Death by Survey

Estimating Adult Mortality without Selection Bias

Emmanuela Gakidou Gary King

July 20, 2005

Sources of Mortality Data

Sources of Mortality Data

- Vital (and sample) registration systems
  (death certificates, mostly in developed countries)

Sources of Mortality Data

- Vital (and sample) registration systems
  (death certificates, mostly in developed countries)
- Demographic surveillance systems
  (a few isolated projects)

Sources of Mortality Data

- Vital (and sample) registration systems
  (death certificates, mostly in developed countries)
- Demographic surveillance systems
  (a few isolated projects)
- Wild guesses
  (typically reported in the media, usually citing other wild guesses as authorities)

Who Uses Mortality Data?

- Medical researchers

Who Uses Mortality Data?

- Medical researchers
- Public health researchers

Who Uses Mortality Data?

- Medical researchers
- Public health researchers
- Public policy makers

Who Uses Mortality Data?

- Medical researchers
- Public health researchers
- Public policy makers
- Demographers

Who Uses Mortality Data?

- Medical researchers
- Public health researchers
- Public policy makers
- Demographers
- Epidemiologists

Who Uses Mortality Data?

- Medical researchers
- Public health researchers
- Public policy makers
- Demographers
- Epidemiologists
- Sociologists

Who Uses Mortality Data?

- Medical researchers
- Public health researchers
- Public policy makers
- Demographers
- Epidemiologists
- Sociologists
- Political scientists:

Who Uses Mortality Data?

- Medical researchers
- Public health researchers
- Public policy makers
- Demographers
- Epidemiologists
- Sociologists
- Political scientists:
  - IR studies: the elite decision to go to war

Who Uses Mortality Data?

- Medical researchers
- Public health researchers
- Public policy makers
- Demographers
- Epidemiologists
- Sociologists
- Political scientists:
    - IR studies: the elite decision to go to war
    - Should also study: more ultimate outcomes, like human misery or mortality

Who Uses Mortality Data?

- Medical researchers
- Public health researchers
- Public policy makers
- Demographers
- Epidemiologists
- Sociologists
- Political scientists:
  - IR studies: the elite decision to go to war
  - Should also study: more ultimate outcomes, like human misery or mortality
  - The Big Problems: uncertainty and selection bias

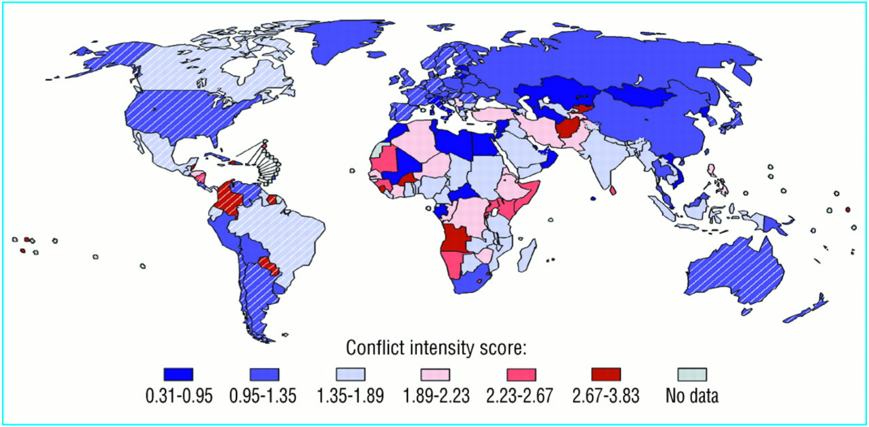
Uncertainty in Death Estimates from Major Sources

The 1990s' Ten Most Deadly Conflicts

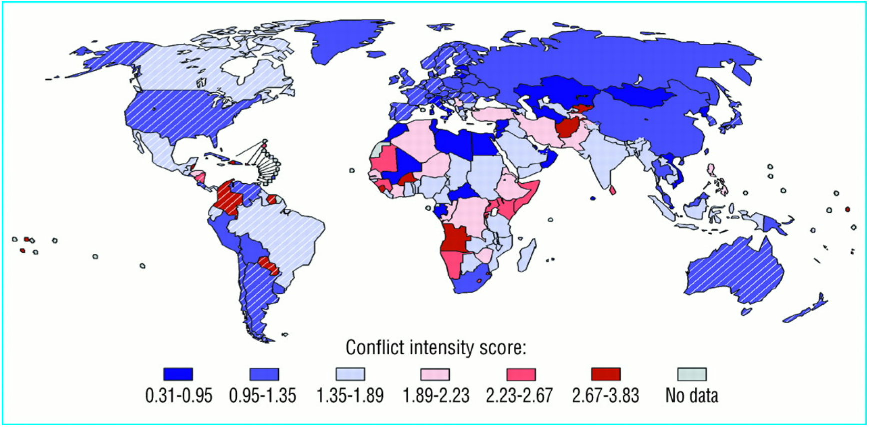
		People Killed (in 1000s)		
Country	Year	Low	High	Range
Rwanda	1994	500	1,000	500
Angola	1992–4	100	500	400
Somalia	1991–9	48	300	252
Bosnia	1992–5	35	250	215
Liberia	1991–6	25	200	175
Burundi	1993	30	200	170
Chechnya	1994–6	30	90	60
Tajikistan	1992–9	20	120	100
Algeria	1992–9	30	100	70
Gulf war	1990–1	4.3	100	95.7

