ESTIMATING THE COMPLETENESS OF REGISTRATION OF DEATHS AND THE RELATIVE UNDERENUMERATION IN TWO SUCCESSIVE CENSUSES

by James Trussell and Jane Menken

William Brass (1979) has suggested a method of simultaneously estimating the completeness of registration of deaths during an intercensal period and the completeness of enumeration of one of the censuses relative to the completeness of the other (see article beginning on page 6 of this issue). The method is based upon examination of the relationship between the intercensal death rates above age *a* estimated in two ways: from intercensal survival of cohorts enumerated in both censuses and from the usual estimate of death rates as registered deaths per estimated population.

In this note we set out in detail the assumptions underlying Brass's new method and seek to clarify the interpretation of the results. We also evaluate the performance of the technique under ideal conditions in which the assumptions are fully satisfied.

Suppose that there are two censuses of a population taken n years apart; this population is assumed to be closed to in- and out-migration. Let

P_a = population over exact age a at the time of the first census;

P'_a = population over exact age a at the time of the second census;

$d_a^\dagger$ = death rate of the population above age a based on census survival estimates; and

d_a^* = death rate for persons above age a based on a true enumeration of deaths between the two censuses.

A hat (^) denotes the *observed* as opposed to *true* rate. Suppose further that some proportion p of the true deaths is actually recorded, and that the completeness of recording does not vary by age or time. Further, let us assume that in one census the population is fully enumerated whereas in the other only some fraction, $1-\delta$, of each age group is counted. Then Brass has proposed that the death rates of the population above age a calculated from census survival and calculated in the normal manner from registered deaths have a simple linear relationship:

$$\hat{d}_a^\dagger = \frac{1}{p}\hat{d}_a^* + \delta.$$

Brass has suggested that the death rate above age a based on census survival be estimated as

$$\hat{d}_a^\dagger = \frac{\hat{P}_{a-n/2} - \hat{P}'_{a+n/2}}{\tfrac{1}{2}(\hat{P}_a + \hat{P}'_a)} \quad . \tag{1}$$

The death rate above age a based on registered deaths is calculated using the same denominator. The relationships between $\hat{d}^\dagger$ and $\hat{d}^*$ vary according to whether underenumeration occurs in the first or second census (or both) and according to the precise definition of the death rates.

Underenumeration in one census only

Central death rates. If in the first census the population is correctly enumerated and in the second it is underenumerated by the same proportion at every age, then

$$\hat{P}_a = P_a$$
$$\hat{P}'_a = (1-\delta)P'_a. \tag{2}$$

From these definitions, the true and observed death rates can be expressed as follows:

$$\text{TRUE:} \qquad d_a^\dagger = \frac{P_{a-n/2} - P'_{a+n/2}}{\tfrac{1}{2}(P_a + P'_a)} \tag{3}$$

$$\text{OBSERVED:} \qquad \hat{d}_a^\dagger = \frac{P_{a-n/2} - (1-\delta)P'_{a+n/2}}{\tfrac{1}{2}(P_a + (1-\delta)P'_a)} \quad . \tag{4}$$

Hence $\hat{d}_a^\dagger = d_a^\dagger R + \delta Z_a$ (5)

$$\text{where} \quad R = \frac{P_a + P'_a}{P_a + (1-\delta)P'_a} \tag{6}$$

$$Z_a = \frac{P'_{a+n/2}}{\tfrac{1}{2}(P_a + (1-\delta)P'_a)} \quad . \tag{7}$$

In order to examine Equation (5) for a simple case, let us assume the population is stationary. Hence $P_a = P'_a = T_a$ if ℓ_0 (the radix of the life table) equals the annual number of births.

Then Equations (6) and (7) simplify to

$$R = \frac{T_a + T_a}{T_a + T_a(1-\delta)} = \frac{2}{2-\delta} \tag{8}$$

$$Z_a = \frac{T_{a+n/2}}{T_a - \frac{\delta}{2}T_a} = R\,\frac{T_{a+n/2}}{T_a} \quad . \tag{9}$$

