EVALUATION OF CENSUS POPULATION COVERAGE BY NETWORK SURVEYS

Monroe G. Sirken and Barry J. Graubard, National Center for Health Statistics, and
Richard W. LaValley, Bureau of the Census

1. INTRODUCTION

The Census Bureau is currently testing and developing the post-enumeration survey methods that it will use to evaluate the completeness of population coverage in the 1980 census. Persons enumerated in this survey, which is conducted soon after the census, are matched with the census enumerations to determine how many of the same people were counted in both. This information is used by a dual system estimator to revise the census population count and to estimate the number of persons missed in the census. The dual system estimator is a consistent estimator of the number of persons missed by the census providing that whether or not a person is missed in the post-enumeration survey is independent of whether or not he had been missed in the census.

It is suspected, however, that the censuses and the post-enumeration surveys of the past have been correlated (i.e., that is both data collection systems have tended to miss the same people). A basis for this concern is the fact that the demographic cohort survival estimates of the number of persons missed in recent censuses were substantially larger than those based on post-enumeration surveys. Revised methods of conducting the post-enumeration survey are being investigated, with the view to reducing the correlation bias of the dual system estimators.

The network survey represents a new approach for designing post-enumeration surveys that is currently being investigated by the Census Bureau. The network sample survey is a relatively new type of survey design that was first proposed [1,3] about a decade ago. The unique feature of the network survey is its counting rule which does not restrict a person from being enumerated more than once in the survey. The network design has been successfully applied to household surveys to estimate disease prevalence [7], as well as birth, death and marriage rates [2,6,8]. The network survey has also been previously proposed [4,5] as one of the data systems for producing dual system estimators of death registration completeness and dual system estimators of census population coverage. The statistical model of dual system estimators presented in these earlier papers [4,5] served as the model for designing a feasibility test of the post-enumeration multiplicity survey that was recently conducted in conjunction with the Oakland, California dress rehearsal for the 1980 census.

Consanguine counting rules were adopted in the Oakland test of the post-enumeration multiplicity survey. According to these rules, a person is eligible to be enumerated at a household if either he or one or more of his close relatives lives there. The hypothesis which motivated the Oakland study is that the correlation bias of the dual system estimator of census undercoverage would be smaller if a person was enumerated at different households instead of at the same household in the post-enumeration survey and in the census. If this hypothesis is valid, the correlation bias would be smaller if the post-enumeration survey was based on a consanguine counting rule rather than on a de jure residence rule. This hypothesis is not testable at this time. This paper is more limited in scope. First we contrast the design features of the conventional post-enumeration survey (PES) and the post-enumeration multiplicity survey (PEMS). Next, we describe the design of the Oakland PEMS and, finally, we discuss one of the implementation problems detected in that study. In the statistical appendix, we define the PES and PEMS dual system estimators and analyze the effect of correlation bias.

2. COMPARISON OF PES AND PEMS

PES and PEMS are both household sample surveys that are conducted after the census. Except for the counting rule, the design features of both surveys are virtually the same. The PES adopts a de jure residence rule, and the PEMS adopts a multiplicity counting rule. The de jure residence rule specifies that people are eligible to be enumerated only at their usual places of residence. In other words, PES adopts the same counting rule that is used in the census. On the other hand, the multiplicity counting rule adopted by PEMS specifies that people are eligible to be enumerated at the households of specified close relatives as well as at their own de jure residences. Consequently, a person is eligible to be enumerated at one and only one household in PES, but he is eligible to be enumerated at one or more households in PEMS, depending on the number of his relatives' residences.

The counting rule difference between PES and PEMS implies important differences between the two surveys in the items of information that are collected in the household interviews. Both surveys count the persons that are enumerated at their de jure residences. In addition, the PEMS

(1) counts the household residents' close relatives who live elsewhere, and

(2) ascertains a counting rule weight for every household resident and for each of their enumerated relatives.

