we preferred to estimate underenumeration in the recount rather than in the initial enumeration, reasoning that the recount adhered to stricter enumeration procedures. Thus, we hoped to avoid the possibility that our estimate of undercounting might be inflated due to a high level of carelessness in the first enumeration.¹

Since an important objective of this project was to assess the quality of the census taking, we decided to cluster our sample in certain geographical areas. A cluster sample permitted us to investigate whether the veracity of census taking varied by census taker. Since census enumerators were assigned to relatively small geographical areas, a clustered sample is the most efficient way of accumulating enough cases to appraise the work of individual marshals. In addition, a geographically concentrated sample makes it possible to characterize the neighborhood context in which the designated households were located and thus to determine whether context is related to various types of error. Finally, we elected to sample household units, not individuals. Since the second census contains very little information on individuals, we maximized both the methodological and substantive value of the data by defining the household as the unit of analysis.

In drawing the sample, we were able to make use of an ecological mapping of Philadelphia carried out by the Philadelphia Social History Project. The city had been subdivided into grid units approximately a

block by a block and a quarter in length (660 feet by 775 feet), for which socio-economic and demographic data were available. In 1870, there were some 1,400 inhabited grid units. These constituted our sampling frame. In almost all cases, a grid unit was enumerated by a single census taker.

Although the population of Philadelphia was predominantly native born, other ethnic groups, particularly Blacks and Irish and German immigrants (constituting only 3.0, 14.4, and 7.4% of the population respectively) were disproportionately concentrated in a small fraction of grid units. In selecting households for our analysis, we elected to oversample Irish, Germans, and Blacks in order to ensure an adequate representation of each major ethnic group in the sample. To accomplish this sampling plan, we first ranked the grid units by ethnic concentration. We selected the 471 grid units (the top third) for each ethnic group that had the highest proportion of that ethnic group. Thus, the initial stage of our sampling plan stratified the grid units according to ethnic density. We then selected from this subsample only those grid units containing 50 or more persons so as to have enough cases for both household and aggregate analysis. Finally, we systematically sampled from the remaining grid units.²

Within each of the 89 designated grid units, using maps of the city compiled in 1867, a listing was made of all streets that ran through the grid unit. One street was selected at random and located in the census manuscripts (which had been cross-listed by address by coders at the PSHP using the Philadelphia city directories). When the street was located, up to 30 households were sampled consecutively. In the event that the street did not contain that many households, a second street from the grid unit was taken and the process was repeated.

Information was extracted for each designated household on the membership of the household, family structure, presence of sub-families, and selected characteristics of the head of the household such as his or her sex, age, place of birth, ethnicity, and amount of personal and real property. It is important to point out that the information derived pertains primarily to characteristics of households, not attributes of individuals, with the exception of the data coded on the household head. Although it might have been desirable to create a separate set of information on each member of the household, the coding task would have exceeded available resources. Moreover, given the limited information in the second census on the individual members of the household, we decided that household level measures would suit our analytic purposes well enough.

The coding operation, nonetheless, was extremely tedious and time consuming. Once a grid unit was identified and households were

sampled, it was necessary to locate the identified area in the November census. The arrangement of addresses within census enumeration districts did not always follow the same schemes in June as were followed in November. Consequently, it frequently took some time to match residential locations in the two enumerations. Once the match was made, a search for particular households in the second enumeration was commenced. If they did not appear at the address expected, a thorough examination was made of the entire census enumeration district, an area of some 1,000 to 3,000 households, to see if the units or any of their members could be located. In fact, this procedure occasionally turned up the household either because a move had occurred in the interim or, more probably, because the address in one or the other census was incorrect. Typically, however, this time-consuming hunt yielded nothing.

The principal problem was one of location, not identification, of households. Unlike the uncertainty inherent in the linking of individuals based solely on one name, or even on the name and other information about the individual, the linkage process of households leaves little uncertainty. This is so because having the names and ages of other members of the household common to both enumerations markedly increases the certainty of links. Only in those rare instances when the recording of names and ages was wholly inaccurate, or when households were initially very small or experienced a very large turnover, was there any ambiguity in identifying the correct household.

Once the link was established, the coders extracted information on the membership of the household unit in the second census, classifying individuals into predesignated categories.³ Since we were primarily interested in the persistence of household members, a coding scheme was developed to record entrances and exits of: 1) the household head; 2) the spouse of the head; 3) children of the head; 4) individuals with the same last name as the household head; and 5) members of the household with different surnames. Each of these categories was further sub-divided by age and, in the nuclear family, by sex. Excepting household heads and spouses, we cannot trace the movement of particular individuals, but we can specify the turnover of specific age, sex, and relational subgroups within the household. Discrepancies between enumerations in the recording of ages and names of household members were also counted in each age and relational subgroup.

FINDINGS

As a prelude to the presentation of the findings, it is important to

consider the extent to which these data from Philadelphia in 1870 can be generalized to other locales in that same year and to other decennial censuses. While we sometimes speak of the results as bearing directly on other censuses, it is wise to exercise some caution in generalizing from our results.

First, Philadelphia in 1870 was one of the largest urban areas in the nation, and it is likely that the difficulties attendant to census taking were compounded by the size of the area to be enumerated. As only one example, at least before 1880 when census enumerators were required to reside in their enumeration district, it is quite likely that the census takers were unfamiliar with the locales within which they worked. Contrast this to a small town or even a small city, where the enumerators probably were very familiar at least with the neighborhoods, if not with individual household members. The differences in recording error between these two extremes should be significant. Second, the quality of census taking depends a good deal on the diligence of the census taker and the supervision of the process by others in the census bureau. While we have no reason, as yet, to suspect that the Philadelphia census takers were more or less careful than any other group of census takers, the possibility should be acknowledged. Finally, it is hazardous to generalize from one census year to the next, if only because enumeration procedures changed over time. For example, as noted above, census takers were required to reside in their enumeration district beginning in 1880. In addition, the 1870 census was, at the time, reputed to be worse than previous censuses. While the census recount in Philadelphia in November of 1870 generally corroborated the results found in the first enumeration, and while some of the results of this study will show that a good portion of what was recorded was accurate, we still cannot assume that the 1870 census was typical of other enumerations.

