THE ADMINISTRATIVE RECORDS EXPERIMENT IN 2000: AN APPLICATION TO POPULATION COUNT ESTIMATION VIA TRIPLE SYSTEMS ESTIMATION

Jana Asher and Stephen E. Fienberg, Carnegie Mellon University *
Jana Asher, Dept. of Statistics, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213

Key Words: Accuracy and Coverage Evaluation survey, Administrative records, Decennial census counts, Multiple systems estimation.

Abstract: The U.S. Census Bureau has a long track-record of efforts to assess the accuracy of decennial census counts, and has, with the help of others, investigated alternatives to traditional census taking such as the use of administrative records. In past censuses and census tests, the Bureau has also used administrative records to supplement traditional counting methods and to explore the possible uses of multiple systems estimation. This paper describes the use of a three-way cross-classification of the the Administrative Records Experiment (AREX) 2000 bottom-up (BARCUF) data file, Census 2000, and the Accuracy and Coverage Evaluation Survey (ACE) to create block-level triple systems population estimates within the five AREX 2000 counties. We present a statistical approach in which the AREX 2000 and Census 2000 results are considered the primary population lists available for all blocks, and the ACE results are treated as an auxiliary random sample of blocks from a third population list.

1 Introduction

Going back at least to the 1940 decennial census, the U.S. Census Bureau has put into place a variety of methods for assessing the accuracy and coverage of the decennial census counts, e.g., using demographic analysis and post-enumeration surveys (e.g., see Anderson, 2000). For the 2000 decennial census, this assessment took the form of the Accuracy and Coverage Evaluation Survey (ACE), a survey of approximately 314,000 housing units intended to find both duplicates and missing households in the decennial census. In March 2001, however, the Executive Steering Committee for A.C.E. Policy (ESCAP) of the U.S. Census Bureau recommended that unadjusted census numbers be used for redistricting, and in October of 2001, ESCAP further recommended that unadjusted Census 2000 numbers be used for purposes other than redistricting, stating that the A.C.E. failed to uncover approximately 3 million erroneous enumerations in the census, causing an over-count of the population.

In a separate effort, the Bureau has investigated alternatives to traditional census taking such as the use of administrative records[1] either as a substitute for or as a supplement to the traditional enumeration. Various panels of the National Research Council have encouraged these efforts and emphasized the important role administrative records could play as part of the decennial census effort (e.g., see Cohen, White, and Rust, 1999, pp. 89-91), as has the U.S. Census Monitoring Board (2001) and congressional committees. Administrative records have also been the focus of efforts by national statistical offices in other countries (e.g., see Scheuren, 1999). Other efforts to explore the use of administrative records in decennial census operations (e.g., see Zanutto and Zaslavsky, 2001) have been intriguing but inevitably have suggested the need for more careful research and followup.

In response to continued interest in the roll administrative records can play in existing survey programs, the U.S. Census Bureau through its Planning, Research, and Evaluation Division (PRED), developed a program to produce, on an annual basis, an administrative records "superlist" called the Statistical Administrative Records System (StARS). StARS is the end result of the careful merging of six or more administrative records sources; the first production of StARS occurred in 2000 using 1999 data. As part of the Census Bureau's evaluation program for Census 2000, PRED developed an administrative records experiment (AREX 2000) in which the feasibility of using StARS for either an administrative

*The authors acknowledge the support and assistance of the Planning, Research and Evaluation Division of the U.S. Census Bureau and the financial support of the Bureau through a contract with the Research Triangle Institute.

[1]By administrative records, we mean government data collected for regulatory or transactional purposes. Examples of administrative records data include state databases such as driver's license records and national databases such as Internal Revenue Service tax records.

records census[2] or to augment existing decennial census practices was explored. As part of AREX 2000, extensive verification and testing, as well as additional processing of the 1999 StARS dataset, was performed for the data for five counties in Colorado and Maryland. Further information on StARS and AREX 2000 can be found in Asher and Fienberg (2001) or in the other papers presented during this session.

In this paper, we outline methodology for estimating a population count at the block level via multiple systems estimation (MSE) using results from the 2000 decennial census, ACE data, and the AREX 2000 "superlist." In this case, two of the lists cover every block (census and AREX 2000), and one list covers a sample of blocks (ACE). Several papers have addressed the issue of population estimation in the case of three or more lists (e.g. Zaslavsky and Wolfgang, 1993), but have not necessarily addressed the issue of missing data within one or more of the lists. Creating models in a situation where there is at least one list that only covers a sample of blocks is addressed via a hierarchical Bayesian approach in Zaslavsky (1989). Zaslavsky's models, however, do not address the issue of lists in which the missingness is not ignorable (Rubin, 1976). There is therefore a need to explore multiple systems estimation in the case of non-ignorable missingness.

