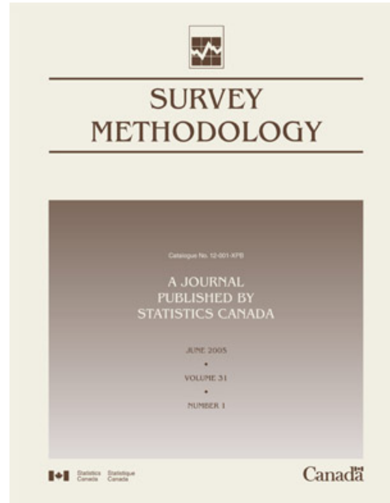
Catalogue no. 12-001-XIE

Survey Methodology

December 2003



Statistics Canada
Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada

Business Survey Methods Division

Survey Methodology

December 2003

Published by authority of the Minister responsible for Statistics Canada

February 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 29, No. 2, pp. 129-138
Statistics Canada, Catalogue No. 12-001

The Accuracy and Coverage Evaluation: Theory and Design

Howard Hogan [1]

Abstract

This paper discusses both the general question of designing a post-enumeration survey, and how these general questions were addressed in the U.S. Census Bureau's coverage measurement planned as part of Census 2000. It relates the basic concepts of the Dual System Estimator to questions of the definition and measurement of correct enumerations, the measurement of census omissions, operational independence, reporting of residence, and the role of after-matching reinterview. It discusses estimation issues such as the treatment of movers, missing data, and synthetic estimation of local corrected population size. It also discusses where the design failed in Census 2000.

Key Words: Dual system estimation; Census adjustment; Undercount.

1. Introduction

The U.S. Census Bureau attempted to correct the initial Census 2000 population figures for measured net undercount (U.S. Census Bureau 2000.) This correction was to be based on the Accuracy and Coverage Evaluation (A.C.E.). The A.C.E. is a post-enumeration survey based on the dual system estimator (DSE). Although seemingly well designed and well executed, the initial A.C.E. production estimates were badly flawed. The A.C.E. produced an estimate of 3.3 million net undercount (378,000 s.e.). This contrasts sharply with the current demographic analysis estimate of only 340 thousand (Robinson 2001) as well as a later revised survey estimate of a 1.3 million overcount (542,000 s.e.) (U.S. Census Bureau 2003).

This paper discusses both the general question of designing a post-enumeration survey (PES), and how these general questions were addressed in the U.S. Census Bureau's plans for the A.C.E. Where applicable, it discusses where the assumptions underlying the design failed in 2000. Throughout, I will use the terms DSE and PES when a general question is discussed and A.C.E. for specific details of the U.S. 2000 design. The next section defines the dual system model as applied to census coverage measurement. Section 3 discusses the definition and measurement of census correct and erroneous enumerations. Section 4 presents the issues in defining and measuring omissions. Section 5 deals with small area estimation. The paper ends with a discussion of some of the problems encountered in implementing the A.C.E together with some concluding remarks.

2. The Dual System Estimation Model

The use of the dual system model is well known either for measuring the completeness of vital events registration (Sekar and Deming 1949; Marks, Seltzer and Krotki 1974) or for use in measuring coverage errors in census data (Marks 1979; Wolter 1986; U.S. Bureau of the Census 1985.) Application of the dual system model in the context of the 1990 Census, including the issue of census adjustment, is documented in Hogan (1992, 1993.)

The standard Petersen (1896), Sekar-Deming or dual system estimator (DSE) can be expressed as:

$$\hat{N}_{++} = N_{+1}\,(N_{1+} / N_{11}) \qquad (1)$$

where

- N_{11} is the number of people counted in both the census and the survey,
- N_{+1} is the number of people correctly counted in the census,
- N_{1+} is the number of people counted in the survey, and
- N_{++} is the total number of people.

That is, the total population is estimated by the number captured in the census multiplied by the ratio of those in the survey to those in both systems (*i.e.*, the inverse of the coverage rate of the census, as measured by the survey).

The DSE will yield a direct estimate of the population of class *j*, as well as any sum of classes. The class *j* might be the household population of a state, of a district, of an ethnic group, or perhaps of an ethnic group within a state.

Requirements for estimating small or local populations, for example, age by sex, by race, by town, often far exceed the capacity of even a very large sample. To meet this need, the DSE is combined with a synthetic assumption to produce estimates for areas of geography smaller than that defined by the domain *j*. The synthetic estimator assumes that a proportion or ratio measured at an aggregate level applies equally to all sub-groupings (Gonzalez 1973; Gonzalez and Hoza 1978.) Using a synthetic assumption, we write

$$\hat{N}^{s}_{jkh} = CCF_j\, C_{jkh} \qquad (2)$$

1. Howard Hogan, Chief, Economic Statistical Methods and Programming Division, Census Bureau, Washington, D.C. 20233.

$$CCF_j = \frac{\hat{N}_j}{C_j} \tag{3}$$

where,

$\hat{N}^s_{jkh}$ is the estimated population in domain j, available at the level of geography k and demographic subclass h.

CCF_j is the net coverage correction factor.