Turning to the substance of our analysis, the principal findings have been divided into two sections, which reflect the two types of error that we are measuring: 1) the underenumeration of entire households; and, 2) the unreliability in the data recorded in both census enumerations. We shall discuss, but cannot resolve, the thorny problem of distinguishing underenumeration from geographical mobility. In attempting to sort out these two phenomena, we shall be interested in the distribution of non-located households among the population. Our analysis of the second form of error, the unreliability of the household information, tests the quality of census information. The measures that we have used to estimate this type of error are undercounting of individuals within households and name and age discrepancies for various age and relational subgroups of household members. A final issue address-

sed in the analysis is the degree to which relationships exist among the different types of error.

Household Underenumeration

The principal technique used to estimate underenumeration was a search in the second enumeration for households that were identified in the first enumeration. Strictly speaking, we will be determining the level of undercounting that occurred in the second enumeration. Because the stated purpose of the second enumeration was to provide a full count of the citizens of Philadelphia, it is likely that, if anything, we are providing a low estimate of the level of underenumeration that normally occurred in 19th century census taking. While we cannot, to our satisfaction, separate underenumeration from residential movement, we have devised a crude method of estimating the upper and lower limits of each factor. In addition, by examining the correlates of non-location, we can point to relationships that probably result from geographical mobility and/or systematic errors in the census.

We shall begin by considering the proportion of households in our sample that were located in the second enumeration. Overall, we managed to find 82.2% of the 2,397 households sampled, indicating that the combination of true movement and underenumeration accounts for less than one-fifth of all cases. Even assuming that most of the 17.8% is underenumeration, this result alone suggests that the coverage of the population in historical censuses might be comparable to that of some modern censuses. Moreover, based upon estimates derived from other urban areas in the late nineteenth century, there is good reason to believe that a substantial part of the 17.8% non-location is due to actual movement, and not to error (Thernstrom, 1973; Chudacoff, 1972; Katz et al., 1976).

Using the Philadelphia city directories, we have developed a technique to estimate the extent to which non-location of households was due to underenumeration or due to geographic mobility. The method involves tracing the unlocated household heads in the 1869, 1870 and 1871 city directories. If the household heads were listed in the 1871 city directory as residing on the same street as in the first census enumeration, then it was assumed that the individuals lived on that street during the recount but were overlooked by the census taker. Similarly, if the household heads were located in either the 1869 or 1870 city directories, but were not located at the same address in the

1871 directory, we assumed instead that residential movement actually took place after the initial enumeration.

The results, crude as they are, provide an upper and lower limit of underenumeration and movement. Overall, 23.3% of the non-located heads were found in the 1871 directory, implying a minimum underenumeration of 4.1% for the whole sample. Similarly, 10.9% of the non-located heads were found in the 1869 or 1870 directories, but not in the 1871 directory, yielding a lower-limit estimate of movement of 1.9% for the total sample. The difference between the total non-location rate (17.8%) and the minimum underenumeration rate (4.1%) provides a maximum movement estimate of 13.7%. Similarly, the difference between the non-location rate and the minimum movement rate (1.9%) provides a maximum underenumeration estimate of 15.9%.

It should be noted that fully 65.8% of the non-located heads (11.7% of the total sample) were found in none of the city directories. Several factors may explain our lack of success in tracing household heads in the directories: serious underenumeration in the compilation of city directories; very rapid transience of a subset of the population; and substantial error in the recording of names in either the census or the directories. While it is difficult to distinguish between these three explanations, it is well known that Blacks, the propertyless, the unemployed and other disadvantaged groups were systematically under-represented in city directories.

Taking this bias into account, we derived a different set of estimates, confining our analysis to those individuals who were most likely to appear in the city directories: household heads in non-manual occupations. The results are striking. While the percentage of non-location due to actual movement was slightly less than previously reported (9.3% compared to 10.9%), the percentage of non-locations assumed to be caused by underenumeration rose from 23.3% to 42.6%. Even granting that this segment of the population would be slightly more likely to persist and thus appear in the 1871 directory, it seems that in the entire sample, underenumeration may be the explanation for at least a large minority of total non-locations.

While the overall rate of non-location and the derived estimates of underenumeration provide some basis for assessing the completeness of census data, it is essential to discover the extent of systematic error (the non-random distribution of underenumeration among population subgroups). Unfortunately, at least at this stage of our research, we are unable to make this determination with any degree of certainty.

However, as an initial approach to determining the distribution of error within the population, we have examined the relationship of non-location (error and movement together) with a variety of social and

demographic variables. The pattern of these relationships, when considered in the context of what is known about the correlates of persistence in the nineteenth century and of the nature of the variables themselves, will provide additional insight into likely correlates of error and/or movement.

We examined the relationship of rates of non-location with several types of variables: 1) characteristics of household heads and of households; 2) social and economic characteristics of geographic units; and 3) other types of error within these geographic units. The first two will be examined, in order, in this section, while the last will be considered near the end of the paper.

Table 1 summarizes the relationships between the rate of non-location and various household level variables. While we report only bivariate relationships, we should note that multivariate analyses confirm that the relationships reported are not spurious. Overall, the variables that are related to higher rates of location are similar to what has been reported in previous historical studies of urban mobility, although the relationships in our analysis appear to be somewhat weaker than in the existing literature. Older (presumably more settled) household heads were found more often than their younger counterparts; heads with higher occupational status and more real property were found more often than those holding unskilled jobs and with little or no real property; and households with many non-relatives (usually boarders) were found more often than households with fewer non-relative members.

