THE ESTIMATION OF CRUDE DEATH RATE WHEN THE DATA AFFECTED BY DIFFERENT UNDER-REGISTRATION

By

Hoda Rashad

Introduction:

The availability and quality of demographic data in developing countries are far from adequate. The introduction and improvements of techniquces for estimating mortality from nontraditional sources of data and for correcting the short comings in traditional data are indispensible.

The under registration of deaths invital statistics data and the reference error associated with data collected in surveys are well known problems facing Demographers. Quite recently, Brass [1] developed a method-Growth balance, method - which makes use of such defective data and provides an estimate of the extent of undergistration.

The basic underlying assumptions of the proposed method are the stability of the age distribution and that underegistration of deaths is equal over all agegroups.

Extensive study revealed that the method is generally robust to patterns of mortality change similar to those in developing countries and also to recent changes in fertility.

The second assumption of equal proportionate underregistration is more likely to apply over the middle age range than for very youngages. Thus, in practise, this method is used for estimating mortality of adult ages only.

It is our purpose in this paper to extend the growth balance method to cover cases when there are two different proportionate underregistration. This is ideally suitable to allow for the different underregistration of young ages since as pointed out by Carrier (1958): 'a substantial proportion of infants die shortly after birth. For a variety of reasons and in a variety of ways this may lead to a proportion of infant deaths being treated differently from deaths at older ages, both as regards disposal of the remains and recording the event. Thus data which gives adequate presentation to deaths at older ages, or at least equal deficiencies at all these ages, are liable to suffer from excessive deficiencies in infant deaths'.

IN principle, of course, the extension of the method may apply to other cases, such as the differential underregistration of old age deaths. The proportionate underregistration of old ages is less or more than the general underregistration according to the significance and role of the older generation in different cultures.

In the following parts we will show that the difference in underregistration may be fully accounted for once the age groups suffering unequal underreport are located. Two numerical applications are illustrated, the first on hypothetical data and the other on actual data for Iraq 1960-1970.

The Method

The general case when the first m age groups suffer from proportionate underregistration ou while age groups from m to M suffer from underregistration u is treated here.

In case $0 > 1$, underregistration for young age groups 1 to m is higher than for age groups m to M. If $0 < 1$ the oppostie occurs.

- The first step is to calculate u:

Using the reported number of deaths and population for ages over m and the relation:

$$\frac{n_y}{p_y} = r + \left(\frac{1}{1-u}\right) \frac{d_y^r}{p_y} \qquad y > m$$

or

using the reported proportions of deaths and population for ages over m and the relations.

$$\frac{N_y}{P_y} = r + CDR'. \frac{D_y^r}{P_y} \qquad y > m$$

$$u = 1 - \frac{\text{total reported deaths}}{\text{total population (CDR)}}$$

where

n_y : number of population aged y. (actual and reported)

p_y : number of population over age y (actual and reported)

r : growth rate.

N_y : proportion of population age y (actual and reported).

P_y : proportion of population over age y.

d_Y^r and D_Y^r denote the number and proportion of reported deaths over age y respectively.

- The second step is to estimate o using the following relations:

$$V_Y = \frac{\frac{N_Y}{P_Y} - r}{CDR'} P_Y - D_Y^r \qquad y < m$$

$$o = \frac{v_Y}{u} + (D_Y^r - D_m^r) \frac{1}{(D_Y^r - D_m^r + v_Y)} \qquad y < m$$

- Finally the actual death rate is equal to:

$$CDR = \left(\frac{\text{Reported death from 1 to m}}{1-ou} + \frac{\text{Reported deaths from m to}}{1-u}\right)$$

total population.

or;

$$CDR = CDR'/k(u,o)$$

where'

$$k(u,o) = 1 - \frac{u(o-1)}{(1-u) + (1-ou) \dfrac{D_m^r}{1-D_m^r}}$$

Mote that when o=1, there is no differential under registration. Then $k(u,o) = 1$ and $CDR = CDR'$.

The proof of this method is given in detail in Appendix (A).

NUMERICAL APPLICATIONS (Using Hypothetical Data):

Application (1)

Starting with a stable distribution, model north, mortality level 11, r = 10.0 corresponding to actual death rate=22.26 given in Coale & Demeny (1966). Subjecting the deaths corresponding to age groups from 0 to 20 to under-report 0.3, while the deaths corresponding to ages from 20 to 80+ are subjected to under-report 0.1 (o =3,u=1). Assuming the total population 100,000 and the total number of actual deaths 2,226, the actual and reported number of deaths and population is presented in Table (1).