Hence $\hat{d}_a^\dagger = R\{d_a^\dagger + \delta S_a\}$ (10)

$$\text{where} \quad S_a = \frac{T_{a+n/2}}{T_a} \quad . \tag{11}$$

Brass has suggested that the death rate based on recorded data be formed by adding all deaths above age a over the n-year period and dividing by the same denominator used in constructing $d_a^\dagger$. Using the same stationarity assumption, the number of deaths above age a during the n-year period is just $n\ell_a$. Thus the two "true" death rates would be equal if $n\ell_a = {}_nL_{a-n/2}$. The suggestion that the deaths be estimated in this way was done for the sake of convenience only. If special tabulations of registered death data were available, then one could estimate properly the numerator of the death rate by counting the deaths during the n-year period to persons who would have been over age $a-n/2$ at the time of the first census. Let us assume that the deaths have been counted in this manner, so that $d_a^* = d_a^\dagger$. Also, by assumption, some proportion p of deaths is not recorded at every age during the n-year interval. Hence, since the number of deaths is $P_{a-n/2} - P'_{a+n/2}$, $\hat{d}_a^* = p d_a^* R$. Since, by assumption, $d_a^* = d_a^\dagger$, then from Equation (10)

$$\hat{d}_a^\dagger = \frac{1}{p}\hat{d}_a^* + R\delta S_a \tag{12}$$

$$= f\hat{d}_a^* + R\delta S_a \qquad \text{where } f = 1/p. \tag{13}$$

This contrasts with the Brass relation

$$\hat{d}_a^\dagger = f\hat{d}_a^* + \delta, \tag{14}$$

which is based on the proposition that

$$\hat{d}_a^\dagger = d_a^\dagger + \delta. \tag{15}$$

Strictly, this proposition implies that $RS_a = 1$, which clearly cannot be true for all δ or all a. Note that a symmetric relationship is obtained if the first census is assumed to be incomplete. In this case

$$\hat{d}_a^\dagger = d_a^\dagger - R'S_a'\delta \quad (16)$$

where $R' = \frac{2}{2-\delta}$ and $S_a' = \frac{T_{a-n/2}}{T_a}$.

Clearly, the desired linear relationship has not materialized.

Cohort death rates. If cohort death rates are calculated instead, then

$$d_a^\dagger = \frac{P_a - P'_{a+n}}{P_a} \quad (17)$$

With this definition of the death rates, different relationships arise depending on whether the first or second census has underenumerated the population. If the first census is underreported, then

$$\hat{d}_a^\dagger = \frac{\hat{P}_a - \hat{P}'_{a+n}}{\hat{P}_a} = \frac{(1-\delta)P_a - P'_{a+n}}{(1-\delta)P_a}$$

$$= \frac{d_a^\dagger}{1-\delta} + \frac{\delta}{1-\delta}. \quad (18)$$

Since d_a^* is assumed to be equal to $d_a^\dagger$, then

$$\hat{d}_a^* = \frac{d_a^\dagger}{f(1-\delta)}.$$

Therefore,

$$\hat{d}_a^\dagger = f\hat{d}_a^* - \frac{\delta}{1-\delta}. \quad (19)$$

If the second census is underreported, then

$$\hat{d}_a^\dagger = \frac{P_a - (1-\delta)P'_{a+n}}{P_a} = d_a^\dagger + \frac{\delta P'_{a+n}}{P_a} \quad (20)$$

or

$$\hat{d}_a^\dagger = f\hat{d}_a^* + \delta\frac{P'_{a+n}}{P_a} = f(1-\delta)\hat{d}_a^* + \delta. \quad (21)$$

Thus, we see that if the first census is underenumerated, then a desired linear relationship emerges; the constant term is, however, $\delta/1-\delta$ instead of the postulated δ. If the second census is underenumerated, then the relationship is still linear but the slope term is $f(1-\delta)$ and the constant term is δ.

If we define the death rate to be the deaths divided by the survivors, then we can obtain similar linear relationships. When the second census is underenumerated,

$$d_a^\dagger = \frac{P_a - P'_{a+n}}{P'_{a+n}} \quad (22)$$

$$\hat{d}_a^\dagger = \frac{\hat{P}_a - \hat{P}'_{a+n}}{\hat{P}'_{a+n}} = \frac{P_a - (1-\delta)P'_{a+n}}{(1-\delta)P'_{a+n}}$$

$$= \frac{d_a^\dagger}{1-\delta} + \frac{\delta}{1-\delta}. \quad (23)$$

Then

$$\hat{d}_a^\dagger = f\hat{d}_a^* + \frac{\delta}{1-\delta}. \quad (24)$$

Similar algebra will show that when the first census is underenumerated

$$\hat{d}_a^\dagger = f(1-\delta)\hat{d}_a^* - \delta. \quad (25)$$

Examination of these results reveals an interesting symmetry. Using the population of the first census as the denominator and assuming the first census is underenumerated produces an equation identical, except in the sign of the constant $(\delta/1-\delta)$, to using the population of the second census as the denominator and assuming that underenumeration occurs only in the second census. The other pair of assumptions (underenumeration in the census not used in the denominator) produces Equations (21) and (25) which have the same symmetry property: same slope and, except for sign, same intercept.