$\hat{N}_j$ is the DSE for domain j.

C_{jkh} is the measure (usually census count) of the population in domain j available at the level of geography k and demographic subclass h, and

$$C_j = \sum_k \sum_h C_{jkh}. \tag{4}$$

C_j need not equal the number of people correctly included in the census (N_{+1}). N_{+1} is estimated from sample data and is not available for all small areas. C is normally the census count, including imputations and erroneous inclusions (duplicates, *etc*.).

Summing over group j and subclass h yields a measured population for the given geographic area k, we have

$$\hat{N}^s_k = \sum_j \sum_h CCF_j C_{jkh}. \tag{5}$$

For example, j may define all 0–17 year–old Asians in owner-occupied housing units while k may define Orange County, California, and h may define 11 year–old girls.

While this produces a small-area and small-group estimate, this calculation can generate fractions. The typical user of census data prefers whole person records. The U.S. Census uses controlled rounding and person record imputation to create integer number of person records for ease of tabulation and data acceptance.

3. Measuring Correct Enumerations

3.1 Defining and Measuring Correct and Erroneous Enumerations

The first step in operationalizing Equation 1 is to define and estimate the set of individuals "correctly" in the census. In this context "correctly" has four dimensions:

1. Appropriateness
2. Uniqueness
3. Completeness
4. Geographic correctness

"Appropriateness" means that the person should be included in the census. People who die before or who were born after the census reference date (April 1 in the U.S.) are not part of the population (universe) to be measured. Similarly, records that refer to fictitious "people," tourists, or animals are out-of-scope.

"Uniqueness" refers to the fact that we wish to measure the number of people included in the census, not the number of census records. If more than one record refers to a single person, the count of records must be reduced for purposes of the DSE.

"Completeness" means that the census record must be sufficient to identify a single person. If it lacks sufficient identifying information, we cannot determine whether the person was appropriately and uniquely included in the census, nor can we determine whether he or she was also included in the survey.

Although completeness is necessary for the DSE, the census count includes imputations and other incomplete enumerations. Census operations normally have a requirement for a "data-defined person record." In Census 2000, the requirement was two characteristics where name counts as a characteristic. The name field must have at least three characters in the first and last name fields combined. The characteristics that are included in the counting are relationship to the householder, sex, race, Hispanic origin, and either age or year of birth. (Childers 2001)

When a record does not meet these requirements census processing substitutes (imputes) a data-defined record. Since the census processing identifies all these whole-person imputations, the quantities are known and need not be estimated. Traditionally, the number of whole person imputations is denoted by II, for "insufficient information."

Additionally, there are person records that are acceptable for census processing but insufficient for use in the DSE. This group includes records with reasonably complete data but without a person's name. Accurate matching or additional interviewing is not possible for these cases. For A.C.E. 2000, the definition for "sufficient information for matching" was complete name and two characteristics. (Childers 2001)

"Geographic correctness" means that people are included in the census where they should be included. Enumerations outside that defined search area (or areas) are counted in the census but not correctly included in the census. This area must be searched during the matching process as well as searched for census duplicates. As the number of addresses in the search area increases, the complexity of matching increases and the chance of matching error grows. This increased complexity and possible levels of error will affect both the matching between the survey and the census and the search for census duplicates. The more addresses that must be searched, the more likely a true match will be missed. Equally importantly, the chance of a false match increases. For example, the chance of finding two people with similar names and ages living in the same block is small. The chances of finding two such people in a large city is considerable.

Two dimensions must be defined to operationalize a search area: (1) correct location and (2) the search area around the correct location.

The "correct location" defines where, under the DSE residence rules, the person should be included in the census. These rules may differ from the rules used in the census. The only requirement is that the location be precisely defined and consistently applied during PES processing. More than one location may be defined as correct so long as the rule is consistently applied. However, usually only one location is defined as correct. This was the rule in the A.C.E.

In the 1990 PES and 2000 A.C.E. the Census Bureau adopted the following rule:

> The person is correctly included in the census if he or she is included at the location where the person considers, at the time of the survey interview, to have been his or her usual residence as of April 1.

This definition generally follows the census rules. However, it makes an explicit allowance for the fact that the concept of "usual residence" is somewhat subjective. Because of this subjectivity, where the person considers his/her usual (April 1) residence may have changed by the time of the survey interview. This, by itself, does not bias the DSE. However, it does require consistent reporting of the "correct location."

The second dimension of geographic correctness is the area of search around the correct location, *i.e.*, the search area. The concept of a search area is to accommodate errors in either the census or survey assignment of residents to a particular geography. It has the effect of lowering the variance and can, in some circumstances, lower the bias as well.

The A.C.E. used the following definition:

> A person was correctly enumerated if the person was counted in the block cluster containing his/her usual residence; or if he/she was included by the census in the housing unit where he/she usually resides, and the housing unit was included in a block adjacent to the correct block cluster.

An important part of this design is that enumerations of people in the "wrong" location are to be classified as erroneous, whether or not the people are also enumerated in the correct location. Thus a person counted only once, but in the wrong location, should be measured, on average, as contributing one erroneous enumeration (in the wrong location) while being missed (one omission) in the correct location. This approach obviates the need to search widely for possible duplicates, but does require that the field interview determine a unique correct location for each person.