The rates of location among ethnic groups provided some surprises. First, contrary to our expectations, Black household heads did not experience a significantly lower rate of location (one must remember, of course, that most persons, Blacks especially, were not household heads). We had assumed, incorrectly, that racial prejudice, unfamiliarity of census takers with Black neighborhoods, and the complexity of Black household units would lead to underenumeration of the Blacks. In contrast, the rate of location of German-headed households was somewhat lower than that of other ethnic groups, perhaps indicating a language barrier that impeded the accurate reporting of names and a disinclination on the part of census takers to concern themselves with communication problems when they were pressured to complete the census with dispatch.

Other correlates of location as well can be explained just as plausibly by underenumeration as by actual movement. Thus, it is possible, as Superintendent Walker observed, that underenumeration of households consisting only of married couples can be explained by the increased likelihood that no one was home when the census taker

passed through. Using the same logic, older heads should have been easier to locate than younger heads, the former typically having more children at home and a non-working spouse. Finally, while lower-status families were undoubtedly less residentially stable, they also may have been more difficult to locate due to the absence of servants and boarders in the house. Thus, there is good reason to believe that both underenumeration and actual movement account for the patterns of non-location described in Table 1. Since we expect the effect of these two factors to be cumulative, it is perhaps surprising that the differences observed were not greater. It may be that random error in the two censuses attenuated the strength of the relationships.

In contrast to Table 1, which arrays individual level data, Table 2 displays the correlates of location rates on an aggregate level. With these aggregate data we can analyze the socio-economic and ecological context of underenumeration. Such an analysis is important because both mobility and underenumeration may have been more highly related to the features of neighborhoods than to characteristics of households.⁵ (For example, if a census taker was prone to underenumerate households of a particular ethnic group, he would more likely shirk his duty on the basis of known characteristics of an entire neighborhood rather than unknown ethnic affiliations within particular households.)

Generally, the results in Table 2 reinforce our earlier analysis of household level variables. At the aggregate level, the greater the mean occupational status, amount of real property, and number of non-relatives, the higher the rate of location. Similarly, higher rates of location occurred in native white-dominated areas, while German dominated areas reveal lower rates. The relationship reported between age of head and location on the individual level disappeared at the aggregate level, and the relationship with real property became somewhat weaker on the aggregate level.

The three variables that characterize the size of the street and the density and complexity of the grid unit are our best measures of physical conditions which might affect the quality of census taking. We had expected that rates of location would be lower in densely populated areas with many back streets. Our expectations were not borne out. Since the effect of migration should only increase the correlation between non-location and these features of the environment, it is reasonable to conclude that census takers were as careful in neighborhoods that were difficult to enumerate as they were in other neighborhoods.

In any case, the analysis of the aggregate level data substantiates the results reported earlier. It still remains difficult to distinguish between

the correlates of movement and of underenumeration. One way of addressing this issue is to explore the relationship between rate of location and the other types of error. We shall turn to this question after we have introduced each of the types of error measured within located households.

Unreliability Within Linked Households

As mentioned earlier, little systematic research has been carried out on the reliability of historical census materials owing to the difficulties of linking data sources. Our examination of unreliability is limited to those few variables that are common to both the first and second enumerations and includes analyses of underenumeration of household members, name discrepancies, and age discrepancies. While we are unable to examine the reliability of other census information (e.g. — place of birth, ethnicity, real property, etc.), these unmeasured and often more obscure variables probably contain similar, if not greater, levels of error.

Underenumeration Within Households

In assessing the extent of turnover within the household, we face the by now familiar problem of distinguishing between actual movement (household entrances and exits in this case) and miscounting in either enumeration. The method adopted here is to compare the entrances and exits of different household members according to their age, sex, and relation to the head of the household. Table 3 shows the numbers of exits and of entrances, the exit rate and the exit/entrance ratio, within 14 household subgroups, for households that were located in the second enumeration.

The table reveals several facts about the quality of census taking. First, exit rates correspond to what would be expected if census taking were done with care. Predictably, there was very little turnover among household heads and their spouses. Beyond the death-prone first year of life, the exit rate is lowest for young nuclear children (3.0%) and gradually increases to an exit rate of 11.7% for adult nuclear children, who presumably had reason to leave home to marry and/or to seek employment (Modell, et al. 1976). And, as one would expect, the exit rate for individuals assumed on the basis of their last name to be extended family members was higher than for those in nuclear categories. Higher still are the rates for household members with different surnames (distant relatives, transient boarders and servants).

The very low exit rates for young nuclear children and for the head of household and spouse provide strong evidence of the capacity of the census to yield reliable data. Even if all of the exits were actually the result of underenumeration, the rate of error in the nuclear family as a whole (excluding children 20 or older) is only 3.3%. Clearly, then, at least this part of the census was performed diligently, and data on the composition of the nuclear family can be used with confidence.

The exit rates in the non-nuclear categories pose greater difficulties for historical researchers, partly because we cannot determine whether household turnover was the result of mobility or census taker error. This problem will be examined in greater detail below. In any case, even if half of the exits are due to error, upwards of a fifth of the teenage and adult non-relatives actually left the household within a period of five months, a very rapid turnover that has enormous implications for the study of family life in the nineteenth century.