Underenumeration in both censuses

We have thus far assumed that either the first or the second census is underenumerated, but not both. Such an assumption is rarely likely to be fulfilled, and more commonly both censuses would underenumerate the population. With underenumeration in both censuses (denoted by δ and δ' respectively), the cohort census survival death rate defined by Equation (18) becomes

$$\hat{d}_a^\dagger = \frac{(1-\delta)P_a - (1-\delta')P'_{a+n}}{(1-\delta)P_a}$$

$$= \frac{(1-\delta)(P_a - P'_{a+n}) + (\delta'-\delta)P'_{a+n}}{(1-\delta)P_a}$$

$$= d_a^\dagger + \frac{P'_{a+n}}{P_a}\left(\frac{\delta'-\delta}{1-\delta}\right) \quad (26)$$

As before,

$$\hat{d}_a^* = \frac{d_a^*}{f(1-\delta)} = \frac{d_a^\dagger}{f(1-\delta)} \quad (27)$$

Hence, with a little algebraic manipulation, it can be shown that

$$\hat{d}_a^\dagger = f(1-\delta')\hat{d}_a^* + 1 - \frac{1-\delta'}{1-\delta}. \quad (28)$$

Examination of the relationship when Equation (22) is used to define the death rate yields an equation similar to (28). In this case

$$\hat{d}_a^\dagger = \frac{(1-\delta)P_a - (1-\delta')P'_{a+n}}{(1-\delta')P'_{a+n}}$$

$$= (1-\delta)f\hat{d}_a^* + \frac{1-\delta}{1-\delta'} - 1. \quad (29)$$

As before, there is a symmetry between the two equations: δ and δ' are reversed and the sign of the constant is reversed.

In Equation (29), the estimated constant term plus 1 can be interpreted as the completeness of enumeration of the first census relative to that of the second census, and the slope term can be interpreted as the completeness of enumeration of the first census relative to the completeness of registration of deaths. The inverse of $(1-\delta)/(1-\delta')$ and $(1-\delta)f$ can be interpreted respectively as the completeness of enumeration of the second census and the completeness of registration of deaths, both relative to the first census. A similar interpretation can, of course, be given to the slope and constant terms in Equation (28). Note that if we wish to interpret f and $1-\delta'$ in Equation (29) as relative to the completeness of enumeration in the first census, then both would be divided by $1-\delta$. The new relationship would become

$$\hat{d}_a^\dagger = f\hat{d}_a^* + \delta'/(1-\delta'), \tag{30}$$

which is identical in form to Equation (24). To obtain (24), we assumed that only the second census was underenumerated. Hence the results based on the assumption of underenumeration in only one census that were obtained earlier can be applied when there is underenumeration in both, provided the estimates are interpreted as being relative to the other census (assumed to be completely enumerated).

It should be noted that Equations (28) and (29) provide a reconciliation of the Brass technique with that proposed earlier by Preston (Preston and Hill, in press), which is based on the accounting identity

$$\frac{\hat{P}_a}{1-\delta} = \frac{\hat{P}'_{a+n}}{1-\delta'} + f\hat{D}_a^* \tag{31}$$

where $\hat{D}_a^*$ is the number of registered deaths during the intercensal period to persons aged a and above at the time of the first census. It can be seen that if both sides of (31) are multiplied by $1-\delta$ and then $\hat{P}'_{a+n}$ is subtracted from both sides and finally both sides are divided by $\hat{P}'_{a+n}$, the result is Equation (29). Similar algebra will produce Equation (28).