The counting rule weights are needed to adjust for the number of times a person is eligible to be enumerated in compliance with the counting rule adopted in the post-enumeration survey. For instance, knowing the number of a person's relatives not living with him is sufficient information to determine the person's counting rule weight. Since the de jure residence rule specifies that every person is enumerable once and only once, the PES estimator does not require collecting the ancillary information listed above.

Both of the PES and PEMS require the census addresses for the people enumerated in the post-enumeration survey. The census address (i.e., address on the census date) is essential because it is a matching item for determining which of the persons enumerated in the survey were also enumerated in the census. This information is readily available for household residents whose addresses were the same in the census and in the post-enumeration survey. When the addresses are different, special efforts are required to obtain the census address. Thus, the census addresses would have to be collected in both the PES and the PEMS for de jure residents who changed their addresses between the census and the post-enumeration survey. Additionally, the PEMS must collect the census addresses for the persons it enumerates at the households of relatives.

For example, if PEMS was based on a counting rule that linked people to their de jure residences as well as their siblings who do not live with them, the PEMS questionnaire would collect the following items of information for every household member.

(1) Where were you living on the census date?

[ ] Here

[ ] Elsewhere→(2) What was your former address?

(3) Do you have any siblings that do not live here?

[ ] No

[ ] Yes→(4) How many?

(5) Where were your siblings living when the census was taken?

Questions (1) and (2) would be asked in both the PES and the PEMS, but Questions (3), (4) and (5) would be asked only in the PEMS.

Since the PEMS involves the collection of supplementary survey data, the survey costs and the respondent burden would be greater for the PEMS than for the PES. The compensatory factor may be that the PEMS estimates of census undercounts are subject to smaller errors than the PES estimates. Since missed people are eligible to be enumerated at more households in PEMS than in PES, it is very likely that the PEMS estimates would be subject to substantially smaller sampling errors. Also, they are probably subject to less correlation bias for the reasons cited earlier. On the other hand, the PES estimates are immune to the measurement errors associated with the items of information that PEMS collects to determine the counting rule weights and to conduct the census match. It remains to be determined how the cost and error effects of the estimates of population coverage relate to the type of counting rule that is adopted in the post-enumeration survey.

3. DESIGN OF THE OAKLAND PEMS

The Oakland PEMS was a two-stage enumeration process. The first stage was based on a stratified cluster sample of 250 households that was selected from the census listing of the persons used in enumerating in the Oakland census about six months earlier. The second stage was based on a sample of about 350 Oakland addresses of the relatives of persons who were enumerated at the first stage addresses. The first stage enumerations collected the information required by the PEMS counting rules. The second stage enumerations were conducted to verify the information reported in the first stage enumerations. It was especially important to verify the census addresses of the relatives reported in the first stage since this information would be of critical importance in deciding whether or not the relatives had been enumerated in the Oakland census.

Some of the basic items of information that were collected in the Oakland PEMS are listed in Figure I. These items were collected from every resident of the stage 1 household. Items A1-A4 enumerated the people that the resident person was eligible to report in compliance with the PEMS counting rules, and Items C1-C4 collected the information required to determine the counting rule weights of every enumerated person. It should be noted that items A1-A4 refer to people who were eligible to be enumerated in the census because they lived in Oakland on the census data, and items C1-C4 refer to people who are eligible reporters in the PEMS because they lived in Oakland when PEMS was conducted. Item B1 determined the census addresses for the enumerated persons.

Figure 1: Types of Questions Asked in Oakland PEMS

Each resident person in HH was asked the following:

A1-Did you live in Oakland on the census date?
A2-How many of your parents lived in Oakland then?
A3-How many of your siblings lived in Oakland then?
A4-How many of your children lived in Oakland then?

For matching purposes, the following was asked for each person reported in A1-A4:

B1-Where in Oakland did (...) live on census date?

To determine multiplicities, the following were asked of each person reported in A1-A4:

C1-Does (...) live in Oakland now?
C2-How many of (...)'s parents now live in Oakland?
C3-How many of (...)'s siblings now live in Oakland?
C4-How many of (...)'s children now live in Oakland?