The definition of "correctly included" does not depend on the correctness of classification j. For example, if a person was really 19 years–old, but was counted in the census as 17, he/she is still considered as correctly included. This is discussed in section 5.2.

To estimate the number of people correctly included in the census, one must take a sample of all data-defined census enumerations. This sample is called the enumeration (or E) sample. Census whole-person imputations (II's) are not part of the E–sample frame.

To maximize correlation with the population sample (see below), the A.C.E. first defines a set of sample areas. These are either a single block or a group of contiguous blocks and are known as block clusters. If a block is sampled, all census records coded to that block, even incorrectly, fall into sample. If the block contains many census housing unit records it may be subsampled.

The records in the E–sample will be checked for completeness. Only records that meet the minimum completeness requirement can be considered as correctly enumerated in the census. Records are then searched throughout the search area to see if the person was counted more than once within the sample block (uniqueness). Duplicate search is done using computer-assisted clerical matching. If more than one record is found, the extra records are coded as duplicates.

Appropriateness and geographic location cannot be determined from the census enumeration alone, but require additional interviewing. If interviewing locates a member of the household, or an acceptable respondent who can confirm the person's existence and that the person had his/her usual residence there on April 1, the enumeration is accepted as correct.

If the respondent reports that the person did not live in the block or search area on April 1, the enumeration is excluded from the correct enumerations. This can occur when the person responded to the census but moved before April 1; the person moved in after April 1 but was enumerated by the census nonresponse follow up operation; or when a parent incorrectly reports a college student as living at home.

The interviewers may determine that the person never existed or was never associated with the block. These records are considered erroneous. It is difficult in some cases to prove that a "person" was not real, especially in a large block. The A.C.E. required the interviewers to find at least three knowledgeable respondents before coding a record as fictitious. However, since the person might have lived somewhere else in the block, it can be difficult in some situations to code the record fictitious.

An important source of error arises from the need to accept proxy responses to verify many enumerations. If the proxy reports a different "correct" residence than the person himself would, an enumeration could be miscoded, since the requirement of a unique "correct" residence would be violated. The A.C.E. used proxy interviews for households that moved between the time of the census and the time of the A.C.E. interviews. Even within a household, different members may hold different views of a person's "correct" residence on Census Day. Proxy respondents, both household and non-household, were responsible for many of the errors in reporting residence in the A.C.E. and thus, the underestimation of census error.

After missing-data estimation and sample weighting, we can estimate the number of people correctly counted in the census as

$$N_{+1} = (C - II)\,\frac{CE}{N_e}. \qquad (6)$$

Where

C = Census total records, including imputed, duplicate, fictitious, *etc.* (the Census count),

II = number of whole-person census imputations,

CE = weighted estimate of appropriate, unique, complete and correct enumerations,

N_e = weighted E–sample estimate of total, including duplicate, fictitious, *etc*.

Occasionally, due to processing errors or timing constraints there may be a group of census enumerations that are excluded from both the E–sample processing and from the searching and matching process. Thus, while these records may be processed in time to be included in the official census results, they arrived too late be included in coverage measurement processing. These cases are sometimes known as "Late Census Adds" (LCA). These cases can be handled analogously to the treatment of census whole person imputations, that is replace (C–II) in Equation 6 with (C–II–LCA). Excluding the LCAs will not affect the DSE of the true population if the number of matches is reduced proportionally to the number of census correct enumerations. Said another way, the assumption is that the probability of a LCA being excluded from the A.C.E. processing must be statistically independent of its inclusion probability in the A.C.E. This is, of course, the traditional dual system independence assumption. (See Hogan 2001 for the supporting theory.) Although there were 2.3 million LCAs in Census 2000, analysis of the A.C.E. results by Raglin (2002) showed a trivial impact on the final DSE results.

In situations where the number of whole person imputations (II) was small, $(CE / N_e - 1)$ would be a measure of census gross overcoverage. That measure, however, is a function of the operational definitions of "correctly enumerated" adopted by the coverage measurement design. Definitions adopted to produce a good measure of net coverage, especially with respect to completeness and geographic correctness, may differ from those most appropriate for studying the quality of Census field operations. In any case, Census 2000 included 5.8 million whole-person imputations, of which 1.2 million were for housing units where the interviewer was unable to obtain even the number of residents (see Table 1 in Nash 2001, and page ii of Wetrogan and Cresce 2001.)

4. Measuring the Proportion of People Correctly Enumerated

Having defined the set of correctly enumerated people, the next step in the DSE is to estimate the census coverage rate, N_{11} / N_{1+}.

Conceptually, estimating the rate entails (1) taking a sample of people, (2) determining whether they should be enumerated in the census, and (3) determining whether they were, indeed, correctly enumerated, using the same definitions as were used to measure N_{+1}. If an unbiased sample can be drawn of people who should have been enumerated and, if we can determine whether they actually were correctly enumerated (included in the census), then the DSE will produce asymptotically unbiased estimates. If each step can be approximately correct, the results will approach an unbiased estimate.

