Evaluating the performance of death distribution methods for estimating death registration completeness: Applications to data from high income countries

Introduction

Conventional approaches to mortality measurement depend on complete vital registration data on deaths and accurate estimates of person years lived in a specified period of time. However, in countries where vital registration systems are underdeveloped, estimates derived from conventional methods can be badly biased. Various alternative approaches have been developed for mortality measurement in these contexts (UN 2003). Among these is the use of analytical methods that estimate levels of death registration completeness (Bennett and Horiuchi 1981; Hill 1987). These estimates can then be used to generate adjustment factors to correct the number of registered deaths to be consistent with population denominators. Death registration methods however rely on a variety of assumptions that can lead to biased estimates when violated.

While death distribution methods have increasingly been used in mortality estimation in recent years (Banister and Hill 2004; Merli 1998; Elo and Preston 1994), few empirical studies have evaluated how they perform. One evaluation strategy is to apply them to settings in which recording of deaths and population is thought to be essentially complete. Such an evaluation does not address issues of performance in the face of substantial age misreporting or the violation of other assumptions of the methodology, but is useful as a test of the methods under ideal circumstances.

Objectives

In this paper we evaluate how three death distribution methods perform with accurate data from the Human Mortality Database, and explore whether additional information can be used to improve performance. Specific objectives are as follows:

(a) Examine disparities in estimates of completeness from the Generalized Growth Balance method (Hill 1987), Synthetic Extinct Generations methods (Bennett and Horiuchi 1981, 1984), and the Adjusted Synthetic Extinct Generations methods (Hill and Choi 2004) in countries with good data.

(b) Evaluate the sensitivity of death registration methods to the age groups used to estimate adjustment factors.

(c) Examine the nature of the association between estimates of registration completeness and estimates of net international migration flows.

Data and Methods

This study uses data from the Human Mortality Database (HMD). This database provides population and mortality data for 28 countries of the now-developed world. For some countries, data recording extends as far back as the 1800's, but in this analysis we only use data for the period 1950 to 2000. Information provided by the HMD indicates that for some countries, the data quality for a number of years is very low. When this is the case, these years are excluded from the analysis. To estimate net international migration, we use net migration data provided by the United Nations Population Division (UN 2003).

Methods

1. The Generalized Growth Balance Method

The General Growth Balance method is a generalization of the Brass Growth Balance method derived from stable populations (Brass 1975). The Demographic Balancing Equation expresses the identity that the growth rate of the population is equal to the difference between the entry rate and the exit rate. This identity holds for open-ended age segments *x*+, and in a closed population the only entries are through birthdays at age *x*. The birth rate *x*+ minus the growth rate *x*+ thus provides a residual estimate of the death rate *x*+. If the residual estimate can be calculated from population data from two population censuses and compared to a direct estimate using registered deaths, the completeness of death recording relative to population recording can be estimated. Hill (1987) shows that

$$\frac{\left({}_5N1_{x-5} * {}_5N2_x\right)^{0.5}}{5*\left(N1_{x+} * N2_{x+}\right)^{0.5}} - \frac{1}{t}\ln\left(\frac{N2_{x+}}{N1_{x+}}\right) \approx \frac{1}{t}\ln\left(\frac{k_1}{k_2}\right) + \frac{\left(k_1 * k_2\right)^{0.5}}{c} * \frac{D(x+)}{t*\left(N1_{x+} * N2_{x+}\right)^{0.5}} \quad (1)$$

where *N1* and *N2* are population counts at two time points separated by *t* years, *D* are intercensal deaths, and k_1 , k_2 , and *c* are the completeness, assumed invariant by age, of the first and second populations counts and the intercensal deaths, respectively.
The recorded death rate x+ is thus

$$\frac{D(x+)}{t*(N1_{x+} * N2_{x+})^{0.5}} \quad (2)$$

and the residual estimate of the death rate based on the age distributions is

$$\frac{({}_5N1_{x-5} * {}_5N2_x)^{0.5}}{5*(N1_{x+} * N2_{x+})^{0.5}} - \frac{1}{t}\ln\left(\frac{N2_{x+}}{N1_{x+}}\right) \quad (3).$$

If the assumptions are met, the points for successive age segments x+ should lie on a straight line, the slope of which, $((k_1 * k_2)^{0.5}/c)$, represents the adjustment factor needed for the recorded death rates to bring them into consistency with the population data. Least squares or other line fitting methods can be used to estimate the slope.

