Lecture 6
Assessing the Completeness of Death Registration
Multiple Census Methods

Michel Guillot
University of Pennsylvania & French Institute for Demographic Studies
Training Workshop on CRVS evaluation, Amman, 28-30 June 2022

Outline

- Generalized Growth Balance (GGB) method
- Synthetic Extinct Generation (SEG) method

Data Requirements and assumptions for both methods

- Data requirements for both methods
  - Age distribution of population at two successive censuses
  - Age distribution of deaths for intercensal period

- Assumptions
  - Coverage of each census is the same for all ages
  - Completeness of reporting of deaths is the same for all ages from a minimum age (usually age 15)
  - The population is closed to migration (or information on migration is available)
  - <u>No assumption of stability</u>

Generalized Growth Balance Method - theory

- Recall from Brass Growth Balance method
- In any closed population, between time t1 and t2:
  - CGR = CBR – CDR (or r = b – d)
- This expression also holds for open-ended age segments *x*+, deaths being the deaths at ages *x* and over, "births" being birthdays at age *x*, and the denominators being the population aged *x* and over
  - r(x+) = b(x+) – d(x+)
    - b(x+) = N(x)/N(x+)
    - d(x+) = D(x+)/N(x+)
  - b(x+) – r(x+) = d(x+)

Generalized Growth Balance Method - theory

- Assume that deaths are registered with a completeness c, constant at all ages: $d^r(x+)=d(x+)*c$
  - $d(x+)$ = true death rate at ages x+
  - $d^r(x+)$ = recorded death rate at ages x+
- $d(x+) = d^r(x+)/c$
- $b(x+) – r(x+) = \{1/c\}*d^r(x+)$
- If we can estimate *b(x+)* and *r(x+)* from census data, there should exist a linear relation between the residual *(b(x+) – r(x+))* and the direct estimate of the death rate $d^r(x+)$. The slope of the line should estimate {1/*c*}, the coverage of deaths.

Generalized Growth Balance Method - theory

- The expression

$$b(x+) – r(x+) = \{1/c\}*d^r(x+)$$

is the equation of a straight line, slope 1/*c*, but no intercept term (i.e. passes through the origin).

- In practice, observed points may indicate that the relation does not pass through the origin. This indicates that the two censuses may have different completeness.

Generalized Growth Balance Method - theory

- Assume the coverage of the first census is k1
- Assume the coverage of the second census is k2
- Assume coverage is constant with age.
- Then equation becomes:

$$b(x+) - r(x+) = \frac{\ln(k_1/k_2)}{t_2 - t_1} + \frac{\sqrt{k_1 k_2}}{c} d^r(x+)$$

- Estimate b(x+), r(x+) and $d^r(x+)$ from data. Estimate intercept *a* and slope *b* of equation using linear regression. Completeness *c* can be derived from *a* and *b*.

Generalized Growth Balance Method – Data Requirements

- Number of women (men), by five year-age group, and for open age interval A+ (with A as high as possible), at two points in time, typically from the results of two censuses.
- Number of deaths of women (men), by five-year age group, and for open age interval A+, over the period between the two censuses.

Generalized Growth Balance Method - steps

- Step 1: Estimate the number of deaths reported in the period between the two census.
  - Apportion annual deaths in the first and last year of the census

- Example: Suppose census dates are April 1, 2000 and April 1, 2010. Then intercensal deaths are:

$$\mathbf{.75} * {}_5D_x(2000) + {}_5D_x(2001) + {}_5D_x(2002) + \ldots + {}_5D_x(2009) + \mathbf{.25} * {}_5D_x(2010)$$

Generalized Growth Balance Method - steps

- Step 2: Cumulate population and deaths downwards

- Population: $N(x+) = \sum_{y=x}^{A-5} {}_5N_y + {}_\infty N_A$

- Same for deaths D(x+)

Generalized Growth Balance Method - steps

- Step 3: Calculate the Person-Years of life lived, PYL(x+)

$$PYL(x+) = (t_2 - t_1)\left(N(x+, t_1) \times N(x+, t_2)\right)^{\frac{1}{2}}$$

- Step 4: Calculate the number of people who turned x in the population, N(x)

$$N(x) = \frac{t}{5}\left({}_5N_{x-5}(t_1) \times {}_5N_x(t_2) \right)^{\frac{1}{2}}$$

Generalized Growth Balance Method - steps

- Step 5: Calculate b(x+), d(x+) and r(x+)