Selection bias in Mortality from War Data

Selection bias in Mortality from War Data

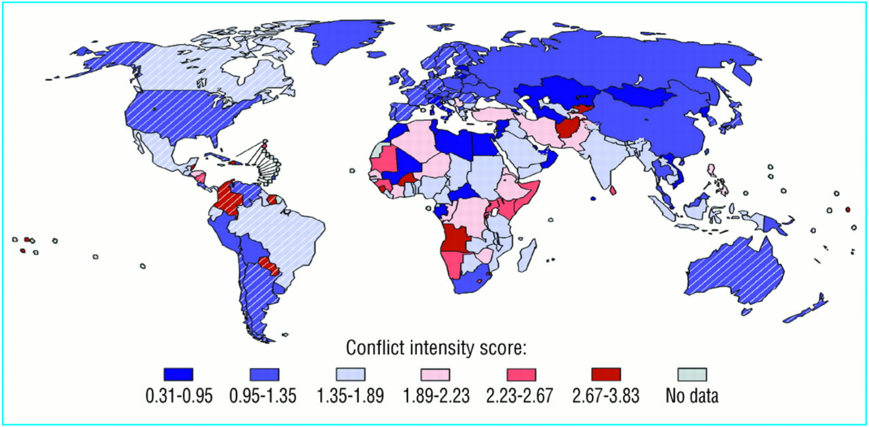


Selection bias in Mortality from War Data



Vital registration areas (cross-hatched): mostly low conflict

Selection bias in Mortality from War Data



Vital registration areas (cross-hatched): mostly low conflict
No registration areas: much higher conflict

Methods for estimating mortality, using surveys of:

Methods for estimating mortality, using surveys of:

- Parents (Brass and Hill, 1973; Hill and Trussell 1977; Timaeus 1991b, 1986)

Methods for estimating mortality, using surveys of:

- Parents (Brass and Hill, 1973; Hill and Trussell 1977; Timaeus 1991b, 1986)
- Siblings (Bicego 1997; Chiphangwi et al. 1992; Danel et al. 1996; Gakidou, Hogan, and Lopez 2004; Garenne and Friedberg 1997; Graham, Brass, and Snow, 1989; Shahidullah 1995; Shiferaw and Tessema 1993; Timaeus and Ali 2001; Walraven and van Dongen 1994; Wirawan and Linnan 1994)

Methods for estimating mortality, using surveys of:

- Parents (Brass and Hill, 1973; Hill and Trussell 1977; Timaeus 1991b, 1986)
- Siblings (Bicego 1997; Chiphangwi et al. 1992; Danel et al. 1996; Gakidou, Hogan, and Lopez 2004; Garenne and Friedberg 1997; Graham, Brass, and Snow, 1989; Shahidullah 1995; Shiferaw and Tessema 1993; Timaeus and Ali 2001; Walraven and van Dongen 1994; Wirawan and Linnan 1994)
- Spouses (Malaker 1986; Stanton, Noureddine and Hill 2000; Singh 2000; Timaeus 1991)

Methods for estimating mortality, using surveys of:

- Parents (Brass and Hill, 1973; Hill and Trussell 1977; Timaeus 1991b, 1986)
- Siblings (Bicego 1997; Chiphangwi et al. 1992; Danel et al. 1996; Gakidou, Hogan, and Lopez 2004; Garenne and Friedberg 1997; Graham, Brass, and Snow, 1989; Shahidullah 1995; Shiferaw and Tessema 1993; Timaeus and Ali 2001; Walraven and van Dongen 1994; Wirawan and Linnan 1994)
- Spouses (Malaker 1986; Stanton, Noureddine and Hill 2000; Singh 2000; Timaeus 1991)
- Household residents and others (Feeney 2001; Graham, Brass, and Snow, 1989)

Extensive Efforts to Develop and Apply These Methods

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography
- Epidemiology
- Public health
- Medicine
- Sociology

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology
- Public health
- Medicine
- Sociology

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology – good statistics in narrow areas
- Public health
- Medicine
- Sociology

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology – good statistics in narrow areas
- Public health – collections of other disciplines
- Medicine
- Sociology

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology – good statistics in narrow areas
- Public health – collections of other disciplines
- Medicine – users, not methodological innovators
- Sociology

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology – good statistics in narrow areas
- Public health – collections of other disciplines
- Medicine – users, not methodological innovators
- Sociology

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology – good statistics in narrow areas
- Public health – collections of other disciplines
- Medicine – users, not methodological innovators
- Sociology

Major global effort to collect data on adult mortality

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology – good statistics in narrow areas
- Public health – collections of other disciplines
- Medicine – users, not methodological innovators
- Sociology

Major global effort to collect data on adult mortality

- World Health Survey (80 countries)

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology – good statistics in narrow areas
- Public health – collections of other disciplines
- Medicine – users, not methodological innovators
- Sociology

Major global effort to collect data on adult mortality

- World Health Survey (80 countries)
- Demographic and Health Surveys (70 countries)

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology – good statistics in narrow areas
- Public health – collections of other disciplines
- Medicine – users, not methodological innovators
- Sociology

Major global effort to collect data on adult mortality

- World Health Survey (80 countries)
- Demographic and Health Surveys (70 countries)
- Gates Grand Challenge surveys ($20M, about to begin)

Extensive Efforts to Develop and Apply These Methods

Several hundred articles published in:

- Demography – numbers without statistics
- Epidemiology – good statistics in narrow areas
- Public health – collections of other disciplines
- Medicine – users, not methodological innovators
- Sociology

Major global effort to collect data on adult mortality

- World Health Survey (80 countries)
- Demographic and Health Surveys (70 countries)
- Gates Grand Challenge surveys ($20M, about to begin)
- Ellison Institute Surveys ($100M, about to begin)

Methodology

Methodology

The Tradeoff

The Tradeoff

Vital registration: small biases but very expensive

The Tradeoff

Vital registration: small biases but very expensive

Surveys: large biases but inexpensive

Methodology

The Tradeoff

Vital registration: small biases but very expensive

Surveys: large biases but inexpensive

- Selection the Dependent Variable: high mortality families less likely to be surveyed

Methodology

The Tradeoff

Vital registration: small biases but very expensive

Surveys: large biases but inexpensive

- Selection the Dependent Variable: high mortality families less likely to be surveyed
- Can biases can be corrected?

Methodology

The Tradeoff

Vital registration: small biases but very expensive

Surveys: large biases but inexpensive

- Selection the Dependent Variable: high mortality families less likely to be surveyed
- Can biases can be corrected?

The Constraint: Statistics with your hands tied behind your back

Methodology

The Tradeoff

Vital registration: small biases but very expensive

Surveys: large biases but inexpensive

- Selection the Dependent Variable: high mortality families less likely to be surveyed
- Can biases can be corrected?