The first step in the process is, normally, to draw a random area sample. The A.C.E. uses the same set of block clusters for this purpose that it uses to define the E–sample.

Interviewers then canvass the block and prepare an independent list of people who should have been enumerated. This list constitutes the population or $P-$sample. The (weighted) sum of the people on this list, denoted $\hat{N}_p$, estimates N_{1+}. However, it is not the number which is of interest, but the ratio of N_{11} to N_{1+}, which we approximate by the ratio of correct matches, $\hat{M}$, to $\hat{N}_p$.

Operationally, the "correctly enumerated" census records are searched to see if the P–sample people were enumerated. The (weighted) number who were matched $(\hat{M})$ estimates N_{11}.

The DSE model will work if we can approximate:

1. Operational independence
2. Consistent reporting of residence
3. Accurate matching
4. Homogeneity within post-stratum

4.1 Operational Independence

Operational independence is the easiest assumption to approximate, but still requires vigilance. In Census 2000, the A.C.E. sample was drawn and the housing units listed before the delivery of the census questionnaires. Although personal contact was minimal, some people may react differently to the census because of their inclusion in survey listing. Early telephone interviews were allowed for independently listed housing units linked to a census address with a completed census questionnaire. This operation occurred while census nonresponse follow up was still being conducted in the area. Personal visit interviewing took place concurrently with some census "coverage improvement" interviewing. Clearly, some contamination could occur. Great care was taken to prevent the same field staff from working the same area in both Census and A.C.E. and to prevent the sharing of information. Still, some people may react differently to the survey because they were enumerated, for example, by a very polite or very surly enumerator. Others may believe that they have a duty to provide the information once, but not twice.

Operational independence must also be preserved in office procedures. Definitions of "nonresponse" or "sufficient information" are sometimes applied differently to matched and non-matched *P*–sample records. The A.C.E. guarded against unnecessarily introducing operational dependence by forcing the processing system to first decide whether a case is acceptable for matching and only then attempt matching. The philosophy is "Do not attempt to find a match unless you would be satisfied that, if no match is found, the person was not enumerated!"

Before beginning the matching, *P*–sample records first are reviewed for:

(1) Appropriateness
(2) Uniqueness
(3) Completeness
(4) Geographic correctness

The A.C.E. contained no obviously fictitious records. One important safeguard is the use of Computer Assisted Personal Interviewing (CAPI). The CAPI instrument makes falsification difficult by "time stamping" the interview and recording every key stroke. We have instituted a quality assurance process to minimize other sloppy or dishonest A.C.E. interviewing. In addition, one important exception to the "no follow up" rule are cases where A.C.E. fabrication is possible, *e.g.*, cases where no one in the household matches, implying possible fabrication.

Out of scope records, *e.g.*, group quarters, are screened out. Occasionally, survey duplicates occur and these are eliminated (uniqueness). Finally, if the survey interview does not meet minimal standards, the case is converted to nonresponse and is later imputed.

4.2 Consistent Reporting of Residence

To measure the number of people correctly in both systems, we must determine whether or not a *P*–sample person was correctly enumerated in the census. This is done by searching the correct census records in the area where the person should have been enumerated.

The same definition of geographic correctness must apply both to whether an enumeration (in the *E*–sample) was correct and to whether the person (in the *P*–sample) was correctly enumerated. Failure to make these concepts agree is termed "balancing error."

Specifically, we must have the same definition of "correct" location and the same search area around the correct location. Errors can result in both erroneous non-matches and erroneous matches. Difficulty comes primarily from two sources. First, both the *P* and *E*–sample accept proxy responses. Thus, even though the person might have a clear and consistent understanding of his usual residence, the proxy respondent may not. Secondly, the way in which the question is posed in each interview could lead to different responses even from the same person. This might result in false non-match/not correctly enumerated status. On the other hand, if the person was incorrectly included by the census, we could incorrectly count the person as "correctly enumerated." Both errors clearly occurred on a relatively large scale in the A.C.E. (See section 6.)

The other dimension of geographic correctness is, again, the extent of search. The same area must be used to define the correct residence for determining both whether an enumeration was correct and whether a person was correctly enumerated. This is achieved by consistently applying the same search area definitions as in section 3.

4.3 Accurate Matching

The purpose of matching is to determine whether a person interviewed in the *P*–sample was also enumerated in the census within the defined search area. Much of the matching is now done by a computerized matching system. The system produces matches, possible matches, and non-matched cases. Repeated tests have shown that cases matched by the computer are nearly certainly correctly linked (Belin 1993). Nearly all clerical matching is now computer-assisted and largely paperless. This new system makes searching easier, including duplicate search. It restricts the codes clerks can apply to only those appropriate for the situation. The almost paperless system eliminated lost and misfiled A.C.E. questionnaires.

The first-level clerks were backed up by a team of 46 technicians. Training for these technicians began in September 1999. They were supported by a team of seven permanent analysts, most of whom have been matching for many years. Each level of matching acts as quality assurance for the level before. In addition, each level could refer problem cases to the next higher level. All matching was done in one location by one staff. The 1980 and 1990 matching operations were done in three and seven sites, respectively.