In the remainder of this paper, we first briefly describe the basic capture-recapture model and the assumptions inherent to that model. Because we are treating the ACE sample as a "missing data" problem, we then describe previous research and theorems related to estimation where there is missing data. Finally we develop models specifically for estimating population counts from the three-way Census/AREX/ACE dataset that allow for non-ignorable missingness in a multiple systems estimation framework.

2 Background Methodology

2.1 Multiple systems estimation

We provide a detailed discussion of multiple systems estimation in Asher and Fienberg (2001) and summarize pieces of that discussion here. The simplest version of this methodology, dual systems estimation, relies on three assumptions: 1) independence of the captures (lists), 2) homogeneous probability of capture in the population of interest, and 3) error-free methods of matching across captures. If these three conditions are met, then a table displaying a cross-classification of population counts for the two lists is as follows:

		List 2 In	List 2 Out	Total
List 1	In	x_{11}	x_{10}	x_{1+}
	Out	x_{01}	x_{00}	x_{0+}
	Total	x_{+1}	x_{+0}	$x_{++} = N$

Here x_{00} represents the count of members of the population that are not captured by either list; the goal is to estimate N, the total number of members of the population. This traditional estimator for N is:

$$\hat{N} = \lfloor \frac{x_{1+}x_{+1}}{x_{11}} \rfloor$$

where $\lfloor \star \rfloor$ is the greatest integer $\leq \star$. Henceforth we simply assume all such estimates are rounded down to the nearest integer.

In practice, the three assumptions outlined above are rarely met. If three lists are available, log-linear modeling techniques can be applied that can account for dependency of the lists. Let x_{ijk}, $i, j, k \in \{0, 1\}$ represent an observed count in a three-way cross-classification table of population counts for three lists. In this notation, x_{000} is the count of those members of the population that are not included on any of the three lists, whereas each of the other 7 counts are in fact observable. Taking these 7 counts together, we observe a total of $n = x_{111} + x_{101} + x_{011} + x_{001} + x_{110} + x_{100} + x_{010}$ persons. Finally, the population total is $N = n + x_{000}$.

Our goal is to estimate x_{000}. As such, we want a model for the observable cells that we then project to the unobserved cell. We do this through a log-linear representation for the expected counts, $m_{ijk} = E\{x_{ijk}\}$:

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$$

with constraints on the u-terms, e.g., that they add to zero across any subscript. This log-linear model is the standard no-second order interaction model, or, in other words, the model that allows for dependency between pairs of lists but not three-way list dependency. But because we are applying it initially to only the 7 observable expected cell counts, the model is in effect "saturated" and fits the data perfectly, i.e., the maximum likelihood estimates for the expected counts are $\hat{m}_{ijk} = x_{ijk}$ for all i, j, k.

Within this framework, we can also fit either reduced models to the data from three cross-classified lists which include only certain pair-wise dependencies or the model of complete independence. Again we estimate the expected cell values under the model and we

[2] An administrative records census is a population count using administrative records only. For more information, see Scheuren (1999).

project the model to the missing $(0, 0, 0)$ cell. For both the saturated and the reduced models we can write the estimate of N, $\hat{N}$, as:

$$\hat{N} = n + \frac{\hat{m}_{111}\hat{m}_{100}\hat{m}_{010}\hat{m}_{001}}{\hat{m}_{110}\hat{m}_{101}\hat{m}_{011}}, \quad (1)$$

The simplest model for three lists that allows for both list dependencies and heterogeneity of capture probabilities is based on models developed by Georg Rasch for scoring examination items in educational testing. For a three-way cross-classification, let $\pi_{k_1k_2k_3}$ be the probability of observing a count in the cell ($k_1k_2k_3$), $k_j \in \{0, 1\}$; in this case $m_{k_1k_2k_3} = N\pi_{k_1k_2k_3}$. Then, under the Rasch model:[3]

$$\log(\pi_{k_1k_2k_3}) = \alpha + k_1\beta_1 + k_2\beta_2 + k_3\beta_3 + (\sum_{j=1}^{3} k_j)^2\gamma \quad (2)$$

In this model, the β terms capture the list dependencies, and the γ term is a measure of the capture heterogeneity. Note that the value of γ is not affected by permutations of k_1, k_2, k_3 and hence we have a quasi-symmetry model (e.g., see Bishop, et al., 1975). This transformed version of the Rasch model ignores moment inequalities implicit in the underlying Rasch model as described in Cressie and Holland (1983) and Darroch, et al. (1993). See Asher and Fienberg (2001) for further details.