Some perspective on underenumeration within households can be gained by comparing the ratios of exits to entrances in each category. In some subgroups there is good reason to expect a nearly equal number of exits and entrances, while in others an imbalance might be anticipated. For example, because of relatively high nineteenth-century mortality rates, it makes sense that there should be more exits than entrances of the head of the household and the spouse. If the exits are attributable to death, it would be quite surprising if a large number of remarriages had occurred within the 5 month period. Similarly, because they are in the marriage-prone years, more adult offspring of the head of the household should have exited than entered. The findings confirm these expectations and thus increase further our confidence in the accuracy of information about nuclear family members.

In contrast to the imbalances expected and found in the nuclear family, there is no a priori reason to assume that such a pattern would exist among non-relatives. While the collective knowledge of historians is admittedly limited concerning the comings and goings of boarders and servants, it seems likely that, on an aggregate level, the total number of these individuals in households should remain approximately the same, particularly over such a short period of time. As Table 3 indicates, they did not, except for the adolescent non-relatives, perhaps because such individuals had stable positions as servants or apprentices. Entrances greatly outnumbered exits for young non-relative children and for non-relative adults.

One explanation for the large surplus of entrances over exits is a seasonal influx of boarders and lodgers into the city for the colder months of the year, possibly following farm labor in rural areas during

the warmer months. An equally likely explanation, in view of the circumstances of the re-enumeration, is that many of the individuals who appeared in the recount were actually present but unrecorded in the first census. This interpretation would apply as well to the extended family members, for whom we also find an imbalance of entrances over exits. To the degree that this surplus of entrances is due to undercounting rather than to seasonal migration, it reflects a bias in the first enumeration. If these individuals were less visible to census takers, they will also be less visible to historical researchers. This in turn will have a decided impact on the success of procedures to link non-nuclear family members, thus impeding longitudinal analysis of this neglected segment of the population.

Name Discrepancies

An examination of name discrepancies between the two enumerations provides additional information which has important implications for linkage procedures. Historians have used different record linkage techniques ranging from simple types of hand linkage to complex forms of automated linkage. (Winchester, 1970; Historical Methods Newsletter, 1976). The choice of technique has depended, in part, upon the size of the population to be linked and on the variables available for use in the linkage procedure.

Typically, especially over longer time intervals (10 years or more), these linkage procedures have yielded low persistence rates, leading historians to conclude that geographic mobility was very high. However, there is a largely unexplored possibility that the individuals did not actually move but were listed under a different name, most likely due to census taker error. Our data provide a unique opportunity to assess the accuracy of linking procedures based on name and age alone. We have this capacity because, in addition to each individual's first name, surname, and age, we have, in both enumerations, the additional information of household address and imputed family relationship. Thus, for example, if an individual's household, first name, age and occupation were all similar but the last name was different, it is clearly a case of actual persistence that would be missed by most linkage procedures.

The rule that we have followed in attributing name discrepancies was that a variant form of a first name or a phonetically similar but differently spelled first or last name was not counted as a discrepancy, while names that were substantially different phonetically were. Thus, Edward and Eddie, Daily and Daley, and Smith and Smythe would not be considered as discrepancies; but John and James or Walsh and

Waters would be. In addition, a different number of syllables in a name was considered to be a discrepancy (Lister, Listerberg).

Clearly, identifying these discrepancies is sometimes a subjective process, and there is a gray area between identifying two names as discrepant and classifying them as an exit of one individual and an entrance of another. We believe that our method of coding provides a conservative estimate of name discrepancies between the two enumerations and that the actual level of discrepancy may have been higher (and mobility, in turn, lower). This is particularly true for non-relatives, because without a familial context it was more difficult to conclude that two variant names were actually the same individual.

Table 4 shows the percentage of individuals, by subgroup, whose names (either first or last) were substantially different in the first and second enumerations. Overall, the table indicates that the rate of name discrepancy is 8.4%. Thus, in a linking procedure that is based solely on name and does not include information about address, occupation, or other household members, one individual in twelve would be missed. Once again, error is more prevalent among the non-nuclear members of the household and relatively uncommon within the nuclear family.

It should be recognized that correctly identifying names of individuals may not have been as important to the census takers in the second enumeration as providing a complete count of the population. Other censuses in the past, seen in this regard, may have been more accurate in recording exact names. Two points, however, militate against this argument. First, the level of overall name agreement, especially in the nuclear family, indicates that the enumerators must have been concerned with correctly naming individuals. Second, if this kind of error were more prevalent in the recount than in other censuses, one might expect the error to occur equally across relational subgroupings. In fact, however, nuclear family members had markedly fewer discrepancies than others in the household. It seems far more plausible, then, that the relatively high rate of name discrepancies among non-relatives may be largely explained by the lack of familiarity of the census informant with the specific names of marginal members of the household rather than to carelessness on the part of the marshal.

Age Discrepancies

The last type of error to be examined is age discrepancy. From previous studies, there is good reason to suspect the reliability of age measures in past censuses, and several historians and demographers have provided assessments of the level of age distortion.⁶ Because

many kinds of demographic and social analyses of historical data rest on a precise measure of age, it is of critical importance to understand the extent of unreliability in the recording of age.

Unfortunately, noting the correct age of the population of Philadelphia was not accorded much importance in the second 1870 enumeration, as a cursory examination of the manuscripts indicates. The extent of age heaping is much larger in the second than the first enumeration, perhaps as a result of official policy. Therefore, the absolute levels of error that we report in this section are, for the most part, less meaningful than the relative levels in the various subgroups. Table 5 arrays, by subgroup, the percentage of individuals found in both enumerations who had age discrepancies of 2 or more years and the percentage who had age discrepancies of 5 or more years (note that these two columns are not mutually exclusive and that the former subsumes the latter).