The use of the A.C.E. procedures for movers also greatly simplified the matching. Information about those who had moved was gathered from current residents. Under the procedures used in 1980 and 1990, movers were interviewed at their residence at the time of the PES interview. It was necessary then to code the reported correct Census Day residence to the correct census geography before beginning matching. This procedure was difficult, especially in rural areas. Mover matching was never before automated. In A.C.E., all matching, including for movers, was done in the *E*–sample block cluster or an adjacent block, using the same computer and computer-assisted clerical matching system. The change in the treatment of movers is discussed below.

4.4 The Role of After-Matching Reinterview

Some cases are sent to the field to gather further information after the initial matching is complete. This after-matching reinterview is often termed "follow up interview."

The follow up interview process, like all PES activities, must fit into the overall framework of the DSE. Specifically, it must account for:

1. Appropriate, unique and correct response
2. Independence between census and survey inclusion probabilities
3. Balancing P and $E-$sample concepts
4. Search area and unique location matching rules
5. Treatment of missing data.

Follow up is only useful if it provides more accurate or consistent responses. Simply obtaining a different response is not justification. Since follow up takes place further from the census reference date than the initial interview, it is more difficult to obtain accurate responses. This is equally true for $E-$sample follow up and $P-$sample follow up. To provide better responses, follow up must use better resources, for example: (1) better respondents (household *vs.* proxy), (2) a better trained, supervised or quality-controlled interviewer, or (3) better questions or interview procedures.

The census data collection period extends from mid-March through mid-summer. Because of the huge scale of the operation, little emphasis is placed on verifying that the people were residents of the household on April 1. Quality assurance reinterview to prevent fabrication is minimal. Because of better training and supervision, and more complete questioning, the A.C.E. follow up interviewing can, in general, obtain more accurate information on residence and location than that gathered during the census process itself. Thus all non-matched $E-$sample cases were sent to follow up.

Follow up can, however, compromise independence. If all cases were sent to follow up, independence would not necessarily be compromised. However, cases that are matched during initial matching are seldom sent to follow up. To do so would stress the resources available for follow up. Instead, only non-matches or "possibly matched" cases are usually selected for follow up. This can introduce operational dependence.

The biases that can be introduced by follow up can occur even if the follow up interview was successfully conducted, since follow up may selectively change the defined "correct location" for non-matches but not for matches. If the follow up operation results in a non-interview, further biases can be introduced depending upon the missing data models applied to these cases.

Choosing cases for follow up requires balancing the need for accurate and consistent information with the need for independence. The $P-$sample only followed up cases when better information was likely. Cases sent to follow up included:

1. Possible matches, since with the information at hand the interviewers can resolve the situation,
2. Initial non-household proxy interviews that result in non-matches. Since we have not spoken to a household member, we have reason to doubt the accuracy,
3. Non-matched cases where, for the same housing unit, the census reports one family and the A.C.E. reports another. In order to ensure consistent reporting of Census Day address between the $P-$sample and the $E-$sample, these cases are sent out together,
4. Partial-household non-matches.

Cases that match and some other non-matched cases were generally not sent to follow up. For example, the A.C.E. did not follow up whole-household nonmatched cases where the census missed the unit, reported it as vacant, or could not obtain an interview (last resort information only).

4.5 Homogeneity Within Post-stratum

The DSE requires that the capture probabilities be independent for all individuals within estimation domains called post-strata. This is approximated by making the post-strata as homogeneous as possible with respect to the census capture probabilities, and then striving for as uniform as possible inclusion probabilities for the survey.

Dividing the population into many relatively small post-strata can increase within strata homogeneity. However, small strata can have high sampling variance and ratio bias. Ratio bias follows from the fact that the DSE is inherently a ratio estimator. This bias tends to decrease as the size of the post-stratum increases. In addition, our treatment of movers adds an additional ratio (see below). For this reason, we designed post-strata with a minimum expected sample size of 100.

For the A.C.E. we post-stratified based on the following variables:

1. Race / Hispanic Origin (7)
2. Age / sex (7)
3. Tenure (2)
4. Metropolitan area size and type of enumeration area (4)
5. Return rates (2)
6. Region (4)

where the number in parenthesis refers to the number of categories. More details on the post-strata are found in Haines (2001).

Coverage differences between racial and ethnic groups is well documented. (See for example Robinson, Ahmed, Das Gupta and Woodrow 1993; Hogan 1993.) Social, cultural, linguistic and economic differences may lead different racial and ethnic groups to react differently to the census procedures.

Demographic analysis and previous coverage surveys have demonstrated that people are differentially missed in different age groups and that the pattern is different for males and females. Most important in this pattern is young adults (Robinson *et al*. 1993.)

The importance of tenure was first measured following the 1980 Census and then implemented in the 1990 post-stratification. Those who live in owner-occupied houses are less mobile. They may feel that they have more of a stake in their community and thus, are more influenced by the census outreach program.