The Constraint: Statistics with your hands tied behind your back

For demographers to accept: methods must be simple and transparent.

Methodology

The Tradeoff

Vital registration: small biases but very expensive

Surveys: large biases but inexpensive

- Selection the Dependent Variable: high mortality families less likely to be surveyed
- Can biases can be corrected?

The Constraint: Statistics with your hands tied behind your back

For demographers to accept: methods must be simple and transparent.

The Opportunity

Methodology

The Tradeoff

Vital registration: small biases but very expensive

Surveys: large biases but inexpensive

- Selection the Dependent Variable: high mortality families less likely to be surveyed
- Can biases can be corrected?

The Constraint: Statistics with your hands tied behind your back

For demographers to accept: methods must be simple and transparent.

The Opportunity

- Joining Demography and Statistics (through political science)

Methodology

The Tradeoff

Vital registration: small biases but very expensive

Surveys: large biases but inexpensive

- Selection the Dependent Variable: high mortality families less likely to be surveyed
- Can biases can be corrected?

The Constraint: Statistics with your hands tied behind your back

For demographers to accept: methods must be simple and transparent.

The Opportunity

- Joining Demography and Statistics (through political science)
- Directing public health dollars (9% of the world's economy)

Methodology

The Tradeoff

Vital registration: small biases but very expensive

Surveys: large biases but inexpensive

- Selection the Dependent Variable: high mortality families less likely to be surveyed
- Can biases can be corrected?

The Constraint: Statistics with your hands tied behind your back

For demographers to accept: methods must be simple and transparent.

The Opportunity

- Joining Demography and Statistics (through political science)
- Directing public health dollars (9% of the world's economy)
- Understanding the consequences of war and the causes of war

A Simplified Quantity of Interest

A Simplified Quantity of Interest

- Consider a cohort of men who turn 20 on 5/1/1980

A Simplified Quantity of Interest

- Consider a cohort of men who turn 20 on 5/1/1980
- Track the individuals for 10 years

A Simplified Quantity of Interest

- Consider a cohort of men who turn 20 on 5/1/1980
- Track the individuals for 10 years
- Quantity of interest: Proportion dead between times 1 and 2

A Simplified Quantity of Interest

- Consider a cohort of men who turn 20 on 5/1/1980
- Track the individuals for 10 years
- Quantity of interest: Proportion dead between times 1 and 2

Mortality =

q =

d_j	Sibship
0	1
1	1
0	1
1	2
1	2

A Simplified Quantity of Interest

- Consider a cohort of men who turn 20 on 5/1/1980
- Track the individuals for 10 years
- Quantity of interest: Proportion dead between times 1 and 2

$$\text{Mortality} = \frac{\text{Deaths}}{\text{“Births”}}$$

$$q = \frac{\sum_{j=1}^{N} d_j}{N}$$

d_j	Sibship
0	1
1	1
0	1
1	2
1	2

A Simplified Quantity of Interest

- Consider a cohort of men who turn 20 on 5/1/1980
- Track the individuals for 10 years
- Quantity of interest: Proportion dead between times 1 and 2

$$\text{Mortality} = \frac{\text{Deaths}}{\text{“Births”}} = \frac{\sum_{j=1}^{N} (\text{Sibship Mortality})_j}{\text{“Births”}}$$

$$q = \frac{\sum_{j=1}^{N} d_j}{N} = \frac{\sum_{j=1}^{N} M_j}{N}$$

d_j	M_j	Sibship
0	1/3	1
1	1/3	1
0	1/3	1
1	2/2	2
1	2/2	2

A Simplified Quantity of Interest

- Consider a cohort of men who turn 20 on 5/1/1980
- Track the individuals for 10 years
- Quantity of interest: Proportion dead between times 1 and 2

$$\text{Mortality} = \frac{\text{Deaths}}{\text{``Births''}} = \frac{\sum_{j=1}^{N} (\text{Sibship Mortality})_j}{\text{``Births''}}$$

$$q = \frac{\sum_{j=1}^{N} d_j}{N} = \frac{\sum_{j=1}^{N} M_j}{N}$$

d_j	M_j	Sibship
0	1/3	1
1	1/3	1
0	1/3	1
1	2/2	2
1	2/2	2

- Applied as estimators to (time 1) samples: unbiased

Estimation Problems with Time 2 Samples

Estimation Problems with Time 2 Samples

Vital registration data

d_j	M_j	Sibship
0	1/3	1
1	1/3	1
0	1/3	1
1	2/3	2
1	2/3	2
0	2/3	2
1	2/2	3
1	2/2	3

Estimation Problems with Time 2 Samples

Vital registration data

d_j	M_j	Sibship
0	1/3	1
1	1/3	1
0	1/3	1
1	2/3	2
1	2/3	2
0	2/3	2
1	2/2	3
1	2/2	3

(Unobserved at time 2)

Estimation Problems with Time 2 Samples

Vital registration data

d_j	M_j	Sibship
0	1/3	1
1	1/3	1
0	1/3	1
1	2/3	2
1	2/3	2
0	2/3	2
1	2/2	3
1	2/2	3

(Unobserved at time 2)

$\Longrightarrow$

Data available to sample at time 2

d_i	M_i
0	1/3
0	1/3
0	2/3

Estimation Problems with Time 2 Samples

Vital registration data

d_j	M_j	Sibship
0	1/3	1
1	1/3	1
0	1/3	1
1	2/3	2
1	2/3	2
0	2/3	2
1	2/2	3
1	2/2	3

(Unobserved at time 2)

$\Longrightarrow$

Data available to sample at time 2

d_i	M_i
0	1/3
0	1/3
0	2/3

- Each respondent reports on entire sibship

Estimation Problems with Time 2 Samples

Vital registration data

d_j	M_j	Sibship
0	1/3	1
1	1/3	1
0	1/3	1
1	2/3	2
1	2/3	2
0	2/3	2
1	2/2	3
1	2/2	3

(Unobserved at time 2)

