Institute of Economic Development and Research
SCHOOL OF ECONOMICS
University of the Philippines System

Discussion Paper NO. 75-15 September 1975

THE EFFECTS OF AGE MISSTATEMENT ON ESTIMATES OF NET MIGRATION

by

Bryan L. Boulier

THE EFFECTS OF AGE MISSTATEMENT ON ESTIMATES OF NET MIGRATION

by

Bryan L. Boulier

There are many procedures for estimating net migration. Among the most widely used are variants of the census survival method. The procedure for estimating net migration to an area using the forward census survival method is to project the initial population of the area to the date of the second census and to subtract the projected population from the enumerated population, the difference being estimated net migration. The survival rates for the projection are calculated from the enumerated national populations of the appropriate age groups at the two census dates (e.g., the ratio of the number age 15-19 in 1970 to the number age 5-9 in 1960.) It has been demonstrated that census survival methods implicitly take into account the relative coverage of the two censuses as well as age misstatement, so that it is very often the case that these methods are preferred to other residual estimation techniques such as vital statistics or life table survival procedures (Hamilton and Anderson, 1944; Zachariah, 1962; and Hamilton, 1966.) There has been much analysis of the biases in the estimates of the volume of net migration rates when the two censuses have differing degrees of enumeration error (Price, 1955; Zachariah, 1962; Hamilton, 1966; and Stone, 1967.) It has been assumed that the consequences of age misstatement are identical to the consequences of enumeration error since the transfer of persons out of an age-group by age misstatement is analogous to an undercount of the population in the age group by a similar percent. This note presents for the first time an explicit derivation of the effects of age misstatement on estimates of the volume of net migration and net migration rates and demonstrates that results which have been obtained in the analysis of enumeration error are not always applicable to age misstatement. The estimates from the census survival rate (CSR) method are compared with those obtained from the use of life table survival rates (LTSR).

Assumptions and Definitions.

We shall analyze a simple case of age misstatement where a certain fraction α of an age group is reported as belonging to an older age group. Such age transfers may arise, for instance, from deliberate misstatement of age, rounding of ages, or inaccuracy of estimation by enumerators who have to guess the ages of respondents not knowing their ages. It will be assumed that there are no [illegible] of an [illegible]

[illegible]

assumed that the consequence [illegible]

identified [illegible] consequence [illegible]

the [illegible]

actions [illegible] the population

[illegible]

[illegible]

[illegible]

not affect [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

differences in mortality among regions in a country, that there is a complete count of the population in both censuses, and that all migration occurs at the end of the intercensal period. Zachariah (1962) examines the consequences of differential mortality, previously cited authors have considered cases of incomplete enumeration, and the last assumption ensures that in the absence of age misstatement the forward estimate of migration correctly estimates the volume of net migration. Siegel and Hamilton (1952) analyze other patterns of the timing of migration, and the results obtained here are easily generalized to those patterns.

We shall consider four age groups in a population when the length of the age interval equals the length of the intercensal period (e.g., five year age groups and five year census intervals.) Let us define the following terms:

$P(x,t)$ = true national population in age group x at time t (x = 1, 2, 3, 4; t = 1, 2),

$p(x,t)$ = true regional population age group x at time t,

$P^*(x,t)$ = reported national population age group x time t,

$p^*(x,t)$ = reported regional population age group x time t,

s_x = true survival rate from age group x to x + 1,

$s_x{}^*$ = census survival rate from age group x to x + 1 calculated from the reported national population,

M_x = true net migration age x,

$M_x{}^*$ = estimated net migration age x from CSR method, and

$\hat{M}_x$ = estimated net migration age x from LTSR method.

It is assumed that migrants move in age group x but are age x + 1 at the date of the second census. Finally, it is assumed that a fraction α of persons age x = 2 are erroneously reported as being age x = 3.

Census Survival Rate Estimates of Net Migration.

Table 1 shows the true and reported national and regional populations at both census dates. The CSR estimates of net migration are:

$$(1) \quad M_x{}^* = p^*(x + 1,\ t + 1) - s_x{}^*(x,t)\ p^*(x,t).$$

The National CSR are:

$$(2) \quad s_1^* = \frac{P^*(2,2)}{P^*(1,1)} = s_1\ (1 - \alpha),$$

$$(3) \quad s_2^* = \frac{P^*(3,2)}{P^*(2,1)} = \frac{s_2}{1 - \alpha} + \frac{\alpha\, s_1}{1 - \alpha}\ \left[\frac{P(1,1)}{P(2,1)}\right], \text{ and}$$

$$(4) \quad s_3^* = \frac{p^*(4,2)}{p^*(3,1)} = \frac{s_2}{1 + a \quad \frac{p(3,1)}{p(2,1)}}$$

The CSR are rather complex, s_1^* is the product of the true survival rate s_1 and one minus the fraction transferred to the next older age group. s_2^* is a combination of the true survival rate, the age misstatement factor, and the national age distribution. s_3^* is also a function of the national age distribution.