The eight counting rules that were investigaged in the Oakland PEMS are listed in the stub of Figure II. Each rule makes persons eligible to be enumerated at their de jure residences. The first listed counting rule, referred to in Figure II as the de jure rule, makes a person eligible to be enumerated once at his de jure place of residence, and makes him ineligible to be enumerated at any other household. Counting rules (2 - 8) make persons eligible to be enumerated at the households of specified relatives in addition to their de jure residences. These rules differ among themselves with respect to which relatives are specified. For instance, according to rule (6) a person would be eligible to be enumerated at his de jure residence and at all other households which are residences of one or more of his children and/or his parents. On the other hand, according to rule (2), a person would be enumerable at his de jure residence and the residences of his children but would not be enumerable at any other residences. Figure II shows the items of information in Figure 1 that are needed for the PEMS estimate associated with each counting rule. Thus, the first column of Figure II lists the items that identify the persons that are eligible to be counted by the rule and the second column lists the items that determine how many times each of the enumerated persons is eligible to be counted in compliance with the counting rule.

Figure II: Items of Information Required for Specified Counting Rule

Counting Rule	Count	Multiplicity
(1) de jure	A1	C1
(2) de jure/children	A1,A2	C1+C4
(3) de jure/sibling	A1,A3	C1+C3
(4) de jure/parent	A1,A4	C1+C2
(5) de jure/children/ sibling	A1,A2, A3	C1+C3+ C4
(6) de jure/children/ parent	A1,A2, A4	C1+C2+ C4
(7) de jure/sibling/ parent	A1,A3, A4	C1+C2+ C3
(8) de jure/children/ sibling/parent	A1,A2, A3,A4	C1+C2+ C3+C4

Midway into the Oakland PEMS another version of counting rules (2)-(7) was proposed by Patricia Royston, National Center of Health Statistics. Instead of linking a person to all of his specified relatives, she suggested that the counting rules be defined so that a person is only linked to those relatives not living with them. [It is feasible to investigate this proposal in the Oakland PEMS because the household listing contains the relationship of every resident person to the household.]

Two types of counting rule weights were investigated in the Oakland PEMS. The household counting rule weight is the inverse of the number of different households that contain either the person himself and/or one of his specified relatives. The person counting rule weight is the inverse of the number of persons, including the person himself and his specified relatives that are eligible to report him in compliance with the counting rule.

4. FINDINGS OF THE OAKLAND PEMS

One of the reasuring findings of the Oakland PEMS was the public's cooperation. There had been concern that respondents might find the questions about relatives confusing and that they might refuse to disclose the names and addresses of their relatives. On the contrary, the PEMS enumerators reported that the consanguine network concept was readily accepted by the respondents. Only one of the 221stage 1 households refused outright to cooperate and less than 10 other households refused to provide the requested information after they reported a relative living elsewhere in Oakland. However, the addresses reported for relatives were often incomplete so that they could not be located in the second stage enumerations.

The 221 stage households reported 422 parents, siblings and children that were living at 347 other Oakland addresses when the census was taken six months earlier. Considerable difficulty was encountered in locating some of these relatives in the second stage enumerations because stage 1 respondents often provided incomplete addresses. In fact, only 191 or slightly less than 50 percent of the relatives' addresses were reported in complete detail by stage 1 respondents. More than 90 percent of the 191 addresses were located in stage 2.

Of the 156 addresses of relatives that were incompletely reported by stage 1 respondents, more than half of them were identified by the telephone directory and by the census address register. Figure III presents a flow chart which shows where the relatives' addresses were obtained from and how many of them were ultimately located in stage 2 enumerations. In summary, of the 347 households of relatives reported by stage 1 respondents, detailed addresses were obtained for about 80 percent and more than 70 percent were located in stage 2.