The error rates in the left hand column, reflecting a stricter measure of age discrepancy, are extremely high and clearly overstate what might be expected in other census reports. A margin of error of 4 years or less is not sensitive to the usually high amount of age heaping in the second enumeration. Thus, the rate of age discrepancies of 5 years or more is perhaps a better indication of age unreliability in other censuses. If that is so, the error rates are still strikingly high for all of the adult categories, exceeding 25% for all but the category of nuclear sons and daughters.

A final method that we adopted to make our data more generalizable to other censuses was to correct for the unusually high amount of age heaping in our data by examining only those grid units that appeared to have limited age heaping in the second enumeration. Therefore, we confined our attention to grid units in which census takers showed at least a minimum of concern with the correct recording of ages. The rates of age discrepancies shown in Table 6 reflect greater census taker diligence; while still quite high, they are substantially lower than the rates shown in Table 5 for the entire sample. In the 2 or more year age discrepancy category, the rates of error for nuclear, relative, and non-relative children were 5.8%, 21.0% and 10.0% respectively. For nuclear, relative, and unrelated adults, the rates of error were 42.5%, 36.5% and 36.3%.

If these data from the more carefully enumerated subsample can be taken to be representative of most nineteenth century censuses, the levels of error were typically quite high for all subgroups except nuclear children. Whether it was because of lower age consciousness or merely an absence of concern on the part of the census marshals in collecting exact ages, we cannot say for sure. Clearly, too, these findings need corroboration from other sources, but the level of error is

sufficiently high to call into question historical analysis that must use age as a precise measure and to present formidable problems for studies which rely on age accuracy in linking procedures.

Correlates of Enumeration Error

Our analysis of the *correlates of error* divides into two parts. First, we investigate how the various measures of unreliability relate to the aggregate level socio-economic and ecological variables. This part of the analysis is analogous to our earlier examination of the relationship between household location and characteristics of the grid units. In both instances, we have attempted to determine if census errors are disproportionately concentrated in certain types of neighborhoods. After making this determination, we then turn to the question of whether the various forms of error are themselves inter-related. This provides a test of whether certain census takers were particularly prone to error.

The correlations between the same contextual variables that were displayed in Table 2 and the different measures of unreliability are generally insignificant. A close examination of the correlation matrix (not shown) reveals that selected social, demographic, and ecological variables at the aggregate level are related to certain forms of error, but the relationships are weak and inconsistent. In the absence of more compelling evidence, we are inclined to conclude that error is more or less randomly distributed throughout the grid units. In other words, we did not find that excessive amounts of household turnover or rates of name and age discrepancies were disproportionately concentrated in rich or poor areas of the city, sections heavily populated or more sparsely settled, or neighborhoods that were predominantly occupied by skilled or unskilled workers. In short, we have no reason to believe that the quality of census taking was strongly affected by the characteristics of the social and physical environment.

We do have some evidence, however, that certain census takers may have produced lower quality work than others. When the various forms of error were inter-correlated, some modest relationships emerged. We discovered that unreliability was slightly higher in areas where household location rates were low. The level of the correlations was generally low, but they were with only a couple of exceptions in the predicted direction.

Our exploration of the patterns of error is not far enough along to draw any firm conclusions, but it would appear at this stage in our analysis that some census takers may have accounted for a disproportionate share of the mistakes. If further investigation confirms this

impression, we shall be able to assess the effect of such census-taker error on various substantive findings by removing the more error-ridden portions of the sample.

CONCLUSIONS

The results described in this paper represent a preliminary and, in some sections, an admittedly provisional assessment of the quality of one census of a single locality, the 1870 census of Philadelphia. To extend these findings to other censuses in the past, or to conclude that our results would necessarily be replicated in non-urban areas in 1870 (or even other cities) is an unwarranted and unwise generalization from a single study. Yet, if our findings are viewed as a first step in appraising the quality of historical census materials, there are some lessons which can be drawn from them.

The rate of underenumeration, as Superintendent Walker predicted, is not as severe or widespread as some critics of the census may have feared. We were able to re-locate, in the manuscript listings compiled in November, 82.2 percent of the households sampled from the June enumeration. The untraced households represent a combination of actual moves and several types of error: 1) the census taker's failure to find the household; 2) inaccurate listings making it difficult to link household members from the first to the second enumeration; and 3) coding error caused by missing households in the linkage search that were actually listed in the census. While we have no reliable way of assessing the relative share of the first two types of errors, we feel reasonably confident that the third form of error is minimal, because all untraced households were independently checked by a second coder.

While it is of some concern to determine the source of error, a more pressing problem is to distinguish movement from mistakes. Beyond the crude technique for estimating minimum error of searching for non-located household heads in the city directories, we could think of no sure method of differentiating errors from residential mobility. While we can safely assume that a substantial portion of the untraced households actually relocated during the period from June to November, we cannot say whether it was a third, half, or two thirds of the total. Certainly, previous studies of residential movement in the 19th century lead us to suspect that the lion's share of the non-located households had actually changed residence, but this is only a guess on our part.

This guess is informed by an analysis of the correlates of non-persistence presented in the second section of the data analysis.

Factors which have been identified in previous studies as correlates of urban migration in the 19th century, such as age of household head, presence of boarders and servants, wealth, and occupational status were also found to be related to non-location in our investigation. While the correlations were not strong, they certainly point to mobility as part of the explanation for our inability to trace a sixth of the households sampled from the first enumeration.

There is also reason to believe that a substantial share of the households sampled were not located due to imperfect census taking. We found lower rates of location in German neighborhoods, implying that language problems may have contributed to poor enumeration. Moreover, toward the end of our analysis, we discovered that, at the level of the grid unit, non-location was somewhat related to unreliable data on household members among the families which were located. This seems to us to be *prima facie* evidence that a share of the untraced households were not found because of census taker carelessness (missing or skipping households) or, perhaps because of faulty information provided by ill-informed household members (obscuring the linking process).

