Probabilistic Indirect Estimation of Under-Five Mortality

Carl P. Schmertmann *
Center for Demography and Population Health, Florida State University

October 31, 2019

Abstract

The probability of death between birth and age five, q(5), is arguably the most important public health indicator in the world's poorest countries. Many countries in which it is critical to measure child mortality lack good vital registration systems, however, and direct estimation of q5 is impossible without age-specific population and death data. In such situations demographers use survey data, such as DHS surveys, and indirect methods. I introduce and test a new, probabilistic approach to the classic Brass problem of indirect estimation of q5 from aggregate data on children ever born and children dead by age of women. Bayesian models exploit the same regularities as classic Brass methods, but without restrictive assumptions about demographic constants. A Bayesian approach does not require fixed temporal patterns in rates, and it can use data from multiple maternal age groups to estimate current q5. It can also produce uncertainty measures that incorporate both sampling error and potential errors in demographic assumptions. The proposed appoach uses simple parametric models to describe fertility and mortality levels and patterns over the period covering approximately 30 years before the survey, and treats the parameters in these models as unknown quantities with prior distributions.

Keywords: indirect estimation, Brass method, under-five mortality, Bayesian

*I thank participants at the IUSSP Workshop on Mortality Monitoring in the Era of the SDGs, Dec 2018 in Rostock, Germany, and especially Katie Wilson and Jon Wakefield, who have shared closely-related work on Bayesian indirect estimation

1 Introduction

The probability of death between birth and age five, which demographers usually denote as q_5 or as the under-five mortality rate (U5MR), is arguably the most important public health indicator in many of the world's poor countries. Many countries in which it is critical to measure child mortality lack good vital registration systems, however, and direct estimation of q_5 is impossible without age-specific population and death data. In such situations demographers use survey data, such as Demographic and Health Surveys, and indirect methods.

There is a large methodological literature on estimating U5MR from surveys, spanning at least half a century (e.g., Brass & Coale 1968, Rajaratnam et al. 2010, Alkema & New 2014, Verhulst 2016, Wilson & Wakefield 2018). In this paper I introduce a new probabilistic approach to the classic Brass problem of indirect estimation from aggregate data on children ever born and children dead by age of women.[1]

2 Notation

Analysis of maternal birth histories depends on the age patterns of fertility and child mortality at the time the survey, and on the recent history of changes in vital rates. In order to establish a vocabulary for discussing these patterns, define

- $f(m,t)$ as the fertility rate at maternal age m and time t
- $\mu(z,t)$ as the mortality rate at age z and time t
- period cumulative mortality at age x and time t as $H_p(x,t) = \int_0^x \mu(z,t)\,dz$
- cohort cumulative mortality at age x and time t as $H_c(x,t) = \int_0^x \mu[\,z,\, t-x+z)]\,dz$

For a survey at time $t = 0$, also define

[1] Data in this form is usually called Summary Birth Histories (SBH). This contrasts with Full Birth History (FBH) survey data, which includes detailed timing information on individual births and child deaths for each respondent.

- $f^*(m) = f(m, 0)$ as the period fertility schedule
- $\mu^*(z) = \mu(z, 0)$ as the period mortality schedule
- period cumulative mortality at age x as $H_p^*(x) = \int_0^x \mu^*(z)\, dz$
- cohort cumulative mortality at age x as $H_c^*(x) = \int_0^x \mu(z, z-x)\, dz$

where an asterisk always denotes a quantity measured on the survey date $t = 0$.

With these definitions the expected fraction dead among all children born x years before the survey (a cohort measure) is

$$q_c^*(x) = 1 - e^{-H_c^*(x)} \tag{1}$$

Among the a-year-old women in a survey, the expected fraction of their children born at maternal age m (these children were born at $t = m - a$; survivors are $a - m$ years old) is

$$w^*(m|a) = \frac{f(m, m-a)}{\int_{12}^a f(m, m-a)\, dm} \tag{2}$$

where I assume that $f(m, t) = 0$ for all ages outside of $m \in [12, 50)$. Note that the denominator in Eq. (2) is the expected parity of a-year-old women in the survey, $P_c^*(a) = \int_{12}^a f(m, m-a)\, dm$.