Table 2 shows the estimated volume of net migration to the region calculated from equation (1) after some simplification of the resulting expressions. There are several interesting results which may be derived from the table. First, consider M_1^*. The estimated volume of migration is a fraction of true migration depending upon the proportion of persons whose ages are misstated. The smaller is , the closer is estimated to true migration. A migration rate which relates migration to the end of the period population
relates migration to the end of the period population $\left(\frac{(1 - \quad)M_1}{(1 - \quad)p(2,2)}\right)$ is unbiased; a migration rate which relates estimated migration to the beginning population $\left(\frac{(1 - \quad)M_1}{p(1,1)}\right)$ and which would be used for population projection is biased towards zero. (See Hamilton (1965) for a discussion of alternative measures of migration rates.)

The results for M_1^* are similar to those obtained in the analysis of enumeration error. If the population is underenumerated by a fraction α for both censuses, the volume of net migration is underestimated by the same proportion but migration rates c culated from the end of period or beginning of period populations will be unbiased. For age misstatement, as was noted, the latter rate is biased.

Now we turn to migration at other ages. M_2^* equals the true volume of net migration plus terms which include the survival rate s_1, the regional and national age distributions, migration into the preceding age bracket, and the extent of age misstatement. If the proportionate age distributions of the regional and national population in the relevant age groups are identical at the time of the first census, the term in square brackets equals zero, and $M_2^* = M_2 + \alpha\, M_1$, so that M_2^* is very close to M_2 if α is small, M_1 is small, or both. If however, the age distribution are identical the bracketed term is non-zero and is weighted by $\alpha\, s_1 p(2,1)$. Even if α is small, the value of the term could be large since $p(2,1)$ is ordinarily large. There is no presumption as to the direction of bias in the estimator M_2^*, and rates calculated from M_2^* will generally be biased. The relationship between M_3^* and M_3 depends upon the age distributions of the regional and national populations at the time of the first census, the survival rate s_3, and α. If the proportionate age distributions of the national and regional populations in the relevant age intervals are equal at the time of the first census, the term in the square brackets equals zero, and $M_3^* = M_3$. If the term is non-zero, then the smaller is α the closer is M_3^* to M_3. But even if α is small, the error may be large since α times the bracketed term is multiplied by $s_3p(3,1)/(1 + \frac{p(2,1)}{p(3,1)})$. As in the case of M_2^*, there is no presumption as to the direction of the bias in the estimator M_3^* and the rates calculated from it will generally be biased.

Life Table Survival Rate Estimates of Net Migration.

Despite the biases note above, it should not be forgotten, however, that the CSR method is likely to provide estimates superior to those calculated from LTSR methods. Suppose we assume that the national population has the same age distribution as the stationary population associated with the life table so that the s_x may be identified with life table survival rates. The forward estimates of migration using LTSR and the enumerated regional population are also shown in Table 2. The estimates include terms of the form $\alpha\ s_x\ p(x,1)$ which may well be large even if α is small. For instance, consider $\hat{M}_1 = (1 - \alpha)\ M_1 - \alpha\ s_1 p(1,1)$. Persons who are transferred to an older age group are counted as outmigrants. The CSR method implicitly adjusts for this age transfer. $\hat{M}_2$ counts as in-migrants persons who transfer from age x = 2 at the time of the second census ($\alpha\ s_1 p(1,1) + \alpha\ M_1$) and the survivors of the persons who misstated their age at the time of the first census ($\alpha\ s_2 p(2,1)$). $\hat{M}_3$ counts as out-migrants the persons who were erroneously enumerated as part of age group x = 3 at the time of the first census ($\alpha\ p(2,1)$) times s_3, the survival rate appropriate for survival from x = 3 to x = 4.

Example.