If the only source of error in the census were in household location, the data from the Philadelphia recount would be generally reassuring. However, a second consideration looms large in our methodological assessment of the census manuscripts, that is, the accuracy of the listings themselves. As we learned earlier, our reliability check is restricted to name and age, because other personal information was excluded from the second enumeration. Whether our reliability estimates based on name and age have bearing on the accuracy of occupation, place of birth, or literacy, we cannot say for sure, but we have no reason to believe that these items were more faithfully recorded than those we were able to examine.

There is both good and bad news to report. The reliability of the data on exits, entrances, ages, and names is reasonably high for nuclear family members, especially for children. We should add here that our data provide quite a conservative test, for the 1870 census was not generally noted for its quality. (In 1880, steps were taken to improve the procedures for collecting data and to standardize the guidelines for recording household information.) Moreover, in the re-enumeration, higher priority was given to counting individuals than to recording their names and ages correctly. Nonetheless, as we said, name discrepancies among nuclear family members are reasonably few. Since most of the population lived in nuclear households or were members of nuclear families embedded in larger households, this finding is reassuring.

However, for the minority who were residing as extended kin, boarders, or servants, we have less reason to feel secure about the data. Age and name discrepancies, especially among adults, were as much the rule as the exception. Similarly, the exit to entrance ratios lead us to suspect the quality of the data on these household members. There are several possible explanations why the data on non-nuclear individuals match so poorly at the two points in time. We have seen that this segment of the population was highly prone to movement. Typically, it seems boarders and extended kin lived in households for short durations. Existing research on the journey to work in Philadelphia during the second third of the 19th century has shown that boarding was a temporary convenience which accompanied a job change, reducing the distance to the new workplace (Hershberg et. al. 1974). Faced with a tight housing market, workers had little choice but to find temporary lodging while they looked for a more permanent residence.

Given the likelihood of rapid turnover, it is not surprising that respondents to the census did not provide accurate responses to the questions about the names and ages of non-nuclear household members. Moreover, if different respondents replied to the two censuses, we might anticipate a high degree of unreliability in the information on the auxiliary members in the household. This may also help to explain why various types of error were found to be slightly correlated. Information provided by boarders and servants on other household members was frequently unreliable, which no doubt explains why the presence of such individuals in the household reduced the quality of the census taking.

Finally, it is possible that less importance was put on precise knowledge of age in the 19th century. In an era before minimum age requirements, the military draft, social security, and retirement, age may have had lower saliency. The less age graded the society, the less concern there was with chronological exactitude (Kett, 1977).

If these factors help to explain the low reliability of the data on non-nuclear household members, they in no way minimize the problems created by such imprecision in the census. Even for nuclear family members, the data meet only modest standards of reliability; and the information on individuals outside the nuclear unit must be used with extreme caution. Currently, a major frontier of quantitative historical research involves record linkage techniques. Our findings help to explain why record linkage techniques using census data have yielded relatively low rates of persistence. Most existing studies have concentrated on household heads, where we would expect the best results. Attempts to link other members of the household are likely to prove frustrating indeed. Until we develop precise techniques for

estimating the level and effect of unreliability in the data, it will be hazardous to draw firm conclusions about the rate of mobility in the population based on record linkage data.

If our data could be shorn of error, we still suspect that we would find a high amount of residential instability in the population. The figures on turnover in the household by age and relation to household head persuade us that there was a good deal of short-term movement, particularly among young adults of working age. Undoubtedly, many left their families to find work in other parts of the city or outside Philadelphia, only to return when they were laid off. Their place in the household was taken by other relatives and boarders who were, in turn, pursuing employment opportunities which required proximity to the work-place. In many respects, then, the characteristic pattern of the 19th century city resembled that of migrant workers or perhaps ghetto youths of today, who must adopt a peripatetic lifestyle out of economic necessity.

We have only begun to examine the substantive implications of error in the 1870 census, but we feel that it provides an unusually useful short-term longitudinal data set for exploring mobility within the household unit. When we are more certain of the sources of error and potential bias in the data, we shall fix our sights on understanding patterns of exit and entrance from the home, trying to amplify the speculative conclusions presented above. Since we have an array of information on the characteristics of the household in the first enumeration as well as supplementary information on the neighborhood context, we expect to be able to explain why certain households experienced a tremendous amount of turnover in membership while others remained fixed in composition.

These plans serve as a reminder that the data presented here are but the initial report in a series of analyses. The methodological questions raised in this paper are part of a systematic effort on the part of the Philadelphia Social History Project to evaluate the data base on which it and many similar undertakings have been built. The enormous promise of quantitative history can only be fulfilled when we are intimately acquainted with the limitations of our sources of information.

SOCIOLOGY AND SOCIAL RESEARCH

Table 1

**Proportion of Households Located in the Second Enumeration
by Selected Characteristics of Households**

	Proportion of Households Located, in Percent	
Household type		
Only nuclear members, no children	67.2	(189)
Only nuclear members, with nuclear children	81.0	(1018)
Other than nuclear members, no children	83.8	(351)
Other than nuclear members, with nuclear children	86.8	<u>(819)</u>
		(2377)
Number of nuclear family members		
1 - 3	80.1	(1076)
4 - 5	84.2	(753)
6 or more	83.6	<u>(530)</u>
		(2359)
Number of non-relatives		
0 - 1	79.7	(1773)
2 - 4	88.6	(482)
5 or more	92.8	<u>(138)</u>
		(2393)
Household headship		
Couple - headed	82.6	(1878)
Male - headed	76.4	(165)
Female - headed	82.5	<u>(342)</u>
		(2385)
Ethnicity of head		
Black	80.9	(256)
Irish	80.7	(471)
German	76.0	(304)
Native White	85.0	(1195)
Other	82.9	<u>(146)</u>
		(2372)
Age of head		
34 or younger	77.5	(690)
35 - 49	81.7	(995)
50 or older	87.6	<u>(703)</u>
		(2388)
Place of birth of head		
Pennsylvania	84.4	(1058)
Other U.S.	93.9	(373)
Foreign born	79.6	<u>(925)</u>
		(2356)
Occupation of head		
Non-manual, including proprietors	85.4	(601)
Skilled manual	82.1	(771)
Higher status laborers	80.3	(507)
Lower status laborers	76.5	<u>(226)</u>
		(2105)
Real property of household		
\$0	78.8	(1621)
\$100 - \$2800	86.3	(262)
\$2801 - \$9999	90.2	(287)
\$10,000 or more	92.4	<u>(223)</u>
		(2393)