$\Longrightarrow$

Data available to sample at time 2

d_i	M_i
0	1/3
0	1/3
0	2/3

- Each respondent reports on entire sibship
- High mortality sibships: underrepresented

Estimation Problems with Time 2 Samples

Vital registration data

d_j	M_j	Sibship
0	1/3	1
1	1/3	1
0	1/3	1
1	2/3	2
1	2/3	2
0	2/3	2
1	2/2	3
1	2/2	3

(Unobserved at time 2)

$\Longrightarrow$

Data available to sample at time 2

d_i	M_i
0	1/3
0	1/3
0	2/3

- Each respondent reports on entire sibship
- High mortality sibships: underrepresented
- No sibships with 0 survivors

Estimation Problems with Time 2 Samples

Vital registration data

d_j	M_j	Sibship
0	1/3	1
1	1/3	1
0	1/3	1
1	2/3	2
1	2/3	2
0	2/3	2
1	2/2	3
1	2/2	3

(Unobserved at time 2)

$\Longrightarrow$

Data available to sample at time 2

d_i	M_i
0	1/3
0	1/3
0	2/3

- Each respondent reports on entire sibship
- High mortality sibships: underrepresented
- No sibships with 0 survivors
- Some families counted more than once

First existing approach: The *Naive* Estimator

First existing approach: The *Naive* Estimator

$$\dot{q} =$$

First existing approach: The *Naive* Estimator

$$\dot{q} = \frac{\text{Reported Deaths}}{\text{Reported ``Births''}} =$$

First existing approach: The *Naive* Estimator

$$\dot{q} = \frac{\text{Reported Deaths}}{\text{Reported ``Births''}} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i}$$

First existing approach: The *Naive* Estimator

$$\dot{q} = \frac{\text{Reported Deaths}}{\text{Reported ``Births''}} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i}$$

Estimator is biased downward:

First existing approach: The *Naive* Estimator

$$\dot{q} = \frac{\text{Reported Deaths}}{\text{Reported ``Births''}} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i}$$

Estimator is biased downward:

- High mortality sibships: underrepresented

First existing approach: The *Naive* Estimator

$$\dot{q} = \frac{\text{Reported Deaths}}{\text{Reported "Births"}} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i}$$

Estimator is biased downward:

- High mortality sibships: underrepresented
- Sibships with 0 survivors: not represented

First existing approach: The *Naive* Estimator

$$\dot{q} = \frac{\text{Reported Deaths}}{\text{Reported ``Births''}} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i}$$

Estimator is biased downward:

- High mortality sibships: underrepresented
- Sibships with 0 survivors: not represented
- Sample includes n "survivors" by design

Second existing approach: The *Standard* Estimator

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} =$$

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} = \frac{\text{Reported Deaths } -0}{\text{Reported “Births” } -n}$$

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} = \frac{\text{Reported Deaths } -0}{\text{Reported ``Births'' } -n} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i - n}$$

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} = \frac{\text{Reported Deaths } -0}{\text{Reported "Births" } -n} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i - n}$$

- Contributing biases:

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} = \frac{\text{Reported Deaths } -0}{\text{Reported "Births" } -n} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i - n}$$

- Contributing biases:
  - High mortality sibships underrepresented: downward bias

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} = \frac{\text{Reported Deaths } -0}{\text{Reported ``Births'' } -n} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i - n}$$

- Contributing biases:
  - High mortality sibships underrepresented: downward bias
  - Sibships with 0 survivors not represented: downward bias

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} = \frac{\text{Reported Deaths } -0}{\text{Reported ``Births'' } -n} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i - n}$$

- Contributing biases:
  - High mortality sibships underrepresented: downward bias
  - Sibships with 0 survivors not represented: downward bias
  - Respondents (always alive) are not counted: upward bias

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} = \frac{\text{Reported Deaths } -0}{\text{Reported "Births" } -n} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i - n}$$

- Contributing biases:
  - High mortality sibships underrepresented: downward bias
  - Sibships with 0 survivors not represented: downward bias
  - Respondents (always alive) are not counted: upward bias
  - Multiply counting of the same siblings, overrepresenting low mortality families: downward bias

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} = \frac{\text{Reported Deaths } -0}{\text{Reported ``Births'' } -n} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i - n}$$

- Contributing biases:
  - High mortality sibships underrepresented: downward bias
  - Sibships with 0 survivors not represented: downward bias
  - Respondents (always alive) are not counted: upward bias
  - Multiply counting of the same siblings, overrepresenting low mortality families: downward bias

A little miracle occurs

Second existing approach: The *Standard* Estimator

Standard = Naive, omitting self-reports:

$$\check{q} = \frac{\text{Reported Deaths } -0}{\text{Reported “Births” } -n} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i - n}$$

- Contributing biases:
  - High mortality sibships underrepresented: downward bias
  - Sibships with 0 survivors not represented: downward bias
  - Respondents (always alive) are not counted: upward bias
  - Multiply counting of the same siblings, overrepresenting low mortality families: downward bias

A little miracle occurs

Trussell and Rodriguez (1990) prove: if mortality is independent of sibship size, all biases cancel: $\check{q}$ is unbiased.

Is mortality independent of sibship size?

Is mortality independent of sibship size?

Peru	2000
Indonesia	1997
Burkina Faso	1998
Benin	1996
Peru	1996
Nigeria	1999
Philippines	1998
Chad	1997
Brazil	1996
Indonesia	1994
Senegal	1999
Philippines	1993
Mali	1996
Tanzania	1996

Guinea	1999
Zimbabwe	1994
Nepal	1996
Cameroon	1998
Cote D'Ivoire	1994
Togo	1998
Eritrea	1995
Ethiopia	2000
Zimbabwe	1999
Colombia	1995
Zambia	1996
Uganda	1995
Madagascar	1997

Is mortality independent of sibship size? No!