Metropolitan area size obviously affects housing patterns and is correlated with the way the Census Bureau builds its address lists. The combined variable "metropolitan area size and type of enumeration area" isolates differences in housing unit coverage. It may, in addition, measure some aspects of social and economic isolation.

The census return rate measures public cooperation with the census, an important predictor of coverage. It also measures directly the proportion of the enumeration that must be done in the census nonresponse follow up. One difficulty in this variable is that not all areas of the country are in the mailback universe. A small proportion is done by direct interview, and obviously have no "return rate." We have chosen to group these areas with "high" mail response areas.

Census Region picks up, among other things, broad differences in settlement patterns and housing stock. "Brown stone walk ups" are more common in the Northeast. Mobile homes are more common in the South.

Obviously, the complete cross-classifications can lead to very small cells. The maximum set of post-strata the sample was designed to support was 448. In fact, after collapsing small cells, there were 416 post-strata.

4.6 Treatment of Movers

People who move between the census reference date and the time of the survey interview present a challenge for designing a DSE for census application. First, people who move are more likely to be missed by the census and by the survey. Secondly, if a person has a different "usual residence" at the time of the survey than he did at the time of the census, one must decide where to sample him.

In the 1990 PES, movers were sampled where they lived at the time of the survey interview. We then searched the census records at, and only at, their April 1 usual residence. This is known as procedure B (Marks 1979). This approach requires both coding the address to the correct Census Day geography and then matching. These activities are complex and time consuming.

The A.C.E. used a different procedure known as procedure C. The A.C.E. estimated the number of movers by the number of people who moved into the sample blocks between April 1 and the time of the A.C.E. interview (in-movers). If the population was closed to international migration, deaths, movement to group quarters, *etc.*, then the number of people who moved in must equal the number who moved out (out-movers). They are the same people in the population, if not in the sample. It is normally easier to find people where they are, so the measured number of in-movers is normally a better estimate of the total number of movers than the measured number of out-movers.

The proportion of movers who are correctly enumerated is estimated by matching the out-movers to the census records for the sample block and extended search area, if appropriate. The estimated number of correctly enumerated movers is then $\hat{M}_t = (\hat{M}_o / \hat{N}_o)\, \hat{N}_i$ where $\hat{M}$ denotes the weighted number of correct matches; $\hat{N}$ denotes the weighted population number; and the subscripts denote total movers (t), out-movers (o) and in-movers (i).

If we denote those who do not move by the subscript n, the overall coverage rate becomes

$$\frac{N_{11}}{N_{+1}} = \frac{\hat{M}_n + \hat{M}_t}{\hat{N}_n + \hat{N}_i}.$$

The effect of procedure C is to increase the effective capture probabilities in the survey for movers and thus increase homogeneity of inclusion in the survey with respect to mover status (*i.e.*, mover/nonmover) (Griffin 2000).

There will be nonresponse and incomplete response at various steps. The goal of the missing data process is to improve the estimate of the number of people correctly counted (from the E–sample) or the estimate of the coverage ratio (from the P–sample). In designing missing data procedures, we choose methods that support the underlying DSE assumptions. Starting with the 1990 PES, the U.S. has estimated the probability a nonresponse record was correct rather than assigned a "zero/one" classification. (Schenker 1988, Belin 1993) The methods used for the A.C.E. are described in Cantwell and Ikeda (in this volume).

5. Synthetic Estimation

5.1 The Synthetic and Dual System Model

To this point, we have been dealing with the actual DSE. However, as noted in section 2, we use a synthetic estimator to distribute the measured net undercount to local areas and small groups.

In the A.C.E. the carrying-down was based on the same post-stratification variables as the DSE itself. The synthetic estimation is based on the assumptions that (1) the DSE estimates the true population, and (2) within post-strata, the true population is distributed proportionally to the pre-adjustment census count.

Clearly, at some level the second assumption can be only true with respect to the expected census counts. That is, even if within post-strata all people had identical probabilities of being enumerated in the census, we would observe different outcomes across blocks. The underlying

DSE explicitly models the undercount as a stochastic process.

As areas get larger, two things happen. First, the stochastic effect, or the random "block effect" begins to average out. Secondly, the effect of the actual undercount from a collection of blocks becomes positively correlated with the post-stratum's coverage correction factor. That is, the larger the area, the more the area's undercount determines the net correction factor.

The stochastic effect would be trivial for all but the smallest areas if Wolter's (1986) autonomous independence assumption held in practice, that is, if each person was included or missed independently of all other people. In fact, it is well known whole families are often missed or duplicated. Indeed, the whole building (or sometimes even block) might be missed or duplicated by the census address listing procedure. The failure of the autonomous independence assumption does not cause a bias in the dual system model as long as the underlying probabilities are equal within post-strata. This failure can mean that observed coverage for a block is inconsistent with the estimated undercount adjustment. However, as attention is turned to larger areas, the stochastic effect diminishes and is replaced with the problem of true heterogeneity of the underlying capture probabilities (see Haines 2001 for synthetic estimation details.)