The expected fraction dead among children born to a-year-old women surveyed is

$$Q^*(a) = \int_{12}^a w^*(m|a)\, q_c^*(a-m)\, dm \tag{3}$$

where this expression depends on recent trends in age-specific fertility (via $f(m, t)$ in Eq. (2)) and mortality (via $\mu(z, t)$ in Eq. (1)). The fundamental problem in indirect estimation is to learn about current period mortality (specifically $q_5^* = 1 - e^{-H_p^*(5)}$) from sample observations of births and child deaths by mother's age ($\{D_a, B_a, a = 12 \ldots 49\}$).

3 Brass Indirect Estimation

The workhorse of indirect estimation methods for q_5 is the Brass child mortality approach (Ch.3 United Nations 1983, Trussell 1975). The Brass method uses robust demographic patterns to estimate q_5 from survey reports on women's ages, parities, and child deaths.

The Brass method begins with the observation that, with continuous rates over age and time, there is an intermediate maternal age $\tilde{m}(a) \in [12, a]$ in Eq. (3) such that

$$Q^*(a) = q_c^*[\, a - \tilde{m}(a)\,] \tag{4}$$

This implies a one-to-one relationship between mother's and children's ages: in a survey, the expected fraction dead among all children of a-year-old mothers equals the probability that a child born exactly $a - \tilde{m}(a)$ years earlier would no longer be alive. In most demographic regimes $a - \tilde{m}(a)$ would be an increasing function of a, because children born to older women would have had longer exposure to mortality risks.

In an important extension, Feeney (1980) noted that if mortality changes approximately linearly over time (and fertility rates are constant) then there is a robust mapping from the cohort mortality indices in Eq. (4) to period indices at specific times in the past:

$$Q^*(a) = q_c^*[\, a - \tilde{m}(a)\,] \approx q_p[\, a - \tilde{m}(a)\,, \tilde{t}(a)\,] \tag{5}$$

where $q_p(x,t) = 1 - e^{-H_p(x,t)}$ refers to the period mortality schedule at time t. Under typical demographic circumstances the allocated time $\tilde{t}(a)$ would usually be a decreasing function of a, implying that data on the survival of older mothers' children is informative about period mortality at earlier times, longer ago.

Thus an observable quantity – the proportion dead among children of surveyed a-year-old women – provides information about past period mortality. The exact relationsips in Eqs. (4) and (5) depend on fertility and mortality patterns by age in the years before the survey, and on their changes over time.

Brass discovered that – when all vital rates are constant – regularities in age schedules[2] imply simple, robust relationships between Q^* values for women's age groups and q_c^* values for integer ages of children. In particular he calculated that with constant vital rates the expected fraction of children dead among mothers aged 30–34 (which must theoretically lie between $q^*(0) = 0$ and $q^*(23)$) would always be quite close to $q^*(5)$. Feeney (1980) further demonstrated that with unchanging fertility rates

[2] in particular, the concentration of death in very early childhood, and the concentration of fertility in a woman's 20s and early 30s

$$Q^*_{30-34} \approx q_p(5, -5) \tag{6}$$

The Brass method and its variants (Trussell 1975, United Nations 1983) fine tune the time allocation with age-specific parity information, but these adjustments are small. As a result, with approximately constant fertility rates it is possible to estimate under-five mortality in the recent past indirectly, using the proportion of children dead among the women who are currently 30–34.

4 A Bayesian Generalization of the Brass Approach

The Brass approach is ingenious. It uses consistent patterns in demographic rates to produce good estimates under a wide range of circumstances. However, it requires demographic stability in age-specific rates, it uses only one age group of women to estimate the recent under-five mortality rate, and it is a deterministic algorithm that does not generate estimates of its likely precision.

A probabilistic Bayesian model can exploit the same regularities without these restrictions. A Bayesian approach does not require fixed temporal patterns in rates, and it can use multiple age groups to estimate current q_5. It can also produce uncertainty measures that incorporate both sampling error and potential errors in demographic assumptions.

In the next subsections, I outline a Bayesian model that can improve the Brass method's accuracy and utility in modern settings where q_5 estimates are needed. This appoach uses simple parametric models to describe fertility and mortality levels and patterns over the period covering approximately 30 years before the survey, and treats the parameters in these models as unknown quantities with prior distributions.

The fundamental idea behind this estimation approach is familiar, even if the Bayesian vocabulary is not. Demographic parameters that affect vital rates are unknown, but they are not complete mysteries. Before seeing the data from any survey, we already have information about plausible demographic rates and patterns – e.g., 25 is a much more likely modal age of childbearing than 35, and a 2% annual decrease in mortality rates is much more likely than a 10% annual decrease. The standard Brass approach chooses fixed

(but demographically plausible) values for fertility and mortality parameters that affect $\frac{D_a}{B_a}$, such as a Coale-Demeny model life table family or a specific pattern of fertility by age. In contrast, a Bayesian approach uses weighted averages over a range of uncertain (but demographically plausible) scenarios rather than picking exactly one. The result is a distribution of possible q_5 values rather than a point estimate, with the most likely values those that are consistent with both survey data and plausible underlying age patterns.