Table 2

Correlates of Rates of Location
of Households in the Second Enumeration*

Percent of households containing only nuclear members	r = -.27
Mean number of nuclear family members	r = +.03
Mean number of non-relatives	r = +.26
Percent Native White	r = +.23
Percent Black	r = -.01
Percent German	r = -.34
Percent Irish	r = -.06
Mean age of head, in years	r = +.03
Percent of heads in non-manual occupations, including proprietors	r = +.21
Percent of heads in skilled manual occupations	r = -.09
Mean real property, in dollars	r = +.11
Mean width of streets in grid unit (from 1 to 3, with 1 narrow, 3 wide)	r = +.06
Density of grid unit (mean number of persons per household)	r = +.04
Number of streets in grid unit	r = +.06

* All of the correlates are proportions or means for all sampled households in grid units, with the exception of mean width of streets in grid unit, density of grid unit, and number of streets in grid unit.

SOCIOLOGY AND SOCIAL RESEARCH

Table 3

Number of Individuals Counted in Both Enumerations, Number of Exits,
Number of Entrances, Rate of Exits, and Ratio of Exits to Entrances,
for 14 Categories of Household Members

	<u>Number of Individuals</u>	<u>Number of Exits</u>	<u>Number of Entrances</u>	<u>Exit Rate</u>	<u>Exit/Entrance Ratio</u>
Head of household	1968	45	22	2.3%	2.05
Spouse	1552	32	14	2.1%	2.29
Nuclear children - aged 1 or younger	401	34	42	8.5%	.81
Male nuclear children, 2 to 12	1088	35	36	3.2%	.97
Female nuclear children, 2 to 12	1090	30	37	2.8%	.81
Male nuclear children, 13 to 19	548	32	28	5.8%	1.14
Female nuclear children, 13 to 19	549	33	32	6.0%	1.03
Male nuclear children, 20 or older	404	47	38	11.7%	1.25
Female nuclear children, 20 or older	385	45	36	11.7%	1.25
Relatives, 19 or younger	98	2	30	22.4%	.73
Relatives, 20 or older	276	69	88	25.0%	.78
Non-relatives, 12 or younger	352	105	166	29.8%	.63
Non-relatives, 13 to 19	348	151	145	43.4%	1.04
Non-relatives, 20 or older	1640	681	897	41.5%	.76
TOTAL	10699	1361	1611	15.1%	.84

Table 4

Proportion of Individuals with Name Discrepancies
for Seven Categories of Household Members

<u>Category</u>	<u>Proportion of Name Discrepancies</u> (in percent)
Nuclear children 19 or younger	8.1% (3,512)
Nuclear members 20 or older, including household head and spouse	6.7% (4,140)
Other relatives 19 or younger	19.7% (76)
Other relatives 20 or older	7.7% (207)
Non-relatives 12 or younger	13.0% (247)
Non-relatives 13 to 19	20.3% (197)
Non-relatives 20 or older	<u>12.6%</u> (959)
Total	8.4% (9,338)

Table 5

**Proportion of Individuals with Age Discrepancies, Using Both
a Liberal Measure (Discrepancies of Two or More Years) and a
Conservative Measure (Discrepancies of Five or More Years)
For Nine Categories of Household Members**

<u>Category</u>	<u>Proportion with Age Discrepancies of 2 or More Years in percent</u>	<u>Proportion with Age Discrepancies of 5 or More Years in percent</u>
Household head	55.1% (1,923)	27.0%*
Spouse	61.5% (1,520)	26.6%
Nuclear children 19 or younger	18.6% (3,512)	2.0%
Nuclear children 20 or older	39.6% (697)	14.8%
Other relatives 19 or younger	19.7% (76)	2.6%
Other relatives 20 or older	53.6% (207)	27.1%
Non-relatives 12 or younger	19.0% (247)	3.6%
Non-relatives 13 to 19	22.8% (197)	4.6%
Non-relatives 20 or older	<u>51.3% (959)</u>	<u>26.9%</u>
Total	37.8% (9,338)	15.3%

* The bases are the same for each of the measures of age discrepancies.

Table 6

Proportion of Individuals with Age Discrepancies, using both a Liberal Measure and a Conservative Measure, for Nine Categories of Household Members, in those Grid Units with Low Levels of Age Heaping

<u>Category</u>	<u>Proportion with Age Discrepancies of 2 or More Years, in Percent</u>	<u>Proportion with Age Discrepancies of 3 or More Years, in Percent</u>
Household head	41.9% (463)	20.5% *
Spouse	43.4% (346)	17.9%
Nuclear children 19 or younger	5.8% (1901)	3.2%
Nuclear children 20 or older	26.0% (145)	6.9%
Other relatives 19 or younger	21.1% (19)	0.0%
Other relatives 20 or older	36.5% (71)	21.1%
Non-relatives 12 or younger	10.0% (100)	2.0%
Non-relatives 13 to 19	12.0% (58)	1.7%
Non-relatives 20 or older	36.3% (276)	14.9%
TOTAL	18.9% (3379)	8.5%

* The bases are the same for each of the measures of age discrepancies.