One of the principal objectives of the second stage enumerations was to validate the census addresses of relatives reported by first stage respondents. Otherwise, it would be impossible to decide whether the relatives not matched in the census enumeration listing were missed by the census or whether they were enumerated at other addresses. Consequently, if the relatives' addresses could not be ascertained from the first stage respondents or from ancillary sources such as telephone directories or the census address register is casts doubt on the feasibility of PEMS.

Two factors were investigated in the Oakland PEMS with respect to their possible effect in the success in locating relatives in the second stage enumerations. Table 1 compares the success in locating relatives in the second stage according to whether the person reported for himself or someone else reported for him in the first stage enumeration. The difference between the success rates is small and it is not statistically significant.

Table I: Success in Locating Relatives at Second Stage Addresses by Type of Respondent at 1st Stage Address

Type of Reporter \ Success in Locating	Sample Size	Total %	Relatives Located	Relatives Not Located
TOTAL	422	100	71%	29%
Self	347	100	70%	30%
Proxy & Other*	75	100	75%	25%

* This includes reports from unidentified reporter.

Table II compares the rate of success in locating relatives on the basis of the type of relative reported in the first stage enumerations. Since the sample sizes are relatively small, the findings need to be interpreted with caution. However, it appears as if the success rates are comparable for fathers, siblings and children, and by comparison mothers have higher success rates and relatives that have multiple links with several people in the first stage households have lower success rates.

Table II: Success in Locating Relatives at Second Stage Addresses by Type of Relative Reported at 1st Stage Address

Type of Relative Reported \ Success in Locating	Sample Size	Total %	Relatives Located	Relatives Not Located
TOTAL	422	100	71%	29%
Mother	60	100	88%	12%
Father	49	100	67%	33%
Sibling	216	100	70%	30%
Children	67	100	73%	27%
Combination*	30	100	37%	63%

* This category indicates those relatives reported that were more than one type of relative to the members of the sample household.

From these findings, we have tentatively concluded that adopting a self respondent rule in PEMS would not improve the success rate in locating relatives in the second stage enumeration, but that limiting the consanguine network of the counting rule to particular relatives such as mothers could enhance the success rate. It is also noteworthy, that many of the relatives who lived at addresses that were not located in the second stage of the Oakland PEMS could be successfully matched with their census enumerations.