The assumption that the value of γ is not affected by permutations of k_1, k_2, k_3 may not hold well. This assumption is equivalent to the belief that the distribution of capture probabilities across individuals is the same for each list. If we believe that this distribution is different for subsets of the lists, we will need to relax this assumption. A vehicle for doing so is through a partial quasi-symmetry model, which, for 3 lists, takes the form:

$$\log(\pi_{k_1 \dots k_J}) = \alpha + k_1\beta_1 + k_2\beta_2 + k_3\beta_3 + \gamma(k_1 + k_2, k_3)$$

where

$$\gamma(s_1, s_2) = \log E\left[e^{(s_1\theta_1 + s_2\theta_2)} | \mathbf{k} = 0\right].$$

In this case, θ_1 represents the random individual effect for lists 1 and 2, and θ_2 represents the random individual effect for list 3; θ_1 and θ_2 are assumed to follow different underlying distributions. Mathematical details for the derivation of the partial quasi-symmetry model can be found in Darroch, et al. (1993).

In our application, both census and AREX 2000 data is available for all blocks, and ACE data is available for a random subset of blocks. All of the models described above, while applicable, need to be modified to account for the missing ACE data. In the next section, we describe methodology for addressing missingness.

2.2 The case of missing data

Missing data can be thought of as unintentionally or intentionally missing, in a pattern that is random or non-random. In our problem; there is a specific sample design from which the final sample of ACE blocks is derived. The missing blocks are purposefully selected, and there is no "non-response" in the sense that data are collected for every ACE block. The data for ACE are intentionally missing and random for the non-ACE blocks, but result from a complex sampling design that must be taken into account in the development of models for population estimation.

Rubin (1976) provides a general framework for handling missing data problems in both likelihood and Bayesian inference which we utilize here. We define M to be a missing value indicator with parameter ϕ and U to be the random variable of interest with pdf $f_\theta(u)$. Then V is the observable random variable. For an observation, $v = u$ if $m = 1$, and v is missing if $m = 0$. Then, taking V to be a vector of n iid random variables, we partition the vector into two subsets, $v_{(0)}$, which is missing, and $v_{(1)}$, which is equal to $u_{(1)}$. Using the observed data $\tilde{v} = (\tilde{v}_{(0)}, \tilde{v}_{(1)})$, we define the following vectors:

$$\begin{aligned} \tilde{m} &= (\tilde{m}_{(0)}, \tilde{m}_{(1)}) \\ \tilde{u} &= (u_{(0)}, \tilde{u}_{(1)}) \end{aligned}$$

Let $g_\phi(m|u)$ be the conditional distribution of M given the unobserved vector u.[4] Then the missing data are *missing at random* if, for each ϕ, $g_\phi(\tilde{m}|\tilde{u})$ is constant for all $u_{(0)}$, and ϕ as being *distinct from* θ if their joint parameter space factorizes into a ϕ-space and θ-space and/or the prior distributions for ϕ and θ are independent. Then if the data are missing at random and ϕ is distinct from θ, both likelihood-based inference and also Bayesian inference can ignore the missing data and be based on the observed data only (Rubin, 1976; Little and Rubin, 1987).

[3] Mathematical details for the derivation of (2) are in Cressie and Holland (1983), Fienberg and Meyer (1983), Holland (1990), Darroch et al. (1993), and Fienberg, Johnson, and Junker (1999).

[4] Note that the marginal distribution $f_\phi(m)$ is of little interest at this point, as any dependence the "missingness" has on the underlying data is intergrated out.