5.2 Misclassification Error

In the discussion so far, we have accepted the post-stratum classification, j, as fixed. In practice, some people will be classified in different post-strata in the census and in the survey. For example, a woman may be reported as age 28 in the census and 31 in the survey, placing her in different post-strata.

Such misreporting is normally not important for matching. Name, address, month and day of birth, relation and household composition are far more important than age, race or sometimes even sex. So, assuming a match, in the above example we would have one correctly enumerated 28 year–old in the E–sample and one correctly enumerated 31 year–old in the P–sample. Misclassification can be seen to have two effects. To the extent the true undercount probabilities are homogeneous with respect to the true characteristics, misclassification introduces heterogeneity (and heterogeneity bias) into the observed estimation cells. This is true even if reporting is consistent between the census and the survey, because it can introduce unobserved subgroups within post-strata where the probabilities of inclusion in each system are correlated.

Inconsistent reporting between the census and the survey poses a problem for the synthetic estimator as well as for the DSE. This is easily seen by ignoring census imputations and erroneous enumerations. In this case, the coverage correction factor is the inverse of the matching rate (N_{11j} / N_{1+j}) where j represents the post-stratum. If the classification into the post-strata is inconsistent between the census and survey, we would be applying the rate, estimated from one group, to a somewhat different group. While misclassification may be ignorable at the poststratum level, it may be important locally. The A.C.E. protected itself against the general problem by avoiding, when possible, post-stratum definitions based on variables with high reporting variability.

6. Failure of the A.C.E. Design and Concluding Remarks

In spite of being seemingly well designed and well executed, the A.C.E. failed to even approximately measure the coverage error in the 2000 U.S. Census. The chief reason seems to have been a failure of the assumption of consistent reporting of Census Day residence. In other words, depending upon when and where and with whom the interview was conducted two or more residences were reported as the correct one for a large number of people in sample.

We know that this happened because, after the both the census and the A.C.E. were completed, we were able to search and match nationally. This allowed us to search for census duplicates, even when the pair was miles apart. This was possible because, for the first time, practically all names in the census were data captured. (See Fay 2002; Mule 2001, 2002.) We could see, for example, how many of the people who were classified by the A.C.E. E–sample as "correctly enumerated" were also enumerated somewhere else, including at an other household or in a group quarters.

In one study, of the 1.3 million (weighted) E–sample people linked to a duplicate census enumeration outside the search area, only 14 percent were coded as erroneous enumerations by the A.C.E. (Feldpausch 2001, Table 1.) Since the A.C.E. E–sample was a random sample, one would expect that for any pair of duplicates it would pick up the erroneous enumeration roughly half the time.

Another 521 thousand E–sample cases (weighted) were linked to census enumerations in group quarters. Of these, only 31 percent were classified as erroneous by the A.C.E. (Feldpausch 2001, Table 3.) Roughly half, 271 thousand, of these linked E–sample cases were linked to an enumeration in a college dormitory. Under census residence rules, those living in a dormitory should be counted there, and not at home. However, the A.C.E. classified only 45 percent of these E–sample cases as erroneous enumerations. Since the proportions coded as correctly enumerated by the A.C.E. are significantly different from what would be reasonable, one must conclude that the A.C.E. had a strong tendency to misclassify enumeration status. Interestingly, many of these misclassified cases, the exact number is hard to determine, must have been A.C.E. matches. This is certainly due to the tendency of respondents to confirm people as living at an address who should be counted as living somewhere else.

We now have clear evidence that large number of parents of college students living in dormitories will consistently report their child as living at home even though census instructions clearly say not to. Further, both parents in a "joint custody" situation may consistently report the child as living in each of two households. Neighbors, no doubt, will report someone as "living there" who is in fact away at college, in the military, in jail, or at a second home. This misreporting occurred in spite of the numerous, detailed and specific probing questions about ususal residence asked by the A.C.E.

The extended search for census duplicates discussed above formed the principal evidence for A.C.E. error. However, other evidence was also gathered, including a re-interview study. These evaluations are discussed in detail in the Census Bureau's "Executive Steering Committee on A.C.E. Policy" (ESCAP) documentation. (See ESCAP I 2001, ESCAP II 2001).

The results of these evaluations is that the A.C.E. failed to correctly identify 4.7 million erroneous enumerations (U.S. Census Bureau 2003, page iv). In addition, it probably mis-identified the residences of large numbers of people in the P–sample, leading to both false matches and false non-matches. An extensive program by the Census Bureau of analysis and estimation produced the 1.3 million overcount estimate cited above. However, this program was uniquely tailored to the special circumstances of the 2000 post-census rematching, reinterviewing and duplicate search. Those interested are directed to U.S. Census Bureau (2003).

This paper has described the theory of the DSE, and has discussed how PES in general, and A.C.E. in particular, have implemented that theory. It has described the approximations necessary in real applications and the types of errors that can occur.

It discussed how carefully each of these approximations must be controlled. Obviously, the A.C.E did not successfully measure the large number of duplicates in the 2000 Census. Failure of even extensive probing questions to elicit accurate reports of ususal residence was the principal cause. However, the theory and design developed here should be of value in any future coverage measurement program.