FOOTNOTES

¹ Our procedure, of course, does not take into account the undercounting which occurred in the initial census, which we might have estimated by working back from the second to the first enumeration. We plan in future work to attempt to estimate the degree of underenumeration in the June census.

² Our initial intent was to sample equally from each of the four major ethnic groups. However, because black household heads were heavily concentrated in only a few grid units, we risked having too few black households in our sample. Therefore, we modified our sampling plan to take this problem into account by selecting more grid units where blacks were disproportionately present, and fewer native white dominated grid units.

Our sample is not precisely representative of the population of Philadelphia. The Blacks, the Germans, and the Irish are all overrepresented. In addition, grid units with a population of less than 50 persons were excluded from the sample to insure that there would be sufficient households at the neighborhood level to conduct an aggregate analysis. These excluded grid squares were few and were disproportionately located in the less developed fringes of the city.

The sampling of grid units that were more highly concentrated with each ethnic group did not confine our selection to ethnically homogeneous grid units. In fact, with the exception of native white dominated grid units and a few black dominated grid units, the

SOCIOLOGY AND SOCIAL RESEARCH

vast majority of grid units from which the sample was drawn had a dominance by one ethnic group of less than 50%. Thus, most of the neighborhoods from which the households were drawn were ethnically diverse even though they contained a higher portion of the designated ethnic groupings.

³ The 1870 census, unlike those in 1880 and thereafter, did not specifically identify relationships within the household. However, census takers were instructed to list the head of the household and spouse first, followed by their children in decreasing order of age. Therefore, relationships within the nuclear family can usually be correctly imputed. The relationship of other household members to the nuclear family is more difficult to determine. While it can be assumed (as we did) that individuals that share the head's last name but who do not fit in the nuclear family are part of the extended family, it is impossible to determine the relationship to the household head of individuals with different surnames, even though some of them are surely the extended kin, particularly the spouse's kin. We know, however, from the 1880 census that the bulk of this different surname category consists of boarders, lodgers and servants. In this essay, we will refer to individuals who appear to be nuclear family members as nuclear, other individuals with the same surname as the nuclear family as other relatives, and individuals with different surnames as non-relatives.

⁴ Although there is reason to believe that mobility was generally high in the 19th century city, we have no way of differentiating migration from the city and migration within the city. As we shall argue later on, based on data produced by Hershberg et al. (1974), before the advent of public transportation most individuals had to live near their place of work. Thus, in a fluid economy, workers were frequently compelled to shift their residence in order to remain employed.

⁵ The aggregate variables that we correlated with rates of location are averages of the sampled households, with the exception of the density of the grid unit and the number of streets in the grid unit. It should be kept in mind that the characteristics of one or two streets within a grid unit do not necessarily reflect the aggregate characteristics of the grid unit.

⁶ We are especially indebted to a paper by Ralph Chiumenti (1975) on discrepancies in the 1870 census. Chiumenti both reviews the literature on age unreliability and does a preliminary analysis of the 1870 recount.

BIBLIOGRAPHY

Chiumenti, Ralph

1975 "A Study of Response Error on Age in the Two Censuses of Philadelphia in 1870", unpublished manuscript.

Chudacoff, Howard P.

1972 *Mobile Americans: Residential and Social Mobility in Omaha*. New York: Oxford University Press.

Heer, David M.

1968 (Ed.) *Social Statistics and the City*. Cambridge, Mass.: Joint Center for Urban Studies.

Hershberg, Theodore; Cox, Harold; Light, Dale Jr., and Greenfield, Richard

1974 *The Journey to Work: An Empirical Investigation of Work, Residence and Transportation, Philadelphia, 1850 and 1880*. (mimeo) rev. ed. in *Toward an Interdisciplinary History of the City: Work, Space, Family and Group Experience in Nineteenth-Century Philadelphia*, Theodore Hershberg, Ed., Oxford University Press, forthcoming.

Historical Methods Newsletter

1976 "A Special Issue: The Philadelphia Social History Project" Vol. 9, No. 2-3.

- Katz, Michael E., Doucet, Michael J. and Stern, Mark J.**
1978 "Migration and the Social Order in Erie County, N.Y.: 1855." *Journal of Interdisciplinary History*, Vol. 8, 4 (Spring), pp. 669-701.
- Kett, Joseph F.**
1977 *Rites of Passage: Adolescence in America, 1790 to the Present*. New York: Basic Books.
- Modell, John; Furstenberg, Frank F., Jr.; and Hershberg, Theodore**
1976 "Social Changes and Transitions to Adulthood in Historical Perspective", *Journal of Family History*, Vol. 1, 1, pp. 7-32.
- Phillips, Derek, Jr.**
1972 *Knowledge from What?: Theories and Methods in Social Science*. Chicago: Rand McNally.
- Sharpless, John V. and Shortridge, Ray M.**
1975 "Biased Underenumeration in Census Manuscripts," *Journal of Urban History*, Vol. 1, 4, pp. 409-437.
- Thernstrom, Stephen**
1973 *The Other Bostonians: Poverty and Progress in the American Metropolis 1880-1970*. Cambridge, Mass.: Harvard University Press.
- Webb, Eugene J.; Campbell, Donald T.; Schwartz, Richard D. and Sechrest, Lee**
1966 *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: McNally.
- U.S. Census Office, Ninth Census**
1870 Vol. 1 (Washington, 1972), XIX-XXXIV.
- Winchester, Ian**
1970 "The Linkage of Historical Records by Man and Computer: Techniques and Problems," *Journal of Interdisciplinary History*, Vol. 1 (Autumn), pp. 107-125.