Table 3 presents estimates of net migration using hypothetical data. The national population is assumed to be stationary with $s_1 = .99$, $s_2 = .98$, and $s_3 = .97$, and it is assumed that five percent of the age group x = 2 is erroneously reported as belonging to age group x = 3. The CSR's calculated from the enumerated national population are $s_1^* = .9405$, $s_2^* = 1.0842$, and $s_3^* = .9229$. The regional population is assumed to have the same true survival rates and the same pattern of age misreporting as the national population. At time t = 1, the regional population is constructed by taking 10 percent of the national population age x = 1, 12 percent of the number x = 2, 11 percent of the number x = 3, and 10 percent of the number x = 4. Part A of the table gives the basic data, part B compares true net migration with migration estimated by the CSR method and by the LTSR method, and part C compares migration rates. It is readily seen from parts B and C of the table that the CSR estimates are superior to the LTSR estimates. $\hat{M}_1$ is less than half of M_1, $\hat{M}_2$ is more than three times M_2, and $\hat{M}_3$, is negative while M_3 is large and positive. In contrast the CSR estimates are all within ten percent of the correct values. Nonetheless, it is important to note that in the presence of modest age misstatement and age distributions of the regional and national populations which are not very dissimilar, the errors in the CSR estimates of net migration and net migration rates are large enough to be important for some analyses.

Summary.

In this paper, we have shown the consequences of age misstatement for census survival rate and life table survival rate estimates of net migration. In the presence of age misstatement, CSR estimates of migration and migration rates are generally biased although the direction of bias is not always clear. Moreover, estimation errors due to age misstatement are not necessarily similar to those which result from underenumeration of a population. Although CSR estimates of net migration are usually superior to those obtained by LTSR methods, the above analysis suggests that in some cases it is probably worthwhile to attempt to adjust the data for age misreporting and then to apply the LTSR method for additional estimates to be compared with those obtained by the CSR procedure.

R E F E R E N C E S

1. Hamilton, C. Horace. 1965. "Practical and Mathematical Considerations in the Formulation and Selection of Migration Rates," Demography, II, 429-463.

2. Hamilton, C. Horace. 1966. "Effect of Census Errors on the Measurement of Net Migration," Demography, III, 393-415.

3. Hamilton, C. Horace and F.M. Anderson, 1944. "Use of Survival Rate Method in Measuring Net Migration," Journal of the American Statistical Association, XXXIX, 197-206.

4. Price, Daniel O. 1955. "Examination of Two Sources of Error in the Estimation of Net Internal Migration," Journal of the American Statistical Association, L, 689-700.

5. Siegel, Jacobs and C. Horace Hamilton. 1952. "Some Considerations in the Use of the Residual Method of Estimating Net Migration," Journal of the American Statistical Association, XLVII, 475-500.

6. Stone, Leroy O. 1967. "Evaluating the Relative Accuracy and Significance of Net Migration Estimates," Demography, IV, 310-330.

7. Zachariah, K.C. 1962. "A Note on the Census Survival Ratio Method of Estimating Net Migration," Journal of the American Statistical Association, LVII, 175-183.

TABLE 1. True and Reported: National and Regional Populations

National Population

True				
t = 1	$P(1,1)$	$P(2,1)$	$P(3,1)$	$P(4,1)$
t = 2	$P(1,2)$	$P(2,2)=s_1P(1,1)$	$P(3,1)=s_2P(2,1)$	$P(4,2)=s_3P(3,1)$
Reported				
t = 1	$P^*(1,1)=P(1,1)$	$P^*(2,1)=(1-\alpha)P(2,1)$	$P^*(3,1)=P(3,1)+\alpha P(2,1)$	$P^*(4,1)=P(4,1)$
t = 2	$P^*(1,2)=P(1,2)$	$P^*(2,1)=(1-\alpha)P(2,2)$ $=(1-\alpha)s_1P(1,1)$	$P^*(3,2)=P(3,2)+\alpha P(2,2)$ $=s_2P(2,1)+\alpha s_1P(1,1)$	$P^*(4,2)=s_3P(3,1)$

Regional Population

True				
t = 1	$p(1,1)$	$p(2,1)$	$p(3,1)$	$p(4,1)$
t = 2	$p(1,2)$	$p(2,2)=s_1p(1,1)++M_1$	$p(3,2)=s_2p(2,1)+M_2$	$p(4,2)=s_3p(3,1)+M_3$
Reported				
t = 1	$p^*(1,1)=p(1,1)$	$p^*(2,1)=(1-\alpha)p(2,1)$	$p^*(3,1)+p(2,1)$	
t = 2	$p^*(1,2)=p(1,2)$	$p^*(2,2)=(1-\alpha)p(2,2)$ $=(1-\alpha)$ [illegible]		