4.1 Period Fertility $f^*(\,)$

I use an empirical model with four "archetype" schedule shapes $\phi_1^* \ldots \phi_4^*$, each estimated from a set of 149 schedules for middle- and low-income countries in the 2003 Census International Database (IDB). Figure 1 illustrates the four archetype shapes, $\phi_j^* \in \mathbb{R}^{38}$. Each represents a fertility pattern over integer ages $12, \ldots, 49$ such that $\phi_{jm} \geq 0$ and $\sum_m \phi_{jm}^* = 1$. These four shapes were calculated empirically (Cutler & Breiman 1994, Eugster & Leisch 2009) to minimize the total squared errors when approximating the 149 IDB schedules with convex combinations of archetypes $\hat{\phi} = \sum_{j=1}^{4} W_j \cdot \phi_j^* \in \mathbb{R}^{38}$.

A population's period fertility schedule at the time of the survey is

$$f^*(m \,|\, TFR, W_1 \ldots W_4) \;=\; TFR \cdot \sum_{j=1}^{4} W_j \cdot \phi_{jm}^* \tag{7}$$

over ages $12 \ldots 49$, where TFR is a non-negative scalar and W is a 4-simplex of non-negative weights that sum to one. In order to average over a wide range of fertility age patterns, I use a prior distribution for weights W that assigns similar probabilities to many different schedule shapes, namely

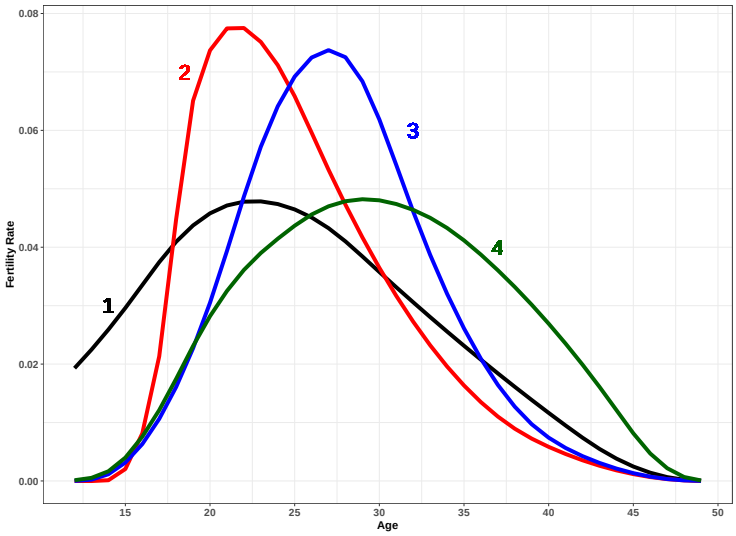
$$(W_1 \ldots W_4) \sim Dirichlet(1, 1, 1, 1)$$

The prior for TFR is

$$TFR \sim Uniform(1, 10)$$

so that all levels of total fertility in [1,10] are equally likely *a priori*.

Figure 1: Archetype age patterns of fertility. Period schedules are convex combinations of these archetypes, multiplied by TFR.



4.2 Period Mortality $\mu^*(\,)$

For the period mortality schedule on the survey date, $\mu^*(\,)$, I use the Wilmoth et al. (2012) mortality model, in which two parameters (q_5, k) determine nine piecewise-constant mortality rates for intervals starting at ages $0, 1, 5, 10, \ldots, 35$. This model is especially convenient for U5MR estimation, because q_5 enters directly as the parameter that affects the overall level of mortality.