2. The Synthetic Extinct Generations Method

The Synthetic Extinct Generations (SEG) method (Bennett and Horiuchi 1981, 1984) uses a distribution of deaths by age above age x together with age-specific growth rates to arrive at an estimate of the population of age *x*, a synthetic analog of Vincent's (1951) method of extinct generations. The completeness of death registration relative to population recording is then estimated by the ratio of the death-based estimate of population aged *x* to the observed population aged *x*. The synthetic estimate of the population aged *x* is given by:

$$\hat{N}(x) = \int_x^{\omega} D(y) e^{\int_x^y r(z)dz} dy \quad (4)$$

where $\hat{N}(x)$ is the estimated population aged *x*, *D(y)* is the observed number of deaths at age *y*, and *r(z)* is the age-specific growth rate of the population at age *z*. The deaths at each age above *x* are adjusted for the cumulative population growth rate between *x* and the age of the deaths to convert them into a stationary population equivalent. Bennett and Horiuchi's 1984 method gives similar results to the 1981 method, but without the diagnostic advantages of estimates of completeness for a range of ages *x*.

3. Adjusted Synthetic Extinct Generations Method

Hill and Choi (2004) used simulations to evaluate how common patterns of data error affect the performance of the GGB and the SEG methods. Their results show that even quite a small change in coverage from one census to the next, if unadjusted for, can seriously bias the completeness estimates from SEG, while SEG estimates seemed to be more robust to a typical pattern of age misreporting than GGB. Taken together, the simulation results suggested that better estimates of death registration completeness can be derived by combining the GGB and SEG methods, first using GGB to estimate any change in census coverage (from the intercept of the fitted straight line), using this estimate to adjust the census counts to be consistent, and then applying SEG to arrive at an adjustment factor for deaths.

Preliminary results

Table 1 shows estimates of completeness for the three methods for 22 countries for the period 1960 to 1970, fitting in each case to the age range 15 to 55. Though the mean and median estimates of completeness are close to 1.0 for all three methods and both sexes, there is substantial variability between countries, from an estimate of completeness (relative to population counts) of 83% (GGB males Switzerland) to 145% (SEG females France). Figure 1 explores the extent to which the variability might be related to migration (the methods all assume that net migration is zero) by plotting the estimate of coverage from Table 1 against the crude net migration rate for the country. Completeness estimates from all three methods are closely associated with net migration. The patterns of association however differ. GGB and the combined GGB/SEG estimates have a negative association with migration, with the estimated completeness becoming lower as net migration increases. Conversely, the association between migration and SEG estimates is positive, with completeness increasing as net migration increases.

Table 2 explores the question of whether the range of ages used to fit points makes a substantial difference to the performance of the methods. The first panel of Table 2 shows the mean, median and standard deviation of completeness estimates by sex from the three methods for all 22 countries, fitting to three different age ranges, 15 to 55, 5 to 65, and 40 to 80. The second, third and fourth panels use the same layout, but dividing the countries into those with significant net emigration (-1 per 1,000 or more), little net migration, and significant net immigration (+ 1 per 1,000 or more).

Estimates fitted to age-group 15 to 55 for all countries show lower mean levels of completeness with both the GGB and GGB/SEG methods. Even though these estimates are lower, they also have higher levels of variability. The lowest completeness levels for estimates fitted to age group 40 to 80 are observed among SEG estimates. In general, estimates for this age range also have the lowest standard deviations.

Table 2 also shows that the performance of different age groups depend on levels of migration. In limited migration countries, estimates fitted to age group 15 to 55 are generally lowest among males and slightly lower among female estimates. While a similar pattern is observed among countries with significant net immigration, mean GGB estimates fitted to age group 15 to 55 are lower in these countries than they are in countries with limited migration. Furthermore, among net emigration countries, mean estimates for age group 15 to 55 are the highest for the GGB and GGB/SEG methods. In these countries also, the means are lowest for age group 5 to 65 when the SEG method is used.