Magnitude of error

The algebra thus far has shown that the precise linear relationship $\hat{d}_a^\dagger = f\hat{d}_a^* + \delta$ does not emerge when the denominators are defined as the average at the two censuses. But we have not given any estimate of the magnitude of the error that would result if one did proceed to interpret directly the results of the exercise. Since we were not sure how Brass estimated the line for the Thailand example he provided, we reestimated it by ordinary least squares (using only points from age 12.5 to age 57.5, as he suggested) to get $\hat{d}_a^\dagger = .0594 + 1.3605\,\hat{d}_a^*$. The estimates are very similar to those obtained by Brass. Next we estimated the relationship using Equations (28) and (29). The denominator of the rate which Brass has labeled age 12.5 would be the $(P_0^{1960} + P_5^{1960})/2$ for Equation (28) and $(P_{10}^{1970} + P_{15}^{1970})/2$ for Equation (29); other rates are defined analogously. The results are as follows:

Based on (28): $\hat{d}_a^\dagger = .0583 + 1.3190\,\hat{d}_a^*$; (32)

Based on (29): $\hat{d}_a^\dagger = .0612 + 1.4017\,\hat{d}_a^*$. (33)

From (32) and (28) we see that the estimated completeness of the second census relative to the first is $-(.0583 - 1) = .9417$; hence the completeness of the first relative to the second is the reciprocal, 1.0619. From (33) and (29) we can calculate the estimated completeness of the first relative to the second as $1 + .0612 = 1.0612$. Hence the two estimates are very similar. From Equation (33) the completeness of death registration is .713 relative to the first census. From Equation (33) we calculate the same coverage rate as $(1/1.3190)(.9417) = .714$; again the results are gratifyingly similar.

Therefore, in this example, the results based on the approximate relationship proposed by Brass are close to those based on more exact equations. It is interesting to note that the Brass estimates are nearly identical to the average (slope = 1.3604, intercept = .0598) of the coefficients in Equations (32) and (33). This finding suggests that the Brass estimate of δ should be interpreted as $\left(\frac{1-\delta}{1-\delta'} - \frac{1-\delta'}{1-\delta}\right)/2$, which in words is half the difference between the completeness of the first relative to the second and the completeness of the second relative to the first. A similar interpretation can be placed on the estimate of the slope.

Summary

Brass has proposed a method for estimating simultaneously the completeness of registration of deaths in an intercensal period and the completeness of enumeration of the second census both relative to the completeness of enumeration of the first. Two death rates, $\hat{d}_a^\dagger$ based on census survival, and $\hat{d}_a^*$ based on registered deaths during the intercensal period to persons aged a and above at the time of the first census, are postulated to have a linear relationship. In this note, the estimators of $d_a^\dagger$ and d_a^* proposed by Brass have been shown to lead to a nonlinear relationship, even in the ideal circumstance where all the assumptions hold and underenumeration is known to occur in only one census. The source of the nonlinearity is shown to arise from the definition of the denominator used to calculate the two death rates. Other estimators which eliminate the nonlinearity were proposed. Finally, in the more realistic event that in both censuses the population is underenumerated, it is shown that the desired linear relationship between $d_a^\dagger$ and d_a^* will emerge if the definition of the denominator is altered. This formulation leads to a reconciliation of the approaches suggested by Brass and Preston. Without this adjustment, the parameters of the Brass procedure are difficult, if not impossible, to interpret directly in terms of relative underenumeration of the two censuses and relative underregistration of death. □

REFERENCES

Brass, William. 1979. A procedure for comparing mortality measures calculated from intercensal survival with the corresponding estimates from registered deaths. *Asian and Pacific Census Forum* 6(2):6–8.

Preston, Samuel, and Kenneth Hill. In press. Estimating the completeness of death registration. To appear in *Population Studies*.

Dr. James Trussell is Assistant Professor of Economics in the Woodrow Wilson School of Public and International Affairs and Faculty Associate at the Office of Population Research (OPR), Princeton University. Dr. Menken is Senior Research Demographer and Assistant Director of OPR. Dr. Trussell is currently working with the World Fertility Survey in London. Although his major research interest is mathematical demography, he has published on such diverse topics as the demography of Afghanistan, teenage pregnancy, and the demography of slaves. Dr. Menken, also a mathematical demographer, has concentrated her research on models of conception and birth.

Both authors participated in the Vital Registration Working Group at the East-West Population Institute in August 1978. Work on this paper began then and continued during and after their return to Princeton. "We found that concentration on complex mathematical derivations prevents the usual boredom of tedious plane rides," said Dr. Menken.