If, however, either the data are not missing at random or ϕ and θ are not distinct, then the joint likelihood for ϕ and θ is given as

$$L(\theta, \phi|\tilde{v}) = \int f_\theta(\tilde{u}) g_\phi(\tilde{m}|\tilde{u}) du_{(0)}$$

or the joint posterior distribution of ϕ and θ is proportional to

$$p(\theta)p(\phi|\theta) \int f_\theta(\tilde{u}) g_\phi(\tilde{m}|\tilde{u}) du_{(0)}.$$

A question of interest then becomes whether the ACE sample is truly missing at random, and if the mechanism by which the missingness occurs is distinct from the population estimation problem. The primary sampling unit for the ACE was a block cluster. As a first step, block clusters were partitioned into three strata by size: 0-2 housing unit (small) block clusters, 3-79 housing unit (medium) block clusters, and (large) block clusters with over 80 housing units. American Indian Reservation block clusters formed a fourth stratum. A systematic sample was taking within each stratum, leading to a total sample of approximately two million housing units. The sample was then reduced as follows: medium and large block clusters were stratified again by estimated demographic composition and differences in housing unit counts between the ACE listing and census address list. The result was five sub-strata within each state, which were differentially subsampled. Small block clusters were subsampled according to sub-strata as well, but oversampled in comparison to the medium and large block clusters. American Indian Reservation block clusters were not subsampled.

The blocks not included in the ACE sample are clearly missing at random. In this case, ϕ can be considered a vector of Bernoulli parameters, where ϕ_i = ϕ_j if block clusters i and j are within the same sub-stratum. u is a vector of population counts, and θ a vector of expected population counts. Within each combination of strata and sub-strata, block clusters have equal probability of selection, therefore the missingness is unrelated to the actual population count of the block cluster. But the second assumption required for ignorable missingness is not met; ϕ and θ are not distinct. To see this, simply note that block clusters with a small number of housing units, and therefore a small population, are more likely to be sampled than other block clusters. Therefore if θ_i is large, ϕ_i is relatively small, and visa versa. Any modeling strategy adopted for these data will therefore need to explicitly address the missingness of the ACE data. One possibility is to model each sub-stratum separately; in this case the missingness could be ignored. Another possibility is to incorporate the covariate information used to develop the strata and sub-strata into the modeling procedure. We discuss these options further in the next section.

3 Methodology

3.1 Developing MSE models for stratified data with ignorable missingness

Let i be a subscript representing inclusion ($i = 1$) or exclusion ($i = 0$) in the census, j be a subscript representing inclusion or exclusion in AREX 2000, k be a subscript representing inclusion or exclusion in ACE, and l be a subscript indicating block (with L blocks total). The situation is that we have data for every block for the census and AREX 2000, but not for ACE. As a results, for $l = 1$ to L blocks, we have x_{10+}, x_{01+}, and x_{11+}. For the subset of blocks for which there is ACE data available, we also have x_{001}, x_{011}, x_{101}, and x_{111}. We wish to balance differences between blocks with the need to borrow strength across them to account for the missingess. To do so, within sub-strata, we build a log-linear model as follows.

First, assume no missing data. A saturated log-linear model would be:

$$\begin{aligned} log(m_{ij0(l)}) &= \alpha_{(l)} + \sum \beta_{ij0(l)} \\ log(m_{ij+(l)}) &= 2\alpha_{(l)} + \sum \beta_{ij1} + 2\sum \beta_{ij0(l)} \end{aligned}$$

where the subscript ijk indicates that β_{ijk} (or $\beta_{ijk(l)}$) is an interaction parameter if $i + j + k > 1$. In this case, expected counts to which ACE data contributes will have a common dependency structure (given by the β_{ij1}'s), while expected counts to which ACE data does not contribute will have block-specific list effects and dependencies (given by the $\beta_{ij0(l)}$'s). These two types of expected counts are connected within the model by a block-effect parameter. Table 2 shows a design matrix for this model in the case of two blocks. For block l, under the assumption of random missingness, the $m_{ij1(l)}$ rows may simply be dropped; the combination of the block-effect parameter and the common β_{ij1}'s across blocks make estimation of these expected values possible given particular constraints on the ratio of missing to non-missing ACE blocks.

For the number of columns in the design matrix in Table 1 to be larger than the number of rows, it must be true that $7(L-L')+3L' \geq 4+4L$ where L is the total number of blocks, and L' is the number of blocks for which ACE sample is missing. For large L, this suggests that there should be more non-missing blocks than missing blocks, which is impossible within the context of this problem. We must reduce the number of parameters to

Table 1: Example "design" matrix for 2 complete blocks, "saturated" log-linear model