Unlike the younger age groups, age group 40 to 80 has higher means and lower standard deviations. In addition, estimates fitted to this age group are the least likely to vary by the levels of net migration shown in Table 2. Figure 2 brings further clarity to the association between net migration and estimates fitted to age-group 40 to 80. In contrast to the results shown in Figure 1, Figure 2 demonstrates that across method, migration has little or no association with estimates derived from age 40 to 80.

Discussion and Conclusions

Our preliminary results have been limited to estimates for the period 1960 to 1970. The final paper examines estimates for all years between 1950 and 2000. These results however point to several issues relevant to the estimation of death registration completeness and thus, the measurement of mortality. First, estimates of completeness are variable across the age groups used for fitting. Second, estimates fitted to younger age-groups are more likely to be affected by levels of net-migration. However, the association between estimated completeness and migration varies across method. Finally, our findings suggest that age group 40 to 80 may be a better option for fitting completeness estimates. Death registration estimates from this age group have the lowest variability across method and are the least likely to be affected by migration patterns.

References

Banister, J. and K. Hill. (2004) Mortality in China 1964-2000. *Population Studies*, March 2004.

Bennett NG and Horiuchi S. 1981. "Estimating the Completeness of Death Registration in a Closed Population." *Population Index* 47(2):207-21.

Bennett NG and Horiuchi S. 1984. "Mortality estimation from registered deaths in less developed countries." *Demography* 21(2):217-33.

Brass, W. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill, NC: University of North Carolina, Chapel Hill.

Elo, I T and Preston S 1994 "Estimating African American Mortality from Inaccurate Data" *Demography*, Vol. 31, No. 3 (Aug., 1994), pp. 427-458

Hill K. 1987. "Estimating Census and Death Registration Completeness." *Asian and Pacific Population Forum* 1(3):8,23-13,24.

Hill K. and Y. Choi. 2004. Death Distribution Methods for Estimating Adult Mortality: Sensitivity Analysis with Simulated Data Errors. Paper prepared for Adult Mortality in Developing Countries Workshop, The Marconi Center, Marin County, California, July 8th to 11th 2004

Merli G M 1998. "Mortality in Vietnam" *Demography*, Vol. 35, No. 3 (Aug., 1998), pp. 345-360

Vincent. 1951. "The Mortality of the Aged". *Population* 16(2): 182-204

United Nations, Population Division. World Population Prospects - the 2002 revision. New York, 2003.

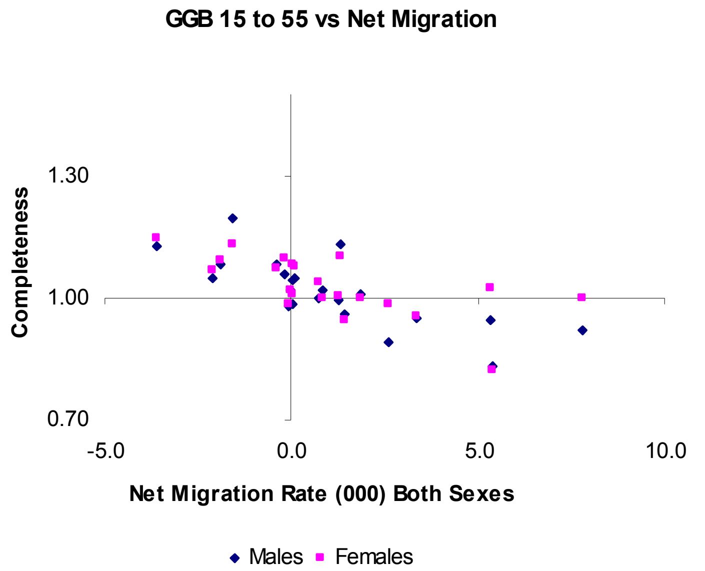
Table 1: Completeness estimates fitted to ages 15 to 55+, by method and net migration rate