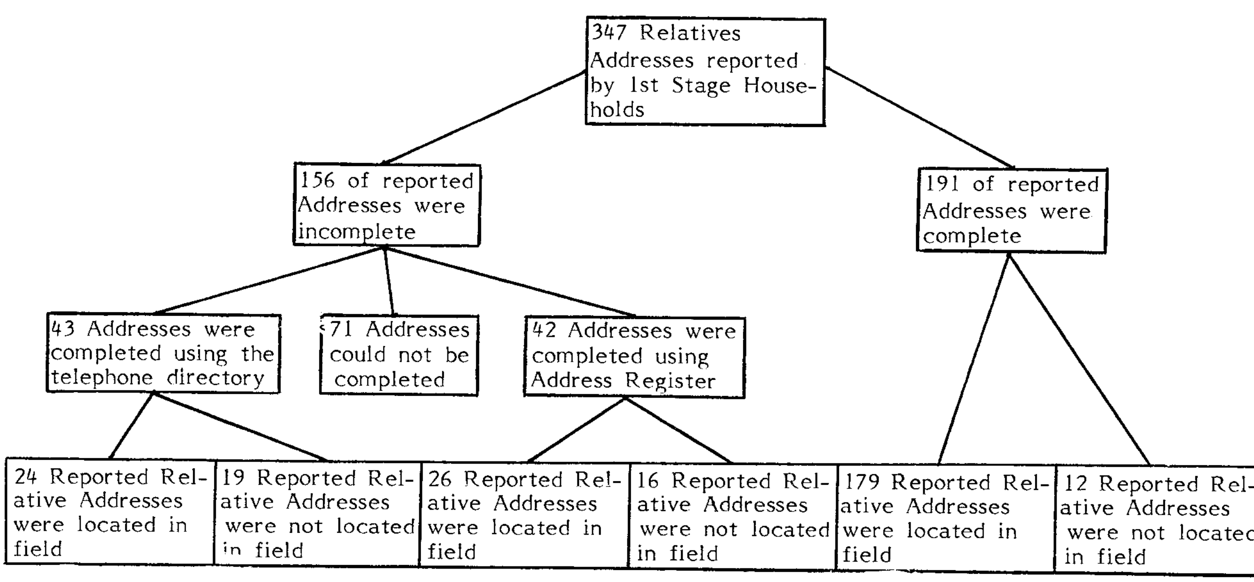
5. CONCLUDING REMARKS

The Oakland PEMS was conducted primarily to test the feasibility of conducting post-enumeration surveys using network counting rules. In this paper, we reported on a PEMS finding relating to the success of locating relatives reported in the first stage enumerations. Although the findings were somewhat disappointing, we believe not substantial improvements will be possible based on the lessons learned in Oakland. Many other findings from the Oakland PEMS reamin to be reported including the completeness of reporting of relatives, and the success in matching relatives with the census enumerations, etc.

In this paper we have discussed the possible attractiveness of network estimators for reducing correlation bias in the dual system estimators of census undercoverage. But other error effects of counting rules need to be investigated including sampling errors, response variance, response bias, matching bias, etc. The ultimate objective of PEMS research is to compare the cost and error effects of counting rules and counting rule weights on dual system estimates of census undercoverage. This information is essential for designing efficient post-enumeration surveys to evaluate census coverage.

The data collected in the Oakland PEMS are being analyzed for the error effects of counting rules and counting rule weights in dual system estimators of census coverage. Additional information will be forthcoming from the further tests of PEMS that are being conducted by the Census Bureau in conjunction with the dress rehearsals for the 1980 censis in Richmond, Virginia and in the southwestern counties in Colorado.

Figure III: Chart of the Success of Locating Reported Relative Addresses by the Source Used to Complete the Address



ACKNOWLEDGEMENTS

Many people made valuable contributions to this project. In particular, we would like to thank Patricia Royston and Kathy Ockay for their technical support and Charles Jones and David Bateman who supervised the Oakland PEMS.

REFERENCES

[1] Birnbaum, Z.W. and Sirken, Monroe G. "Design of Sample Surveys to Estimate the Prevalence of Rare Diseases; Three Unbiased Estimates," NCHS Series 2, No. 11, October 1965.

[2] Nathan, Gad; Schmelz, U.O.; and Kenvin, J. "Multiplicity Study of Marriages and Births in Israel," NCHS Series 2, No. 70, June 1977.

[3] Sirken, Monroe G. "Household Surveys with Multiplicity," Journal of the American Statistical Association, March 1970, Vol. 65, pp 257-266.

[4] ________, "Design of Household Sample Surveys to Test Death Registration Completeness," Demography, Vol. 10, No. 3, August 1970 pp 469-478.

[5] ________, "Dual System Estimators Based on Multiplicity Surveys," Developments in Dual System Estimation of Population Size and Growth, The University of Alberta Press, 1978 pp 81-88.

[6] Sirken, Monroe G. and Royston, Patricia N. "Underreporting of Births and Deaths in Household Surveys of Population Change," American Statistical Association, Proceedings of the Social Statistics Section, pp 412-415.

[7] Sirken, Monroe G.; Inderfurth, G.P.; Burnham, C.E.; and Danchik, C.M. "Household Sample Surveys of Diabetes: Design Effects of Counting Rules," American Statistical Association, Proceedings of the Social Statistics Section, 1975, pp 659-663.