	β_{001}	β_{011}	β_{101}	β_{111}	$\alpha_{(1)}$	$\beta_{100(1)}$	$\beta_{010(1)}$	$\beta_{110(1)}$	$\alpha_{(2)}$	$\beta_{100(2)}$	$\beta_{010(2)}$	$\beta_{110(2)}$
$\log(m_{01+(1)})$	1	1	0	0	2	0	2	0	0	0	0	0
$\log(m_{10+(1)})$	1	0	1	0	2	2	0	0	0	0	0	0
$\log(m_{11+(1)})$	1	1	1	1	2	2	2	2	0	0	0	0
$\log(m_{001(1)})$	1	0	0	0	1	0	0	0	0	0	0	0
$\log(m_{011(1)})$	1	1	0	0	1	0	1	0	0	0	0	0
$\log(m_{101(1)})$	1	0	1	0	1	1	0	0	0	0	0	0
$\log(m_{111(1)})$	1	1	1	1	1	1	1	1	0	0	0	0
$\log(m_{01+(2)})$	1	1	0	0	0	0	0	0	2	0	2	0
$\log(m_{10+(2)})$	1	0	1	0	0	0	0	0	2	2	0	0
$\log(m_{11+(2)})$	1	1	1	1	0	0	0	0	2	2	2	2
$\log(m_{001(2)})$	1	0	0	0	0	0	0	0	1	0	0	0
$\log(m_{011(2)})$	1	1	0	0	0	0	0	0	1	0	2	0
$\log(m_{101(2)})$	1	0	1	0	0	0	0	0	1	1	0	0
$\log(m_{111(2)})$	1	1	1	1	0	0	0	0	1	1	1	1

something reasonable for the data at hand; specifically, the number of block-specific parameters, if reduced by one per block, allows the model to operate with two non-missing blocks and an arbitrarily large number of missing blocks. In other words, the model will work if the number of block-specific parameters is limited to the number of non-missing datapoints for every block (there are three such datapoints). Removing the block effect is out of the question, as this is what is used to link the blocks without ACE data to the data available for that block. So one (or all three) of the other parameters must become identical for each block for the model to work. It may not be plausible to assume a fully saturated model for these data in any case. Modifying the log-linear model by dropping terms equates to removing columns from the design matrix in Table 2 with care to avoid creating a non-hierarchical model. Luckily, the $\beta_{110(k)}$ parameters will be frequently dropped, as we believe there is little dependency between AREX 2000 and the decennnial census.

Rasch-like models can be considered in this framework quite easily. Again, we begin in the frame that there is no missing data. Then:

$$\begin{aligned} log(m_{ij0(l)}) &= \alpha_{(l)} + \sum \beta_{ij0(l)} + \gamma_{(l)} \\ log(m_{ij+(l)}) &= 2\alpha_{(l)} + \sum \beta_{ij1} + 2\sum \beta_{ij0(l)} + 2\gamma_{(l)} \end{aligned}$$

Note that the symmetry term is within block, and missingness will present the same problems for this model as for the fully saturated one, due to a greater number of parameters than datapoints for blocks without ACE sample. Again, the solution is to allow at least one parameter to be constant across blocks.

Results from several of the models described thus far can be compared within a model selection procedure that relies on goodness-of-fit and model parsimony. The missingness for these data will result in a model selection problem that will test different configurations of dependency and heterogeneity relationships *mixed* between lists mixed with block dependency, as parameters are moved from block-specific to non-block-specific and potentially dropped altogether. Whatever pattern of parameters prevails, the potential models will all be hierarchical, with the block parameters as random effects and the β's as fixed effects. Additionally, in the Rasch-like models, the γ parameters are considered random effects at the individual level. As a result, standard statistical software for generalized linear models is ill-prepared to incorporate either the moment constraints for the heterogeneity term nor the uncertainty of the block effect. Expanding the hierarchical model into full Bayesian format will allow us to address these issues.

To develop the Bayesian model, we choose the following framework:

$$\begin{aligned} X_{(l)} &\sim \text{Multinomial}(\sum_{ijk} e^{\alpha_{(l)}+\sum \beta_{ijk(l)}}, \\ & \qquad p_{ijk(l)} = \frac{e^{\alpha_{(l)}+\sum \beta_{ijk(l)}}}{\sum_{ijk} e^{\alpha_{(l)}+\sum \beta_{ijk(l)}}}) \\ \alpha_{(l)} &\sim \text{Normal}(\mu_\alpha, \sigma^2_\alpha) \\ \beta_{ijk(l)} &\sim \text{Normal}(\mu_{ijk}, \sigma^2_{ijk}) \end{aligned}$$

The assumption that the β terms follow common prior distributions across blocks removes the concerns about identifiability of the model, assuming a sufficient prior.