The detailed calculations for estimating u are given in Table (2).

Using least square fit, $\frac{1}{1-u} = \frac{\Sigma XY - \bar{X}\,\Sigma Y}{\Sigma X^2 - n\bar{X}^2}$

where $X = \frac{d_y^r}{p_y}$, $Y = \frac{n_y}{p_y}$

$\frac{1}{1-u} = 1.120$, then u = .107.

To estimate v_i, we need to calculate r and CDR'. r is the intercept of the straight line whether using proportion or numbers.

$r = \bar{Y} - 1.120$, $\bar{X} = .009 \cong .01$.

Table (1) The actual and reported number of population and deaths in case of differential underregistration of deaths.

age	Actual data		Reported data
	population	deaths	deaths
0-	2880	487.27	341.08
1-	9870	307.63	215.34
5-	10950	109.29	76.50
10-	10030	51.86	36.30
15-	9280	54.98	38.48
20-	8520	71.89	64.70
25-	7760	69.00	62.10
30-	7050	66.78	60.10
35-	6370	68.78	61.90
40-	5710	74.34	66.91
45-	5060	79.69	71.72
50-	4400	88.14	79.33
55-	3740	97.72	87.94
60-	3040	111.52	100.37
65-	2320	123.99	111.58
70-	1590	130.44	117.39
75-	920	113.30	101.97
80+	510	119.53	107.58
Total	100000	2226.14	1801.29

Table (2) The details of calculating u

age y	number around age y (n_y)	pop. beyond age y (p_y)	reported deaths beyond age y (d_y^r)	X d_y^r/p_y	Y $\frac{n_y}{p_y}$
20	1780	56990	1093.377	.0191	.031
25	1628	48470	1028.708	.0212	.033
30	1481	40710	966.741	.0237	.036
35	1342	33660	906.647	.0269	.039
40	1208	27290	844.732	.0309	.044
45	1077	21580	777.849	.0360	.049
50	946	16520	706.137	.0427	.057
55	814	12120	626.805	.0517	.067
60	678	8380	538.862	.0643	.080
65	536	5340	438.508	.0821	.100
70	391	3020	326.931	.1082	.129
75	251	1430	209.536	.1465	.175

$$CDR' = \frac{\text{Total reported deaths}}{\text{Total population}} \left(\frac{1}{1-u}\right)$$

$$= \frac{1801\ (1.12)}{100,000} = .020$$

The detailed calculations of v_i using ages less than 20 are given in Table (3).

Table (3) The detailed calculation of v_i

age	Y_i	P_i	D_i^r	$\frac{Y_i - r}{CDR'} . P_i$ r= .01, CDR'=.02	$v_i = \frac{Y_i - r}{CDR'} . P_i - D_i^r$
5	.0265	87.25	69.077	71.981	2.904
10	.0274	76.3	64.837	66.381	1.544
15	.0291	66.27	62.827	63.287	.46

To calculate o, the following relation is used:

$$o = \left(\frac{v_i}{u} + (D_i^r - D_m^r)\right) \frac{1}{(v_i + D_i^r - D_m^r)}$$

$$D_m^r = D_{20}^r = 60.697$$

o (using v_1 and D_1^r corresponding to age 5)$= \frac{29.04 + 8.38}{2.904 + 8.38}$

$= 3.31$

o (using v_2 and D_2^r corresponding to age 10)= 3.44

o (using v_3 and D_3^r corresponding to age 15)= 2.60

The mean of the previous values is used as an estimate for O=3.11

$$K(u,o) = 1 - \frac{u(o-1)}{(1-u) + (1-ou)\dfrac{D_m^r}{1-D_m^r}} = .892$$

and finally,

$$\text{actual death rate} = \frac{CDR'}{K(u,o)} = 22.40$$

Thus instead of a reported death rate 18.01%, this method results an estimated death rate = 22.40% which is quite close to the actual death rate = 22.26%.

Numerical Application (2) Using Actual Data):

Brass (1976) applied the growth balance method to vital registration and census statistics for Iraq. It was noticed that the points at higher ages were quite close to linearity; those at younger ages were erratic and displayed a peculiar curvature upwards at the lower end of the graph. Brass suspected different underregistration of deaths at young ages (up to 30 years).

Ignoring the upturn of the lower points, the estimate of f was reached as 1.88 and used to inflate the reported deaths over the range for which the correction was taken as applicable.