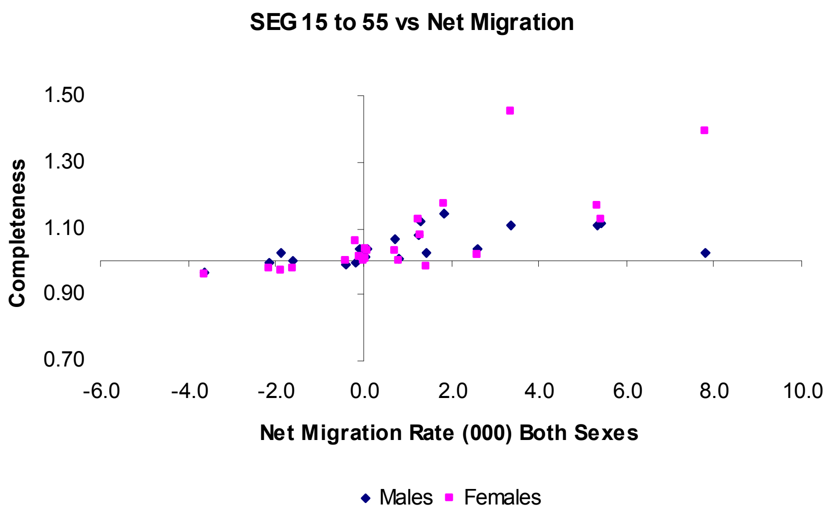
Country	Males			Females			Net Migration Rate Per Thousand
	GGB	SEG	GGB/SEG	GGB	SEG	GGB/SEG	
Australia	0.92	1.03	0.90	1.00	1.39	0.97	7.82
Austria	1.02	1.01	1.02	1.00	1.01	1.00	0.82
Belgium	0.96	1.03	0.95	0.95	0.99	0.94	1.42
Bulgaria	1.06	0.99	1.05	1.10	1.06	1.07	-0.18
Canada	0.94	1.11	0.93	1.03	1.17	1.01	5.34
Czech Republic	1.08	0.99	1.08	1.07	1.00	1.07	-0.41
Denmark	0.99	1.03	0.98	1.01	1.01	1.01	0.04
England and Wales	1.02	1.01	1.02	1.02	1.00	1.02	-0.01
Finland	1.13	0.97	1.13	1.15	0.96	1.13	-3.61
France	0.95	1.11	0.92	0.96	1.45	1.06	3.38
Hungary	1.05	1.04	1.04	1.08	1.03	1.07	0.08
Iceland	1.05	1.00	1.04	1.07	0.98	1.06	-2.13
Italy	1.19	1.00	1.17	1.13	0.98	1.10	-1.59
Japan	0.98	1.04	0.98	0.98	1.01	0.98	-0.08
Netherlands	1.00	1.07	1.00	1.04	1.03	1.03	0.74
New Zealand	1.01	1.14	0.99	1.00	1.17	0.99	1.87
Norway	1.04	1.01	1.03	1.08	1.04	1.06	0.06
Spain	1.08	1.02	1.04	1.09	0.97	1.07	-1.89
Sweden	0.89	1.04	0.89	0.98	1.02	0.91	2.62
Switzerland	0.83	1.12	0.82	0.82	1.13	0.82	5.42
Ukriane	1.13	1.12	1.08	1.10	1.08	1.03	1.32
USA	1.00	1.08	0.99	1.01	1.13	0.99	1.26
Mean	1.015	1.043	1.002	1.030	1.073	1.018	
Median	1.014	1.031	1.010	1.023	1.026	1.026	
Std. Dev.	0.083	0.049	0.081	0.072	0.129	0.069	