3.2 Developing MSE models with covariates for data with non-ignorable missingness

Covariate information derived about the housing units within the blocks is used in the stratification of the ACE sample. By regressing these characteristics on the logit of ϕ within a Bayesian framework, we can account for the missingness within the model. We let M be the vector of missing value indicators, and $Y_{(l)}$ be a vector of covariates for block l. Then:

$$\begin{aligned}
X_{(l)} &\sim \text{Multinomial}(\sum_{ijk} e^{\alpha_{(l)}+\sum \beta_{ijk(l)}}, \\
&\quad p_{ijk(l)} = \frac{e^{\alpha_{(l)}+\sum \beta_{ijk(l)}}}{\sum_{ijk} e^{\alpha_{(l)}+\sum \beta_{ijk(l)}}}) \\
M_{(l)} &\sim \text{Bernoulli}(\frac{e^{\tau Y_{(l)}}}{1+e^{\tau Y_{(l)}}}) \\
\alpha_{(l)} &\sim \text{Normal}(\mu_\alpha, \sigma^2_\alpha) \\
\beta_{ijk(l)} &\sim \text{Normal}(\mu_{ijk}, \sigma^2_{ijk}) \\
\tau &\sim \text{Normal}(\mu_\tau, \sigma^2_\tau)
\end{aligned}$$

To implement this model, the distribution of $m_{(l)}x_{ij1(l)}$ is used, and the covariate information and combined priors again makes the model "identifiable" in a Bayesian sense.

4 Future Work

At this time, the creation of a three-way cross-classification of the decennial census, ACE, and AREX 2000 is nearly complete. Our future work will involve fitting these models to this dataset and refining our methodology.

References

Anderson, Margo J., ed. (2000). *Encyclopedia of the U.S. Census.* Congressional Quarterly Press, Washington, DC.

Asher, Jana L. and Fienberg, Stephen E. (2001). Statistical Variations on an Administrative Records Census. *Proceedings of the Annual Meeting of the American Statistical Association*, 2001.

Bishop, Yvonne M. M. and Fienberg, Stephen E. and Holland, Paul H. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA.

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

Cressie, Noel and Holland, Paul W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, **48**, 129–141.

Cohen, Michael L., White, Andrew A., and Rust, Keith F., eds. (1999). *Measuring a changing nation: Modern Methods for the 2000 Census.* National Academy Press, Washington, DC.

Darroch, John N., Fienberg, Stephen E., Glonek, Gary F. V., and Junker, Brian W. (1993). A Three-Sample Multiple-Recapture Approach to Census Population Estimation With Heterogeneous Catchability. *Journal of the American Statistical Association*, **88**, 1137–1148.

Fienberg, Stephen E., Johnson, Mathew S., and Junker, Brian W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society, Series A*, **162**, 383–405.

Little, Roderick J.A., Rubin, Donald B. (1987). *Statistical Analysis With Missing Data (Wiley Series in Probability and Mathematical Statistics)*. Wiley, New York.

Rubin, Donald B. (1976). Inference and Missing Data. *Biometrika*, **63**, 581–592.

Scheuren, Fritz (1999). Administrative Records and Census Taking. *Survey Methodology*, **25**, 151–160.

U.S. Bureau of the Census Executive Steering Committee for Accuracy and Coverage Evaluation Policy (2001). *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses.*

U.S. Bureau of the Census Executive Steering Committee for Accuracy and Coverage Evaluation Policy (2001). *Report of Tabulations of Population to States and Localities. Federal Register*, Vol. 66, No. 46, March 8, 2001, 14004-14046. `http://www.census.gov/dmd/www/EscapRep.html`).

U.S. Census Monitoring Board, Congressional Members (2001). *Report to Congress, September 1, 2001.*

Zanutto, Elaine, and Zaslavsky Alan M. (2001). Using Administrative Records to Impute for Nonresponse. *Survey Nonresponse*, edited by R. Groves, R.J.A. Little, and J. Eltinge. New York: John Wiley.

Zaslavsky, Alan M. (1989). *Representing the census undercount: reweighting and imputation methods.* Unpublished Ph.D. thesis, Massachusetts Institute of Technology.

Zaslavsky, Alan M. and Wolfgang, Glenn S. (1993). Triple-System Modeling of Census, Post-Enumeration Survey, and Administrative-List Data. *Journal of Business and Economic Statistics*, **11**, 279–288.