$$b(x+) = \frac{N(x)}{PYL(x+)}$$

$$d(x+) = \frac{D(x+)}{PYL(x+)}$$

$$r(x+) = \frac{{}_{\infty}N_x(t_2) - {}_{\infty}N_x(t_1)}{PYL(x+)}$$

Generalized Growth Balance Method - steps

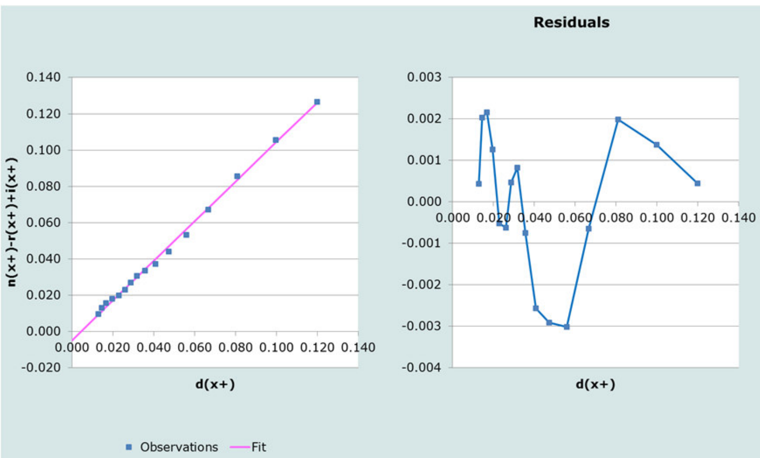
- Step 6: Plot graph of b(x+)-r(x+) against d(x+)
- Step 7: Fit line, estimate a and b, and then c
  - If intercept *a* is negative:

$$c = \frac{e^{a(t_2 - t_1)}}{b}$$

  - If intercept *a* is positive:

$$c = \frac{e^{-a(t_2 - t_1)}}{b}$$

Example: South Africa, Males, 2001-07



Intercept a = -.00467
Slope b = 1.0907

Completeness c = .9175 = **92%**

Limitations

- Migration
- Fluctuation of completeness of death registration by age

Synthetic Extinct Generation method - theory

- Extension of the Preston and Coale method for populations that are not stable and for which two censuses are available.
- In a stable population:

$N(80,t) = D(80,t) + D(81,t) \times e^{r} + D(82,t) \times e^{2r} + \ \ldots$

- In any closed population:

$N(80,t) = D(80,t) + D(81,t) \times e^{\ {}_1r_{80}} + D(82,t) \times e^{\left(\ {}_1r_{80} + \ {}_1r_{81}\right)} + \ \ldots$

where ${}_1r_x$ is the growth rate of the population aged x, x+1

Synthetic Extinct Generation method - theory

- In reality we have D_a^r (=recorded deaths) rather than D_a (=true deaths)
- If deaths are not completely recorded, then $D_a^r = c*D_a$, or $D_a = D_a^r/c$
- If we use D_a^r instead of D_a when summing deaths, we obtain $\widehat{N}_x$ instead of the true population N_x
- The ratio $\frac{\widehat{N}_x}{N_x}$ gives an estimate of completeness c

Generalized Growth Balance Method – Data Requirements

- Number of women (men), by five year-age group, and for open age interval A+ (with A as high as possible), at two points in time, typically from the results of two censuses.
- Number of deaths of women (men), by five-year age group, and for open age interval A+, over the period between the two censuses.

Synthetic Extinct Generation method - steps

- Step 1: Estimate the number of deaths reported in the period between the two census.
  - Apportion annual deaths in the first and last year of the census
- Step 2: estimate age-specific growth rates between the two censuses, adjusted for change in census coverage
- $_5r_x = \frac{ln\left({}_5N_x(t_2)/ {}_5N_x(t_1)\right)}{t_2 - t_1} + \delta$
  - δ = intercept *a* estimated in Generalized Growth Balance method

Synthetic Extinct Generation method - steps

- Step 3: Estimate life expectancy at age A and five-year age intervals down to 65
  - Use estimates from Generalized Growth Balance method (or other sources if available)

Synthetic Extinct Generation method - steps

- Step 4: Estimate the number of people who turned x, and the number of people aged x, x+5, from the reported deaths

$$\widehat{N}_x = \widehat{N}_{x+5}\exp(5\,{}_5r_x) + {}_5D_x\exp(2.5\,{}_5r_x)$$

$$\widehat{N}_A = {}_\infty D_A\left(\exp({}_\infty r_A \times e_A) - ({}_\infty r_A \times e_A)^2/6\right)$$

$${}_5\widehat{N}_x = 2.5\left(\widehat{N}_x + \widehat{N}_{x+5}\right)$$

Synthetic Extinct Generation method - steps

- Step 5: Estimate the number of people aged x, x+5 during the period between the two censuses, from the census populations

$${}_5N_x = (t_2 - t_1)\,({}_5N_x(t_1) \times {}_5N_x(t_2))^{1/2}$$

- Step 6: Calculate the ratios of estimated over observed ${}_5N_x$ as well as and ratios of estimated over observed ${}_{A-x}N_x$.

$$\frac{{}_5\widehat{N}_x}{{}_5N_x}\ and\ \frac{{}_{A-x}\widehat{N}_x}{{}_{A-x}N_x}$$

Synthetic Extinct Generation method - steps

- Step 7: Estimate the completeness of reported deaths
  - Adjust δ if necessary, using data/solver in spreadsheet
  - Decide about age range of ratios

- More complex procedure using iteration is described in IUSSP volume.

Synthetic Extinct Generation method - limitations

- Incorrect estimate of relative coverage of the censuses δ
- Exaggeration of reported age
- Age misstatement in population estimates and age-specific miscounting

GGB vs. SEG

- Both methods robust to age misreporting
- GGB more sensitive to coverage errors that change with age
- SEG sensitive to time change in census coverage
    - Importance of using δ adjustment
- Fit using 5+ to 65+ better than 15+ to 55+
- Both GGB and SEG are very sensitive to net migration
- If little migration: Use GGB first, estimate δ, then apply SEG
- If significant migration: Apply both GGB and SEG, fit 30+ to 65+, and average results of both methods