I assume *a priori* that all q_5 values between 0 and 500 per thousand are equally plausible

$$q_5 \sim Uniform(0, .50)$$

The k parameter in the Wilmoth et al. (2012) model affects the shape of the mortality schedule, with values typically in the [-2,+2] range. k has only weak effects on predicted Q_a^* values in the model, so I use a simple prior

$$k \sim N(0, 1)$$

Empirical approximations that follow require calculating $q(x)$ values at ages 0.5, 1.5, ..., 37.5, so I further divide the $[0, 1)$ interval in this model. I assume that the mortality rate in the first six months of life is ${}_1\mu_0^* + \Delta$, and that in the second six months it is ${}_1\mu_0^* - \Delta$. Under the reasonable assumption $q(0.5) = 0.80\, q(1)$, this means that $\Delta = -2\ln(.20e^{+0.5\cdot{}_1\mu_0^*} +$

$.80e^{-0.5 \cdot {}_1\mu_0^*}$). With this additional subdivision of infant mortality, the two parameters (q_5, k) determine a set of piecewise-constant rates for ten age intervals that start at lower limits $L = 0,\ 0.5,\ 1,\ 5,\ 10, \ldots, 35$.

4.3 Fertility and mortality change

Fertility changes at rate r

I assume that age-specific fertility rates have been changing exponentially at rate r throughout the reproductive lifetimes of the women in the survey (approx. the 35 years), so that

$$f(m,t) = f^*(m) \cdot e^{rt}$$

where $f^*(m)$ is the period schedule on the survey date ($t = 0$). This is a very simple assumption, but it allows changing rates in a way that the original Brass method does not. Schmertmann et al. (2013) investigated indirect estimators with this simple model of fertility change and found it effective. If $r < 0$ then a greater fraction of births would have happened longer ago, implying longer exposures and a higher expected fraction dead at any given level of current q_5.

Under this assumption the fertility term in eq. (2) is $f(m, m-a) = f^*(m) \cdot e^{-r(a-m)}$, so that

$$w^*(m, r|a) \propto f^*(m) \cdot e^{rm} \tag{8}$$

For the rate of fertility change I use the prior

$$r \sim N(0, .025^2)$$

which assigns approximately 90% prior probability to the range between -4% and +4% annual change. Results are insensitive to the exact specification.

Mortality changes at rate s

I make a parallel assumption for mortality rates, namely

$$\mu(z,t) = \mu^*(z) \cdot e^{st}$$

where s is the annual rate of mortality change. As with fertility, if $s < 0$ and mortality is declining then this model allows for higher mortality rates in the past, which implies a higher expected fraction dead at any given level of current q_5.

For the rate of mortality change I use the same prior distribution as fertility

$$s \sim N(0, .025^2)$$

As with r, results are insensitive to the exact specification.
Under these assumptions the expected proportion dead among children born x years ago in Eq. (3) is

$$\begin{aligned} q_c^*(x,s) &= 1 - \exp\left[-\int_0^x \mu^*(z)\, e^{-s(x-z)}\, dz\right] \\ &= 1 - \exp\left[-e^{-sx} \int_0^x \mu^*(z)\, e^{sz}\, dz\right] \end{aligned} \tag{9}$$

Finally, under these assumptions about changing rates Eq. (3) for the expected proportion dead among children ever born to a-year-old women becomes

$$Q^*(a,r,s) = \int_{12}^{a} w^*(m, r|a)\, q_c^*(a-m, s)\, dm \tag{10}$$

which depends on the rates of fertility and mortality change (r and s), and on the period schedules at the time of the survey (f^* and μ^*).

4.4 Likelihood

Observed births by age $B_{20} \ldots B_{49}$[3] are informative about the level, change, and age pattern of fertility (TFR, r, and $W_1 \ldots W_4$, respectively). Conditional on births, child deaths by mother's age $D_{20} \ldots D_{49}$ are informative about the level, change, and age pattern of mortality (q_5, s, and k, respectively).

I assume that births are independent Poisson variables at each age

$$B_a \sim Poisson(W_a\, P_{ca}^*)$$

[3]I omit women under 20 from the calculations because the implicit assumption that child mortality risks are independent of mother's age are much less likely to hold for very young mothers. This is a well-known problem.

Table 1: Parameters and effects on expected fraction of children dead by maternal age

Parameters	Interpretation	Effects.on.Q.	Strength.of.Effect
$TFR \sim U[1, 10]$	Fertility Level	None	Zero
$W_1 \ldots W_4 \sim$ $Dirichlet(1, 1, 1, 1)$	Fertility Timing	later fert $\rightarrow$ lower Q^* at all ages	Weak
$q_5 \sim U[0, .500]$	Mortality Level	higher $q_5 \rightarrow$ higher Q^* at all ages	**Very Strong**
$k \sim N(0, 1)$	Mortality Timing	higher $k \rightarrow$ higher Q^* at higher maternal ages	Weak
$100r \sim N(0, 2.5^2)$	Rate of Fert Change	higher $r \rightarrow$ lower Q^* at higher maternal ages	Very Weak
$100s \sim N(0, 2.5^2)$	Rate of Mort Change	higher $s \rightarrow$ lower Q^* at higher maternal ages	**Very Strong**