[8] Schmelz, U.O.; Nathan, Gad; and Kenvin, J. "Multiplicity Study of Births and Deaths in Judea-Samaria and Gaza Strip-North Sinai," Technical Publications, Series No. 44, Central Bureau of Statistics, Jerusalem, 1977

APPENDIX

PES and PEMS Estimators

1. Introduction

Denote the population by $I = \left\{I_1, \ldots, I_\alpha, \ldots, I_N\right\}$ and the frame of housing units at which the N persons reside by $H = \left\{H_1, \ldots, H_i, \ldots, H_L\right\}$. In the survey that is conducted to estimate N, people in I are eligible to be enumerated at the households in H that are specified by the survey counting rule. Let the indicator variable ${}_r\delta_{\alpha i}$ specify the links between the people and households eligible to report them in compliance with counting rule r. Then

$$ {}_r\delta_{\alpha i} = \begin{cases} 1 \text{ if } I_\alpha (\alpha=1,\ldots,N) \text{ is eligible to be enumerated at } H_i (i=1,\ldots,L) \\ 0 \text{ otherwise.} \end{cases} $$

The multiplicity estimator of N based on counting rule r is

$$ N_r = \frac{L}{\ell} \sum_{i=1}^{L} \sum_{\alpha=1}^{N} t_i \, {}_rW_{\alpha i} \, {}_r\delta_{\alpha i}, \qquad (1.0) $$

where $\ell \leq L$ is the household sample size, the ${}_rW_{\alpha i}$'s are the counting rule weights, and the Bernoulli variable

$$ t_i = \begin{cases} 1 \text{ if } H_i (i=1,\ldots,L) \text{ is selected in the sample} \\ 0 \text{ otherwise.} \end{cases} $$

The estimator N_r is unbiased if and only if

$$ \sum_{i=1}^{L} {}_r\delta_{\alpha i} \, {}_rW_{\alpha i} = 1, \quad (\alpha=1,\ldots,N). \qquad (1.1) $$

Alternative counting rule algorithms exist that satisfy the unbiasedness conditions. For instance, the conditions are satisfied if

$$ {}_rW_{\alpha i} = \frac{{}_rR_{\alpha i}}{{}_rR_\alpha} \; (\alpha=1,\ldots,N), \; (i=1,\ldots,L) \qquad (1.2) $$

where ${}_rR_{\alpha i}$ = the number of persons residing in $H_i (i=1,\ldots,L)$ that are linked to $I_\alpha (\alpha=1,\ldots N)$ by rule r, and ${}_rR_\alpha = \sum_{i=1}^{L} {}_rR_{\alpha i}$ = the number of persons in population I that are linked to $I_\alpha (\alpha=1,\ldots,N)$ by rule r. If the weights are assigned such that ${}_rW_{\alpha i} = {}_rW_\alpha$, the unbiasedness conditions are satisfied by

$$ {}_rW_\alpha = \frac{1}{{}_rS_\alpha} \; (\alpha=1,\ldots,N), \qquad (1.3) $$

where

$$ {}_rS_\alpha = \sum_{i=1}^{L} {}_r\delta_{\alpha i} = \text{the number of households} $$

in H at which $I_\alpha (\alpha=1,\ldots,N)$ is eligible to be enumerated by rule r. The ${}_rW_\alpha$'s and ${}_rW_{\alpha i}$'s are referred to in the text as the household counting rule weights and the person counting rule weights respectively. If the weights are assigned such that ${}_r W_{\alpha i} = {}_r W$, the unbiasedness conditions are

$$ {}_r W = \frac{1}{{}_r \bar{s}} \qquad (1.4)$$

where

$$ {}_r \bar{s} = \frac{1}{N} \sum_{\alpha=1}^{N} {}_r s_\alpha = \text{average number of times}$$

that a person in I is eligible to be enumerated at households in H.

2. Survey Estimators

In the survey process, some people are erroneously enumerated at eligible households, and some people are not enumerated at eligible households. Since the first problem is relatively unimportant it will be ignored in the following analysis. Let the random variable

$$ {}_r \delta'_{\alpha i} = \begin{cases} 1 \text{ if } I_\alpha (\alpha=1,\ldots,N) \text{ is enumerated at } H_i \ (i=1,\ldots,L) \\ 0 \text{ otherwise.} \end{cases}$$