To allow for differential underregistration at young ages, the same previous adjustment was extended to ages over 5. It was pointed out that since mortality over 5 was so low, little overall error was expected by this adjustment. To estimate the deaths corresponding to ages less than 5, the south set of Coale & Demeny model life table was used. Level 14 mortality was estimated to correspond to a population with the Iraq distribution and the adjusted death rates over age 10. The adjusted crude death rate for the Iraq age distribution was then estimated as 15.5%. Brass commented that this rate is somewhat lower than expected.

The adjustment procedure - to allow for the differential underregistration is applied using the same data for Iraq. Table (4) presents the original data foe Iraq.

Table (4) Data for Iraq 1960-70, females

age group	number (thousands)	deaths (thousands)
0-4	766.7	2.13
5-9	603.0	.36
10-14	491.2	.34
15-19	343.4	.31
20-29	531.4	.74
30-39	459.2	.87
40-49	315.5	.95
50-59	227.6	1.02
60-69	155.8	1.90
70-79	75.9	4.76
81 over	24.0	

* reproduced from Brass (1976), table 6.

Using the points corresponding to ages over 30:

$\frac{1}{1-u} = 1.883$ $u = .4689$

$CDR' = .0063$ $r = .0262$

The detailed calculations for estimating o are presented in Table (5).

Table (5) The detailed calculations for estimating o

age (Y) (1)	D^r_y (2)	P_y (3)	$\frac{N_y}{P_y} - r/CDR'$ (4)	$v_y = (4).P_y - D^r_y$	o
5	.8404	.8080	2.5873	1.2497	2.07
10	.8139	.6567	2.4603	.8017	2.00
15	.7884	.5340	2.0634	.3134	1.91
20	.7653	.4480	1.9047	.0880	1.70

Thus o = 2.

The adjusted death rate, assuming the underregistration under age 30 is twice the underregistration over age 30= 20%.

In view of the previous discussion and the near constancy of o, the adjusted death rate seems much more reasonable than the reported rate of 3.35%.

Note that, assuming the underregistration over age 5 is the same as the underregistration over age 30, the adjusted death rate = 14%. This is quite close to the estimate provided by Brass.

APPENDIX (A)

A.1 DEFINITIONS

M : total number of age groups.

d_i: actual number of deaths in age group i. i=1,2, ..., M.

u : proportionate under-registration in age groups m to M ($0 < u < 1$).
u = (under-registered deaths/actual deaths).

ou: proportionate under-registration in the remaining age groups (1 to m).

m : number of age groups experiencing under-report ou.

N_y: actual and reported population proportion per year of age around the point y.

P_y: actual and reported population proportion over age y.

D^r_y: reported proportion of deaths over age y. (reported deaths over age y/total reported deaths for all ages).

D_y: actual proportion of deaths beyond age y.

Y_y: N_y/P_y.

X^r_y: D^r_y/P_y.

r : growth rate.

CDR: actual death rate.

X_y: D_y/P_y

A.2 RESULTS

In section (4) we will prove that:

For $i > m$:

1. $Y_i = r + CDR' \cdot X_i^r$

where

$CDR' = CDR.K(u,o)$

$$K(u,o) = 1 - \frac{u(o-1)\sum_{x=1}^{m} d_x}{(1-u)\sum_{x=1}^{M} d_x}$$

$$= 1 - \frac{u(o-1)}{(1-u) + (1-ou)\dfrac{D_m^r}{1-D_m^r}}$$

Also,

2. $u = 1 - \dfrac{\text{Total reported deaths}}{CDR'.\ \text{Total population}}$

For $i < m$:

3. $v_i = \dfrac{(D_i^r - D_m^r)\, u(o-1)}{(1-ou)}$

Also,

4. $v_i = \frac{Y_i - r}{CDR'} P_i - D_i^r.$

A.3 METHOD

Using the reported population and deaths for age groups m to M, CDR', r and u are estimated using (1) and (2).