		Correlation			Correlation
Peru	2000	0.97	Guinea	1999	0.80
Indonesia	1997	0.96	Zimbabwe	1994	0.76
Burkina Faso	1998	0.95	Nepal	1996	0.75
Benin	1996	0.95	Cameroon	1998	0.75
Peru	1996	0.95	Cote D'Ivoire	1994	0.75
Nigeria	1999	0.93	Togo	1998	0.74
Philippines	1998	0.93	Eritrea	1995	0.70
Chad	1997	0.93	Ethiopia	2000	0.71
Brazil	1996	0.92	Zimbabwe	1999	0.69
Indonesia	1994	0.91	Colombia	1995	0.52
Senegal	1999	0.90	Zambia	1996	0.47
Philippines	1993	0.88	Uganda	1995	-0.06
Mali	1996	0.86	Madagascar	1997	-0.19
Tanzania	1996	0.82			

An Alternative Estimator

An Alternative Estimator

- No assumption about sibship size and mortality.

An Alternative Estimator

- No assumption about sibship size and mortality.
- Two problems addressed separately:

An Alternative Estimator

- No assumption about sibship size and mortality.
- Two problems addressed separately:
  1. Underrepresentation of high mortality families

An Alternative Estimator

- No assumption about sibship size and mortality.
- Two problems addressed separately:
  1. Underrepresentation of high mortality families
  2. Nonrepresentation of families with zero survivors

Underrepresentation of high mortality families

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}}$

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}} = \frac{B_i - S_i}{B_i}$

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}} = \frac{B_i - S_i}{B_i}$
- Sampling sibships at time 1: proportional to “Births” (B_i)

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}} = \frac{B_i - S_i}{B_i}$
- Sampling sibships at time 1: proportional to “Births” (B_i)
- Sampling sibships at time 2: proportional to Survivors (S_i)

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}} = \frac{B_i - S_i}{B_i}$
- Sampling sibships at time 1: proportional to “Births” (B_i)
- Sampling sibships at time 2: proportional to Survivors (S_i)
- Estimation at time 1: simple average is unbiased

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}} = \frac{B_i - S_i}{B_i}$
- Sampling sibships at time 1: proportional to “Births” (B_i)
- Sampling sibships at time 2: proportional to Survivors (S_i)
- Estimation at time 1: simple average is unbiased
- Estimation at time 2:

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}} = \frac{B_i - S_i}{B_i}$
- Sampling sibships at time 1: proportional to “Births” (B_i)
- Sampling sibships at time 2: proportional to Survivors (S_i)
- Estimation at time 1: simple average is unbiased
- Estimation at time 2:
  - Weight sample from $\propto S_i$ to $\propto B_i$ by: $W_i = B_i / S_i$

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}} = \frac{B_i - S_i}{B_i}$
- Sampling sibships at time 1: proportional to “Births” (B_i)
- Sampling sibships at time 2: proportional to Survivors (S_i)
- Estimation at time 1: simple average is unbiased
- Estimation at time 2:
  - Weight sample from $\propto S_i$ to $\propto B_i$ by: $W_i = B_i / S_i$
  - The weighted average is unbiased:

$$\frac{\sum_{i=1}^{n} W_i M_i}{\sum_{i=1}^{n} W_i}$$

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}} = \frac{B_i - S_i}{B_i}$
- Sampling sibships at time 1: proportional to “Births” (B_i)
- Sampling sibships at time 2: proportional to Survivors (S_i)
- Estimation at time 1: simple average is unbiased
- Estimation at time 2:
  - Weight sample from $\propto S_i$ to $\propto B_i$ by: $W_i = B_i / S_i$
  - The weighted average is unbiased:

$$\frac{\sum_{i=1}^{n} W_i M_i}{\sum_{i=1}^{n} W_i}$$

- This portion of the problem is solved exactly

Underrepresentation of high mortality families

(Temporarily assuming no families with 0 survivors)

- Sibship mortality: $M_i = \frac{\text{“Births”} - \text{Survivors}}{\text{“Births”}} = \frac{B_i - S_i}{B_i}$
- Sampling sibships at time 1: proportional to “Births” (B_i)
- Sampling sibships at time 2: proportional to Survivors (S_i)
- Estimation at time 1: simple average is unbiased
- Estimation at time 2:
  - Weight sample from $\propto S_i$ to $\propto B_i$ by: $W_i = B_i/S_i$
  - The weighted average is unbiased:

$$\frac{\sum_{i=1}^{n} W_i M_i}{\sum_{i=1}^{n} W_i}$$

- This portion of the problem is solved exactly
- Weights are common; quantities of interest that serve as their own weights are not.

Correcting for Families with 0 Survivors

Correcting for Families with 0 Survivors

- Missing information: ζ, number of siblings in families with zero survivors

Correcting for Families with 0 Survivors

- Missing information: ζ, number of siblings in families with zero survivors
- Mortality estimator, if we had an estimator for ζ:

$$\hat{q} = \frac{\sum_{i=1}^{n} M_i W_i + \hat{\zeta}}{\sum_{i=1}^{n} W_i + \hat{\zeta}}.$$

Correcting for Families with 0 Survivors

- Missing information: ζ, number of siblings in families with zero survivors
- Mortality estimator, if we had an estimator for ζ:

$$\hat{q} = \frac{\sum_{i=1}^{n} M_i W_i + \hat{\zeta}}{\sum_{i=1}^{n} W_i + \hat{\zeta}}.$$

- Direct information in our data about families without survivors: None

Correcting for Families with 0 Survivors

- Missing information: ζ, number of siblings in families with zero survivors
- Mortality estimator, if we had an estimator for ζ:

$$\hat{q} = \frac{\sum_{i=1}^{n} M_i W_i + \hat{\zeta}}{\sum_{i=1}^{n} W_i + \hat{\zeta}}.$$

- Direct information in our data about families without survivors: None
- Approach: we extrapolate from families with survivors

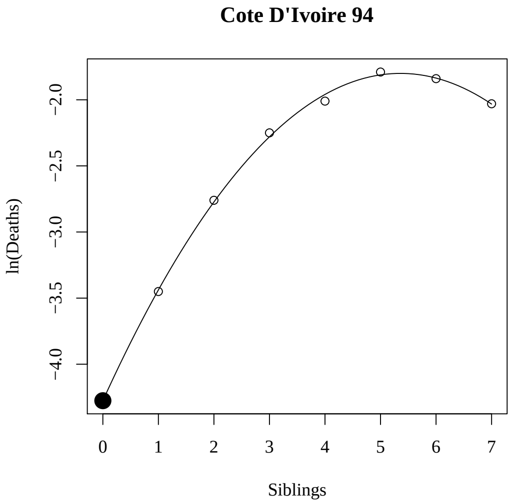
Correcting for Families with 0 Survivors