TABLE 1. True and Reported: National and Regional Populations

		National Population		
True				
$t = 1$	$P(1,1)$	$P(2,1)$	$P(3,1)$	$P(4,1)$
$t = 2$	$P(1,2)$	$P(2,2)=s_1P(1,1)$	$P(3,1)=s_2P(2,1)$	$P(4,2)=s_3P(3,1)$
Reported				
$t = 1$	$P^*(1,1)=P(1,1)$	$P^*(2,1)=(1-\alpha)P(2,1)$	$P^*(3,1)=P(3,1)+\alpha P(2,1)$	$P^*(4,1)=P(4,1)$
$t = 2$	$P^*(1,2)=P(1,2)$	$P^*(2,1)=(1-\alpha)P(2,2)$ $=(1-\alpha)s_1P(1,1)$	$P^*(3,2)=P(3,2)+\alpha P(2,2)$ $=s_2P(2,1)+\alpha s_1P(1,1)$	$P^*(4,2)=s_3P(3,1)$
		Regional Population		
True				
$t = 1$	$p(1,1)$	$p(2,1)$	$p(3,1)$	$p(4,1)$
$t = 2$	$p(1,2)$	$p(2,2)=s_1p(1,1)++M_1$	$p(3,2)=s_2p(2,1)+M_2$	$p(4,2)=s_3p(3,1)+M_3$
Reported				
$t = 1$	$p^*(1,1)=p(1,1)$	$p^*(2,1)=(1-\alpha)p(2,1)$	$p^*(3,1)+p(2,1)$	$p^*(4,1)=p(4,1)$
$t = 2$	$p^*(1,2)=p(1,2)$	$p^*(2,2)=(1-\alpha)p(2,2)$ $=(1-\alpha)s_1p(1,1)+M_1$	$p^*(3,2)=p(3,2)+p(2,2)$ $=s_2p(2,1)+\alpha\left[s_1p(1,1)+M_1\right]+M_2$	$p^*(4,2)=s_3p(3,1)+M_3$

TABLE 2. Estimated Net Migration

A. Census Survival Method

$$M_1^* = (1 - \alpha)$$

$$M_2^* = M_2 + \alpha M_1 + \alpha s_1\ p(2,1) \cdot \left[\frac{p(1,1)}{p(2,1)} - \frac{P(1,1)}{P(2,1)}\right]$$

$$M_3^* = M_3 + \frac{\alpha s_3\ p(3,1)}{1 + \alpha\ \frac{P(2,1)}{P(3,1)}} \cdot \left[\frac{P(2.1)}{P(3,1)} - \frac{p(2,1)}{p(3,1)}\right]$$

B. Life Table Survival Method

$$M_1 = (1 - \alpha) M_1 - \alpha s_1\ p\ (1,1)$$

$$M_2 = M_2 + \alpha M_1 + \alpha s_1\ p(1,1) + \alpha s_2\ p(2,1)$$

$$M_3 = M_3 - \alpha\ s_3\ p(2,1)$$

TABLE 3. An Illustrative Example of Age Misstatement and Estimation of Net Migration by Census Survival Rate and Life Table Survival Rate Methods

A. Data

Age/Time	National Population True (t=1,2)	National Population Reported (t=1,2)	Regional Population True t=1	Regional Population True t=2	Regional Population Reported t=1	Regional Population Reported t=2
x = 1	100,000	100,000	10,000	10,000	10,000	10,00
x = 2	99,000	94,050	11,880	10,900	11,286	10,35
x = 3	97,020	101,970	10,672	12.142	11,266	12,68
x = 4	94,109	94,109	9,411	10,852	9,411	10,85

B. Number of Net Migrants

	x=1	x=2	x=3
True (Mx)	1,000	500	500
Census Survival Rate Method (Mx*)	950	451	455
Life Table Survival Rate Method (Mx)	455	1,627	-76

C. Net Migration Rates (1)

	True M_1	True M_2	True M_3	Census Survival M_1^*	Census Survival M_2^*	Census Survival M_3^*	Life Table M_1	Life Table M_2	Life Table M_3
Beginning Period	.100	.042	.047	.095	.040	.040	.046	.144	-.007
End of Period	.092	.041	.046	.096	.036	.042	.044	.128	-.007

(1) Beginning of period rates are calculated from $M_x/p(x,1)$ and end of the period rates are calculated from $M_x/p(x+1,2)$. The true rates use true migration and the true regional population; the CSR and LTSR estimates of migration rates use the reported regional population as denominators.