Acknowledgements

This paper reports the results of research and analysis undertaken by Census Bureau staff. The opinions expressed are those of the author and do not necessarily reflect those of the Census Bureau.

References

Belin, T.R. (1993). Evaluating sources of variation in record linkage through a factorial experiment. *Survey Methodology*, 19, 13-29.

Cantwell, P., and Ikeda, M. (2003). Handling missing data in the 2000 accuracy and coverage evaluation survey. *Survey Methodology*, 29, 2, in press.

Childers, D. (2001). Accuracy and Coverage Evaluation: The Design Document. DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1 (Revised).

ESCAP I (2001). Report of the Executive Steering Committee for Accuracy and overage Evaluation Policy. March 1, 2001. (See www.census.gov/dmd/www/pdf/Escap2.pdf)

ESCAP II (2001). Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses. October 17, 2001. (See www.census.gov/dmd/www/pdf/Recommend2.pdf)

ESCAP II (2001). Census Person Duplication and the Corresponding A.C.E. Enumeration Status. Executive Steering Committee for A.C.E. Policy II, Report 6.

Fay, R. (2002). Probabilistic models for detecting census person duplication. *Proceedings of the Survey Research Methods Section*, American Statistical Association.

Feldpausch, R. (2001). ESCAP II: Census Person Duplication and the Corresponding A.C.E. Enumeration Status. Executive Steering Committee for A.C.E. Policy II, report 6.

Gonzalez, M. (1973). Use and evaluation of synthetic estimators. *Proceedings of the Social Statistics Section*, American Statistical Association, 73, 7-15.

Gonzalez, M., and Hoza, C. (1978). Small-area estimation with application to unemployment and housing estimates. *Journal of the American Statistical Association*, 73, 361, 7-15.

Griffin, R. (2000). Accuracy and Coverage Evaluation Survey: Dual System Estimation. DSSD Census 2000 Procedures and Operations Memorandum Series Q-20.

Haines, D. (2001). Accuracy and Coverage Evaluation Survey: Computer Specifications for Person Synthetic Estimation (U.S.) Re-issue of Q-30. DSSD Census 2000 Procedures and Operations Memorandum Series Q-46.

Hogan, H. (1992). The 1990 Post-Enumeration Survey: An overview. *The American Statistician*, 46, 261-269.

Hogan, H. (1993). The Post-Enumeration Survey: Operations and Results. *Journal of American Statistical Association*, 88, 423.

Hogan, H. (2001). Accuracy and Coverage Evaluation Survey: Effect of Excluding 'Late Census Adds'. DSSD Census 2000 Procedures and Operations Memorandum Series Q-43. http://www.census.gov/ dmd/www/pdf/Q-43.pdf

Marks, E.S., Seltzer, W. and Krotki, K.J. (1974). *Population Growth Estimation*. New York: Population Council.

Marks, E.S. (1979). The Role of Dual System Estimation in Census Evaluation. In Recent Developments in PGE, (K. Krotki). University of Alberta Press. 156-188.

Mule, T. (2001). ESCAP II: Person Duplication in Census 2000. Executive Steering Committee for A.C.E. Policy II, Report 20.

Mule, T. (2002). Further Study of Person Duplication Statistical Matching and Modeling Methodology. DSSD A.C.E. Revision II Memorandum Series PP-51.

Nash, F.F. (2001). ESCAP II: Analysis of Census Imputations. Executive Steering Committee for A.C.E. Policy II, Report 21.

Petersen, C.G.J. (1896). The Yearly Immigration of Young Plaice into the Limfjord from the German Sea. Report of the Danish Biological Station. 6, 1-48.

Raglin, D. (2002). ESCAP II: Effect of Excluding Reinstated Census People from the A.C.E. Person Process. Report 13, http://www.census.gov/dmd/www/pdf/Report13.PDF

Robinson, J.G., Ahmed, B., Das Gupta, P. and Woodrow, K. (1993). Estimates of population coverage in the 1990 united states census based on demographic analysis. *Journal of the American Statistical Association*. 88, 1061-77.

Robinson, J.G. (2001). ESCAP II: Demographic Analysis Results. Executive Steering Committee for A.C.E. Policy II, Report 1.

Schenker, N. (1988). Handling missing data in coverage estimation with application to the 1986 test of adjustment related operations. *Survey Methodology*, 14, 87-97.

Sekar, C.C., and Deming, W.E. (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association*, 44, 101-115.

U.S. Census Bureau (1985). Evaluating Census of Population and Housing, Statistical Training Document, ISP-TR-5, Washington, D.C.

U.S. Census Bureau (2000). Statement on the Feasibility of Using Statistical Methods to Improve the Accuracy of Census 2000.

U.S. Census Bureau (2003). Technical Assessment of A.C.E. Revision II, March 12, 2003, http://www.census.gov/dmd/www/pdf/ACETechAssess.pdf

Wetrogan, S.I., and Cresce, A.R. (2001). ESCAP II: Characteristics of Census Imputations. Executive Steering Committee for A.C.E. Policy II, Report 22.

Wolter, K.M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, 81, 338-346.