- Missing information: ζ, number of siblings in families with zero survivors
- Mortality estimator, if we had an estimator for ζ:

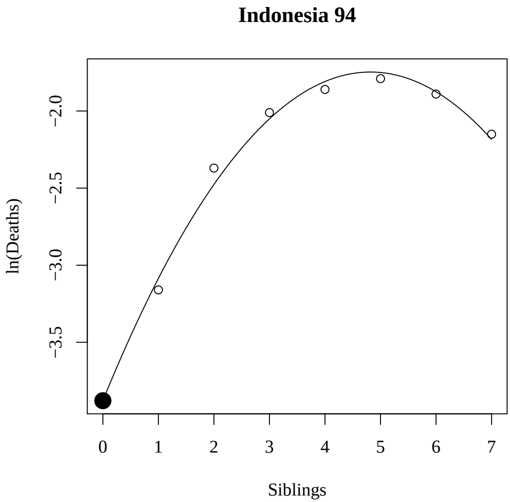
$$\hat{q} = \frac{\sum_{i=1}^{n} M_i W_i + \hat{\zeta}}{\sum_{i=1}^{n} W_i + \hat{\zeta}}.$$

- Direct information in our data about families without survivors: None
- Approach: we extrapolate from families with survivors
- Thus, this part of the answer is more uncertain

Extrapolation to Deaths in Families with 0 Survivors

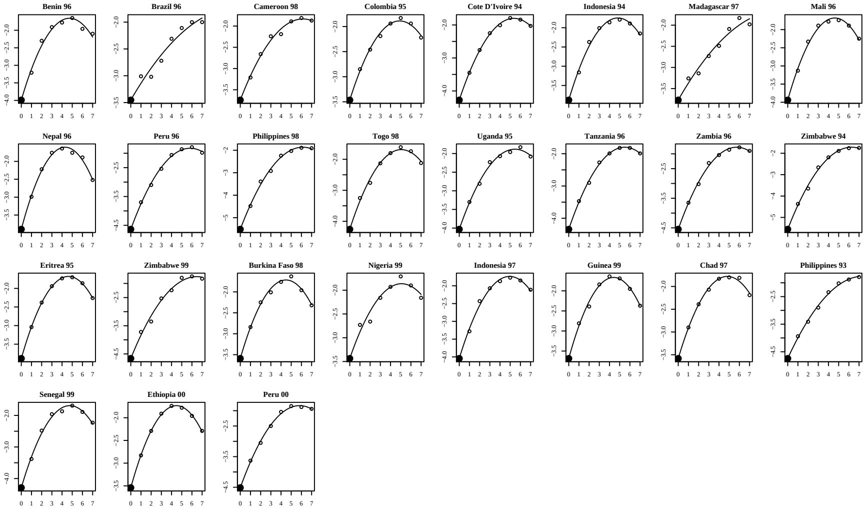


Extrapolation to Deaths in Families with 0 Survivors



Extrapolation to Deaths in Families with 0 Survivors

Extrapolation to Deaths in Families with 0 Survivors



Monte Carlo Simulation Setup: Create 27 Populations

Monte Carlo Simulation Setup: Create 27 Populations

Formed from cross-classifications of:

Monte Carlo Simulation Setup: Create 27 Populations

Formed from cross-classifications of:

- Average mortality: low (0.1), medium (0.2), high (0.3)

Monte Carlo Simulation Setup: Create 27 Populations

Formed from cross-classifications of:

- Average mortality: low (0.1), medium (0.2), high (0.3)
- Average fertility levels: low (Kazakstan, 2.56), medium (Turkey, 3.07), high (Kenya, 4.26 children)

Monte Carlo Simulation Setup: Create 27 Populations

Formed from cross-classifications of:

- Average mortality: low (0.1), medium (0.2), high (0.3)
- Average fertility levels: low (Kazakstan, 2.56), medium (Turkey, 3.07), high (Kenya, 4.26 children)
- Correlation(sibship size, mortality): positive, zero, negative

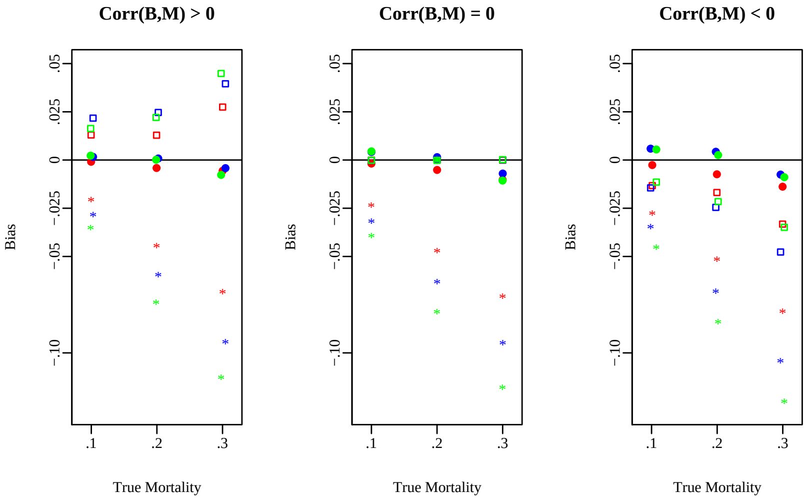
Monte Carlo Simulation Setup: Create 27 Populations

Formed from cross-classifications of:

- Average mortality: low (0.1), medium (0.2), high (0.3)
- Average fertility levels: low (Kazakstan, 2.56), medium (Turkey, 3.07), high (Kenya, 4.26 children)
- Correlation(sibship size, mortality): positive, zero, negative