Using relation (3) and (4) o is estimated as:

$$o = (\frac{v_i}{u} + (D_i^r - D_m^r)) \frac{1}{(D_i^r - D_m^r + v_i)}$$

Finally, the reported deaths are adjusted using u and o, thus:

$$CDR = (\frac{\text{Reported deaths from 1 to m}}{(1-ou)} + \frac{\text{Reported deaths from m to M}}{(1-u)}$$

/ total population.

or

$CDR = CDR'/K(u,o).$

A.4 PROOF

For $i > m$:

$$1. \quad D_i = \frac{\sum_{x=i}^{M} d_x}{\sum_{x=1}^{M} d_x} = \frac{\sum_{x=i}^{M} d_x(1-u)}{\sum_{x=1}^{M} d_x(1-u)} \qquad (a.1)$$

$$D_i^r = \frac{\sum_{x=i}^{M} d_x(1-u)}{\sum_{x=1}^{m} d_x(1-ou) + \sum_{x=m}^{m} d_x(1-u)}$$

$$D_i^r = \frac{\sum_{x=i}^{M} d_x(1-u)}{\sum_{x=1}^{m} d_x(1-u+u-ou) + \sum_{x=m}^{M} d_x(1-u)}$$

then

$$D_i^r = \frac{\sum_{x=i}^{M} d_x(1-u)}{\sum_{x=1}^{M} d_x(1-u) + u(1-o-) \sum_{x=1}^{m} d_x}$$

dividing the nominator and denominator by $\sum_{x=1}^{M} d_x(1-u)$

and using (a.1) we get:

$$D_i^r = \frac{D_i}{1 - \frac{u(o-1) \sum_{x=1}^{m} d_x}{(1-u) \sum_{x=1}^{M} d_x}}$$

Thus,

$$D_i^r = D_i / K(u,o) \qquad (a.2)$$

Where

$$K(u,o) = 1 - \frac{u(o-1)\sum_{x=1}^{m} d_x}{(1-u)\sum_{x=1}^{M} d_x}$$

but

$$x_i^r = \frac{D_i^r}{P_i}$$

then

$$X_i^r = \frac{D_i}{K(u,o)P_i} = \frac{X_i}{K(u,o)}, \qquad (a.3)$$

since

$$Y_i = r + CDR.X_i$$

using (a.3)

$$Y_i = r + CDR.K(u,o).X_i^r.$$

Finally,

$$Y_i = r + CDR' \cdot X_i^r \quad ,$$

where CDR' = CDR.K(u,o)

$$K(u,o) = 1 - \frac{u(o-1)\sum_{x=1}^{m} d_x}{(1-u)\sum_{x=1}^{M} d_x} \quad .$$

To show that K(u,o) may be re-expressed in terms of the reported deaths as:

$$K(u,o) = 1 - \frac{u(o-1)}{(1-u) + (1-ou)\dfrac{D_m^r}{1-D_m^r}}$$

since

$$K(u,o) = 1 - \frac{u(o-1)\sum_{x=1}^{m} d_x}{(1-u)\sum_{x=1}^{M} d_x}$$

$$= 1 - \frac{u(o-1)\sum_{x=1}^{m} d_x}{(1-u)\left[\sum_{x=1}^{m} d_x + \sum_{x=m}^{M} d_x\right]}$$

$$= 1 - \frac{u(o-1)}{(1-u)\left[1 + \frac{\sum_{x=m}^{M} d_x}{\sum_{x=1}^{m} d_x}\right]}$$

$$= 1 - \frac{u(o-1)}{(1-u)\left[1 + \frac{(1-ou)}{(1-u)} \cdot \frac{\sum_{x=m}^{M} d_x(1-u)}{\sum_{x=1}^{m} d_x(1-ou)}\right]}$$

and

$$K(u,o) = 1 - \frac{u(o-1)}{(1-u) + (1-ou)\,\frac{D_m^r}{1 - D_m^r}}$$

This completes the proof of (A.1).

2. Using the reported number of deaths for ages over m, and the relation (2.3)

$$\frac{n_y}{p_y} = r + f\,\frac{d_y^r}{p_y}$$

where f is the ratio of the true deaths over age m to the reported deaths over age m.

Thus, $f = \dfrac{\text{true deaths over age m}}{\text{true deaths over age m - under-registered deat}}$
over age m

$$f = \frac{1}{1-u}$$

since

$$Y_i = r + CDR'.X_i^r.$$

Then

$$\frac{CDR'(\text{total population})}{(\text{total reported deaths}} = \frac{1}{1=u}$$

and

$$u = 1 - \frac{\text{total reported deaths}}{CDR'.\text{total population.}}$$

For $i < m$:

3. $$D_i^r = \frac{\sum_{x=i}^{m} d_x(1-o\ u)+ \sum_{x=m}^{M} d_x(1-u)}{\sum_{y=1}^{m} d_x(1-o\ u)+ \sum_{x=m}^{M} d_x(1-u)} \qquad (a.4)$$

$$D_i^r = \frac{\sum_{x=i}^{m} d_x(1-u+u-ou)+ \sum_{x=m}^{M} d_x(1-u)}{\sum_{y=1}^{m} d_x(1-u+u-ou)+ \sum_{x=m}^{M} d_x(1-u)}$$

$$= \frac{\sum_{x=i}^{M} d_x(1-u) - u(o-1)\sum_{x=i}^{m} d_x}{\sum_{x=1}^{M} d_x(1-u) - u(o-1)\sum_{x=1}^{m} d_x}$$

dividing the nominator and denominator by $(1-u)\sum_{x=1}^{M} d_x$, we get:

$$v_i = \frac{\dfrac{\sum_{x=i}^{m} d_x}{\sum_{x=1}^{m} d_x}}{\dfrac{(1-u)\sum_{x=1}^{M} d_x}{u(o-1)\sum_{x=1}^{m} d_x} - 1}$$

$$v_i = \frac{\dfrac{\sum_{x=i}^{m} d_x(1-ou)}{\sum_{x=1}^{m} d_x(1-ou)}}{\dfrac{(1-u)}{u(0-1)}\left\{1 + \dfrac{(1-ou)\sum_{x=m}^{M} d_x(1-u)}{(1-u)\sum_{x=1}^{m} d_x(1-ou)}\right\} - 1} \qquad (a.7)$$

but

$$\frac{\sum_{x=i}^{m} d_x(1-ou)}{\sum_{x=1}^{m} d_x(1-ou)} = \frac{\sum_{x=i}^{m} d_x(1-ou) + \sum_{x=m}^{M} d_x(1-u) - \sum_{x=m}^{M} d_x(1-u)}{\sum_{x=1}^{m} d_x(1-ou) + \sum_{x=m}^{M} d_x(1-u) - \sum_{x=m}^{M} d_x(1-u)}$$

$$= \frac{D_i^r - D_m^r}{1-D_m^r}$$

substituting in (a.7)

$$v_i = \frac{\dfrac{D_i^r - D_m^r}{1-D_m^r}}{\dfrac{(1-u)}{u(o-1)} \quad 1+ \dfrac{(1-ou)D_m^r}{(1-u)(1-D_m^r)} \quad -1}$$

$$D_i^r = \frac{D_i - \dfrac{u(o-1) \sum_{x=i}^{m} d_x}{(1-u) \sum_{x=1}^{M} d_x}}{1-\dfrac{u(o-1) \sum_{x=1}^{m} d_x}{(1-u) \sum_{x=1}^{M} d_x}}$$

$$D_i^r = \frac{D_i}{K(u,o)} - v_i$$

Thus

$$v_i = \frac{u(o-1) \sum_{x=1}^{m} d_x}{K(u,o)(1-u) \sum_{x=1}^{M} d_x} \qquad (a.5)$$

$$v_i = \frac{D_i}{K(u,o)} \doteq D_i^r \qquad (a.6)$$

Rewriting (a.5)

$$v_i = \frac{\dfrac{u(o-1) \sum_{x=1}^{m} d_x}{(1-u) \sum_{x=1}^{M} d_x}}{1 - \dfrac{u(o-1) \sum_{x=1}^{m} d_x}{(1-u) \sum_{x=1}^{M} d_x}}$$

Multiplying the nominator and denominator by

$$\frac{(1-u) \sum_{x=1}^{M} d_x}{u(o-1) \sum_{x=1}^{m} d_x}, \text{ we get}$$

$$v_i = \frac{(D_i^r - D_m^r)}{\frac{(1-u)(1-D_m^r)}{u(o-1)} + \frac{(1-ou)\ D_m^r}{u(o-1)} - (1-D_m^r}$$

and finally,

$$v_i \doteq \frac{(D_i^r - D_m^r)u(o-1)}{(1-ou)}$$

4. From (a.6)

$$v_i = \frac{D_i}{K(u,o)} - D_i^r$$

$$\frac{Y_i - R}{CDR'} = \frac{(Y_i - r)}{CDR.K(u,o)} = \frac{X_i}{K(o,u)} = \frac{D_i}{P_i.K(o,u)}$$

then

$$\frac{D_i}{K(o,u)} = \frac{Y_i - r}{CDR'} .P_i$$

and

$$v_i = \frac{Y_i - r}{CDR'} \ P_i - D_i^r.$$