The biased count of N based on the survey using rule r is

$$E(N''_r) = \sum_{i=1}^{L} \sum_{\alpha=1}^{N} {}_r W_{\alpha i} \, E({}_r \delta''_{\alpha i}) \qquad (2.0)$$

where

$$ {}_r \delta''_{\alpha i} = {}_r \delta_{\alpha i} \cdot {}_r \delta'_{\alpha i}.$$

Since the census adopts a de jure residence rule which uniquely links every person to his de jure residence, ${}_c W_{\alpha i} = 1$. Thus, the expected census estimate of N is

$$E(N''_c) = \sum_{i=1}^{L} \sum_{\alpha=1}^{N} E({}_c \delta''_{\alpha i}). \qquad (2.1)$$

After the census is conducted, the population in in ℓ sample housing units is re-enumerated using a rule r. The survey estimate of N is

$$\hat{N}''_r = \frac{L}{\ell} \sum_{i=1}^{L} \sum_{\alpha=1}^{N} t_i \ {}_r W_{\alpha i} \ {}_r \delta''_{\alpha i}. \qquad (2.2)$$

Three potential counting rules (r=1,2,3) for the post-enumeration survey are:

r	Counting Rule
1	I_α (α=1,...,N) is linked to his de jure residence.
2	I_α (α=1,...,N) is linked to the residences of specified relatives who do not live with I_α.
3	I_α(α=1,...,N) is linked to his de jure residence, and to other residences containing his specified relatives.

Thus

$$\hat{N}''_1 = \frac{L}{\ell} \sum_{i=1}^{L} t_i \sum_{\alpha=1}^{N} {}_1 \delta''_{\alpha i}. \qquad (2.3)$$

Also

$$\hat{N}''_3 = \frac{L}{\ell} \sum_{i=1}^{L} t_i \sum_{\alpha=1}^{N} \left({}_1 \delta'_{\alpha i} + {}_2 \delta''_{\alpha i} \right) {}_3 W_{\alpha i}. \qquad (2.4)$$

since r_1 and r_2 are complimentary rules, that is

$$ {}_3 \delta''_{\alpha i} = {}_1 \delta''_{\alpha i} + {}_2 \delta''_{\alpha i}. \qquad (2.5)$$

$\hat{N}''_3$ and $\hat{N}''_1$ respectively represent the PEMS and PES estimators of N. However, $\hat{N}''_2$ would not be seriously considered as an estimator of N unless there was assurance that every I_α(α=1,...,N) was eligible to be enumerated in PEMS because he had a type of relative specified by r_2.

3. Dual System Estimators

The dual system estimator of N based on the census and the post-enumeration survey using r is

$$\hat{N}^*_{c/r} = \frac{N''_c \hat{N}''_r}{\hat{N}''_{c/r}} \qquad (3.0)$$

where

$$\hat{N}''_{c/r} = \frac{L}{\ell} \sum_{\alpha=1}^{N} \sum_{i=1}^{L} t_i \ {}_r W_{\alpha i} \ {}_r \delta''_{\alpha i} \sum_{j=1}^{L} {}_c \delta''_{\alpha j} \qquad (3.1)$$

is the estimator of the number of persons in I that are enumerated in both the census and the survey. Determining $\hat{N}''_{c/r}$ involves matching the persons that were enumerated in the survey with those that were enumerated in the census.

The correlation bias of $\hat{N}^*_{c/r}$ is

$$B\left(\hat{N}^*_{c/r}\right) = 1 - \frac{E(N''_r) \ E(N''_c)}{N \cdot E(N''_{c/r})} \qquad (3.2)$$

If $|B(\hat{N}^*_{c/1})| \geq |B(\hat{N}^*_{c/2})|$ it follows from (2.4) and (3.2) that the correlation bias of the dual system estimator would be smaller for the PEMS based on r_3 than the PES based on r_1. The correlation bias is equal to zero if and only if there is independence between being enumerated in the census and the post-enumeration survey. In other words, $B(\hat{N}^*_{c/r}) = 0$, if and only if,

$$N \cdot E\ (N''_{c/r}) = E(N''_r) \ E(N''_c).$$