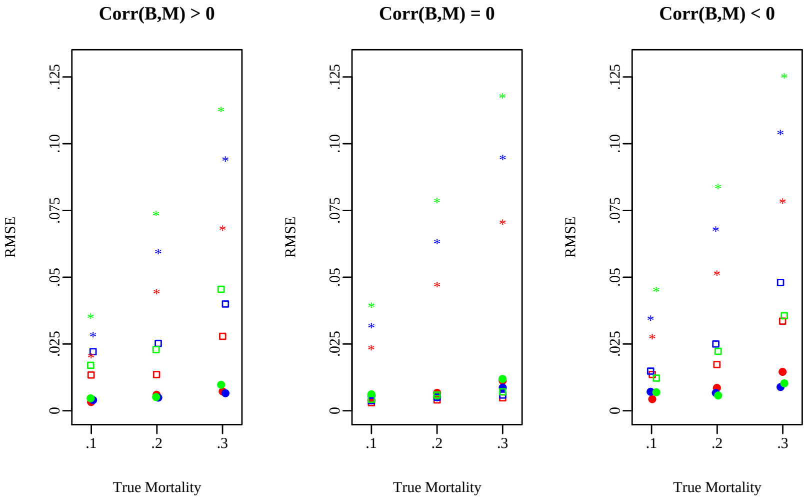
For each, create 1,000 data sets, each with $n = 1,000$ randomly drawn time 2 survey respondents

Bias



Corr(B,M) > 0
Corr(B,M) = 0
Corr(B,M) < 0
Bias
.05
.025
0
−.025
−.05
−.10
.1
.2
.3
True Mortality

Root Mean Square Error



Sources of Bias Reduction

Across the scenerios:

Sources of Bias Reduction

Across the scenerios:

- Most bias reduction is due to weighting

Sources of Bias Reduction

Across the scenerios:

- Most bias reduction is due to weighting
    - 72% of bias corrected on average

Sources of Bias Reduction

Across the scenerios:

- Most bias reduction is due to weighting
  - 72% of bias corrected on average
  - Largest reduction in countries with high positive correlations, high fertility, and low mortality (leading to fewest deaths in families with 0 survivors)

Sources of Bias Reduction

Across the scenerios:

- Most bias reduction is due to weighting
  - 72% of bias corrected on average
  - Largest reduction in countries with high positive correlations, high fertility, and low mortality (leading to fewest deaths in families with 0 survivors)
- Weighting plus extrapolation for families with 0 survivors: 92% corrected

Alternative Characterization of the New Estimator

Alternative Characterization of the New Estimator

- Weighted average = Unweighted average in an appropriate pseudo-sample

Alternative Characterization of the New Estimator

- Weighted average = Unweighted average in an appropriate pseudo-sample
- Arbitrarily choose: "Survivors" $= n$

Alternative Characterization of the New Estimator

- Weighted average = Unweighted average in an appropriate pseudo-sample
- Arbitrarily choose: "Survivors" $= n$
- Each survivor counts for D_i/S_i deaths in the pseudo-sample

Alternative Characterization of the New Estimator

- Weighted average = Unweighted average in an appropriate pseudo-sample
- Arbitrarily choose: "Survivors" $= n$
- Each survivor counts for D_i/S_i deaths in the pseudo-sample
- Thus, compute: "Deaths" $= \sum_{i=1}^{n}(D_i/S_i) + \hat{\zeta}$

Alternative Characterization of the New Estimator

- Weighted average = Unweighted average in an appropriate pseudo-sample
- Arbitrarily choose: "Survivors" $= n$
- Each survivor counts for D_i/S_i deaths in the pseudo-sample
- Thus, compute: "Deaths" $= \sum_{i=1}^{n}(D_i/S_i) + \hat{\zeta}$
- Thus, an equivalent expression for $\hat{q}$:

$$\hat{q} = \frac{\text{Deaths}}{\text{Deaths} + \text{Survivors}} = \frac{\left[\sum_{i=1}^{n}(D_i/S_i) + \hat{\zeta}\right]}{\left[\sum_{i=1}^{n}(D_i/S_i) + \hat{\zeta}\right] + n},$$

Generalization: Asking Women about Male Mortality

Generalization: Asking Women about Male Mortality

- Drop females with no male siblings

Generalization: Asking Women about Male Mortality

- Drop females with no male siblings
- Each female respondent represents $R_i = (S_i/B_i)/(S_i^f/B_i^f)$ males in the pseudo-sample (the sibship survival rate ratio)

Generalization: Asking Women about Male Mortality

- Drop females with no male siblings
- Each female respondent represents $R_i = (S_i/B_i)/(S_i^f/B_i^f)$ males in the pseudo-sample (the sibship survival rate ratio)
- So set: "Survivors" = "males" + "females"

Generalization: Asking Women about Male Mortality

- Drop females with no male siblings
- Each female respondent represents $R_i = (S_i/B_i)/(S_i^f/B_i^f)$ males in the pseudo-sample (the sibship survival rate ratio)
- So set: "Survivors" = "males" + "females" = $n_m + \sum_{i=1}^{n_f} R_i$

Generalization: Asking Women about Male Mortality

- Drop females with no male siblings
- Each female respondent represents $R_i = (S_i/B_i)/(S_i^f/B_i^f)$ males in the pseudo-sample (the sibship survival rate ratio)
- So set: "Survivors" = "males" + "females" $= n_m + \sum_{i=1}^{n_f} R_i$
- Example: if male survival is twice females, each female represents two males in the pseudo-sample; so $R = 2$.