Table 2: Summary estimates of completeness, sex, method, and net migration rate

		Males			Females		
All countries		Mean	Median	Std. dev.	Mean	Median	Std. dev.
GGB	15 to 55	1.014	1.015	0.083	1.023	1.030	0.072
	5 to 65	1.026	1.019	0.051	1.034	1.023	0.053
	40 to 80	1.035	1.028	0.043	1.047	1.037	0.038
SEG	15 to 55	1.031	1.043	0.049	1.026	1.073	0.129
	5 to 65	1.036	1.049	0.065	1.024	1.057	0.109
	40 to 80	1.022	1.028	0.019	1.023	1.035	0.047
GGB/SEG	15 to 55	1.010	1.002	0.081	1.026	1.018	0.069
	5 to 65	1.021	1.009	0.056	1.029	1.013	0.054
	40 to 80	1.031	1.017	0.052	1.032	1.028	0.044
Net Immigration counties (Net. Mig. Rate > +1.00)							
GGB							
	15-55	0.960	0.952	0.083	0.982	0.997	0.074
	5-65	0.980	0.981	0.047	0.991	1.014	0.061
	40-80	1.016	1.043	0.066	1.029	1.019	0.054
SEG							
	15-55	1.085	1.106	0.044	1.169	1.129	0.157
	5-65	1.103	1.093	0.067	1.131	1.135	0.140
	40-80	1.031	1.036	0.026	1.049	1.025	0.071
GGB/SEG							
	15-55	0.940	0.934	0.075	0.969	0.986	0.072
	5-65	0.965	0.973	0.052	0.974	1.000	0.059
	40-80	0.996	1.017	0.076	1.009	1.017	0.058
Limited Migration Countries							
GGB							
	15-55	1.027	1.019	0.034	1.042	1.036	0.041
	5-65	1.028	1.027	0.020	1.034	1.035	0.032
	40-80	1.040	1.046	0.015	1.047	1.049	0.023
SEG							
	15-55	1.021	1.013	0.025	1.021	1.013	0.020
	5-65	1.022	1.013	0.025	1.020	1.016	0.019
	40-80	1.025	1.022	0.013	1.030	1.033	0.019
GGB/SEG							
	15-55	1.023	1.023	0.033	1.036	1.034	0.035
	5-65	1.024	1.027	0.021	1.030	1.034	0.028
	40-80	1.039	1.045	0.016	1.046	1.034	0.026
Net Emmigration countries (Net. Mig. Rate > -1.00)							
GGB							
	15-55	1.113	1.105	0.064	1.110	1.112	0.035
	5-65	1.084	1.092	0.031	1.069	1.060	0.021
	40-80	1.025	1.023	0.010	1.035	1.034	0.016
SEG							
	15-55	0.997	1.000	0.024	0.973	0.976	0.010
	5-65	0.989	0.995	0.024	0.971	0.974	0.018
	40-80	1.027	1.028	0.015	1.015	1.015	0.010
GGB/SEG							
	15-55	1.093	1.084	0.066	1.091	1.087	0.034
	5-65	1.077	1.088	0.031	1.061	1.050	0.025
	40-80	1.015	1.007	0.019	1.032	1.029	0.021

Figure 1: Estimates fitted ages 15 to 55 by sex, net migration, and method





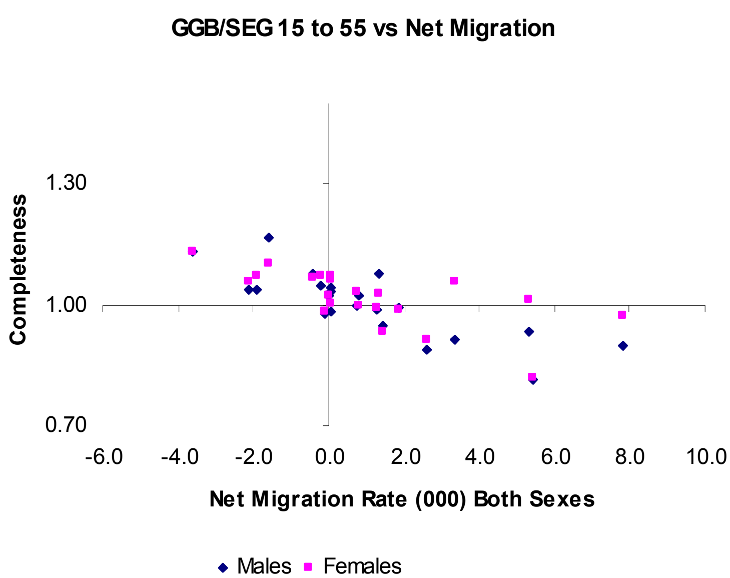


Figure 1: Estimates fitted ages 40 to 80 by sex, net migration, and method