Generalization: Asking Women about Male Mortality

- Drop females with no male siblings
- Each female respondent represents $R_i = (S_i/B_i)/(S_i^f/B_i^f)$ males in the pseudo-sample (the sibship survival rate ratio)
- So set: "Survivors" = "males" + "females" = $n_m + \sum_{i=1}^{n_f} R_i$
- Example: if male survival is twice females, each female represents two males in the pseudo-sample; so $R = 2$.
- Calculate:

$$\text{“Deaths”} = \sum_{i=1}^{n_m} \frac{D_i}{S_i} + \sum_{i'=1}^{n_f} R_{i'} \frac{D_{i'}}{S_{i'}} + \hat{\zeta}$$

Generalization: Asking Women about Male Mortality

- Drop females with no male siblings
- Each female respondent represents $R_i = (S_i/B_i)/(S_i^f/B_i^f)$ males in the pseudo-sample (the sibship survival rate ratio)
- So set: "Survivors" = "males" + "females" = $n_m + \sum_{i=1}^{n_f} R_i$
- Example: if male survival is twice females, each female represents two males in the pseudo-sample; so $R = 2$.
- Calculate:

$$\text{"Deaths"} = \sum_{i=1}^{n_m} \frac{D_i}{S_i} + \sum_{i'=1}^{n_f} R_{i'} \frac{D_{i'}}{S_{i'}} + \hat{\zeta}$$

- The estimator is still Deaths/(Deaths + Survivors).

Generalization: Asking Women about Male Mortality

- Drop females with no male siblings
- Each female respondent represents $R_i = (S_i/B_i)/(S_i^f/B_i^f)$ males in the pseudo-sample (the sibship survival rate ratio)
- So set: "Survivors" = "males" + "females" = $n_m + \sum_{i=1}^{n_f} R_i$
- Example: if male survival is twice females, each female represents two males in the pseudo-sample; so $R = 2$.
- Calculate:

$$\text{“Deaths”} = \sum_{i=1}^{n_m} \frac{D_i}{S_i} + \sum_{i'=1}^{n_f} R_{i'} \frac{D_{i'}}{S_{i'}} + \hat{\zeta}$$

- The estimator is still Deaths/(Deaths + Survivors).
- Can also survey parents, neighbors, teachers, coworkers, etc.

Generalization: Asking Women about Male Mortality

- Drop females with no male siblings
- Each female respondent represents $R_i = (S_i/B_i)/(S_i^f/B_i^f)$ males in the pseudo-sample (the sibship survival rate ratio)
- So set: "Survivors" = "males" + "females" = $n_m + \sum_{i=1}^{n_f} R_i$
- Example: if male survival is twice females, each female represents two males in the pseudo-sample; so $R = 2$.
- Calculate:

$$\text{“Deaths”} = \sum_{i=1}^{n_m} \frac{D_i}{S_i} + \sum_{i'=1}^{n_f} R_{i'} \frac{D_{i'}}{S_{i'}} + \hat{\zeta}$$

- The estimator is still Deaths/(Deaths + Survivors).
- Can also survey parents, neighbors, teachers, coworkers, etc.
- May also estimate maternal mortality, the mortality of parents from data on (adult) children, sisters, etc.

Person-Year Quantities

Person-Year Quantities

- Thus far, all quantities defined for a fixed cohort

Person-Year Quantities

- Thus far, all quantities defined for a fixed cohort
- Men who turn 20 on 5/1/1980, followed for 10 years

Person-Year Quantities

- Thus far, all quantities defined for a fixed cohort
- Men who turn 20 on 5/1/1980, followed for 10 years
- Demographers define person-years, counting any fraction of time a respondent spends in the designated cohort and time interval.

Person-Year Quantities

- Thus far, all quantities defined for a fixed cohort
- Men who turn 20 on 5/1/1980, followed for 10 years
- Demographers define person-years, counting any fraction of time a respondent spends in the designated cohort and time interval.
- A man who turns 20 on 5/1/1985 counts as half a respondent

Person-Year Quantities

- Thus far, all quantities defined for a fixed cohort
- Men who turn 20 on 5/1/1980, followed for 10 years
- Demographers define person-years, counting any fraction of time a respondent spends in the designated cohort and time interval.
- A man who turns 20 on 5/1/1985 counts as half a respondent
- All the methods we discussed generalize to person-years

Concluding Remarks

Concluding Remarks

Required parts of the new estimator

Concluding Remarks

Required parts of the new estimator

1. a weight variable: constructed without external information

Concluding Remarks

Required parts of the new estimator

1. a weight variable: constructed without external information
2. an extrapolation, riskier but apparently reasonable

Concluding Remarks

Required parts of the new estimator

1. a weight variable: constructed without external information
2. an extrapolation, riskier but apparently reasonable

Future research

Concluding Remarks

Required parts of the new estimator

1. a weight variable: constructed without external information
2. an extrapolation, riskier but apparently reasonable

Future research

1. Reducing recall bias, such as via prompting questions, the time of last contact with (and physical distance to) relatives

Concluding Remarks

Required parts of the new estimator

1. a weight variable: constructed without external information
2. an extrapolation, riskier but apparently reasonable

Future research

1. Reducing recall bias, such as via prompting questions, the time of last contact with (and physical distance to) relatives
2. Reducing temporal recall, such as by using memorable events such as wars, famines, etc.

Concluding Remarks

Required parts of the new estimator

1. a weight variable: constructed without external information
2. an extrapolation, riskier but apparently reasonable

Future research

1. Reducing recall bias, such as via prompting questions, the time of last contact with (and physical distance to) relatives
2. Reducing temporal recall, such as by using memorable events such as wars, famines, etc.
3. Validation of surveys in areas with valid vital registration systems or established demographic surveillance.

Concluding Remarks

Required parts of the new estimator

1. a weight variable: constructed without external information
2. an extrapolation, riskier but apparently reasonable

Future research

1. Reducing recall bias, such as via prompting questions, the time of last contact with (and physical distance to) relatives
2. Reducing temporal recall, such as by using memorable events such as wars, famines, etc.
3. Validation of surveys in areas with valid vital registration systems or established demographic surveillance.
4. Extensive applications worldwide

Concluding Remarks

Required parts of the new estimator

1. a weight variable: constructed without external information
2. an extrapolation, riskier but apparently reasonable

Future research

1. Reducing recall bias, such as via prompting questions, the time of last contact with (and physical distance to) relatives
2. Reducing temporal recall, such as by using memorable events such as wars, famines, etc.
3. Validation of surveys in areas with valid vital registration systems or established demographic surveillance.
4. Extensive applications worldwide
5. Ideally, a new subfield within IR predicting mortality and human misery with war and the predictors of war.

This paper and other information

http://GKing.Harvard.edu