and that the number of deaths conditional on births is also Poisson

$$D_a|B_a \sim Poisson(B_a\, Q_a^*)$$

Table 1 summarizes the probabilistic model parameters, and includes a brief description of the direction of change and sensitivity of Q_a^* (or equivalently, $D_a|B_a$) with respect to each parameter. Sensitivity experiments, summarized in the last column, indicate that expected fractions dead Q_a^* change very little with fertility parameters or with mortality shape parameter k. Indeed, these insensitivities were a main rationale for Brass's original approach. Expected fractions dead are very sensitive to q_5 and, importantly, to the rate of mortality change s.

```
## Loading required package: kableExtra
```

```
## Warning: package 'kableExtra' was built under R version 3.5.3
```

4.5 Markov Chain Monte Carlo (MCMC) Estimation in Stan

I programmed the model in the *Stan* MCMC language (Carpenter et al. 2017), as implemented in the *rstan* package in *R* (Stan Development Team 2016, R Core Team 2016). The empirical posterior density of (q_5, s) values is the basis for the main calculations below.

5 Some Preliminary Results

I estimated posterior distributions for $(q_5, s, k, TFR, r, \mathbf{W})$ for data from each of 137 Demographic and Health Survey datasets, downloaded from the IPUMS-DHS website (Boyle et al. 2019). Here I briefly illustrate model output and show some examples of promising (and problematic) preliminary results.

5.1 Example 1: Ethiopia 2016 DHS

5.1.1 Indirect Estimates

The Ethiopia 2016 DHS survey included detailed birth history data from 15683 women 15–49 years old, of whom 12185 were age 20+. Because full birth histories are available, it is possible to calculate period child mortality rates directly. These direct estimates, which were downloaded from the UN-IGME website (UN Inter-agency Group for Child Mortality Estimation n.d.) serve as a comparison for the indirect, Bayesian approach based only on summary data.

Estimating the model described above in Stan produces a sample of parameter values, with frequencies proportional to posterior probabilities. That is, likely values are sampled more often and vice versa. Figure 2 below illustrates draws from 4 MCMC chains for parameters TFR, q_5, r, and s.[4]

[4] The figure includes discarded "warmup" values at the beginning of each chain, indicated with a grey background.

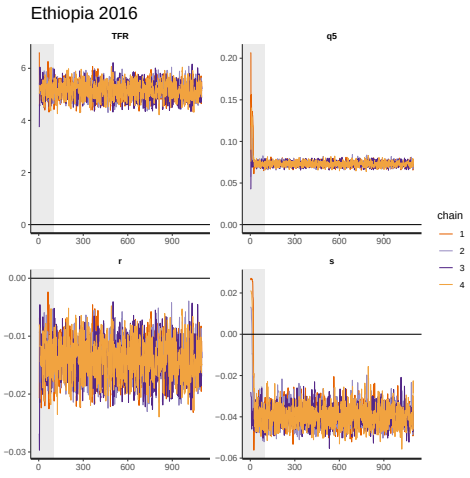


Figure 2: MCMC parameter traces, Ethiopia 2016 DHS

and Table 2 provides a summary of the marginal posterior distributions for these parameters, including the median, 10th, and 90th percentiles of the draws illustrated in Figure 2.

Table 2: Posterior Summary, Ethiopia 2016

Parameter	**Median**	Q10	Q90
TFR	**5.15**	4.73	5.57
1000 $\mathbf{q_5^*}$	**73**	69	77
r	**-0.014**	-0.018	-0.010
s	**-0.040**	-0.045	-0.033

Thus the analysis indicates that 2016 data on Ethiopian births and deaths by mother's age is consistent with a 2016 period q_5^* of 73 per 1000, and that q_5^* is in the [69,77] per 1000 range with 80% probability. Further, age patterns in parity and fertility by age provide evidence that mortality had been falling rapidly in the period over which the children represented in the sample were born: the posterior median of s is -4.0% per year, and there is an 80% probability that the rate of mortality change is in [-4.5%,-3.3%].

Figures 3 and 4 show the model fits to average parity and to proportion dead in the Ethiopia 2016 data, and Figure 5 shows the estimated childhood mortality schedule. All three figures include 80% posterior intervals as well as point estimates.

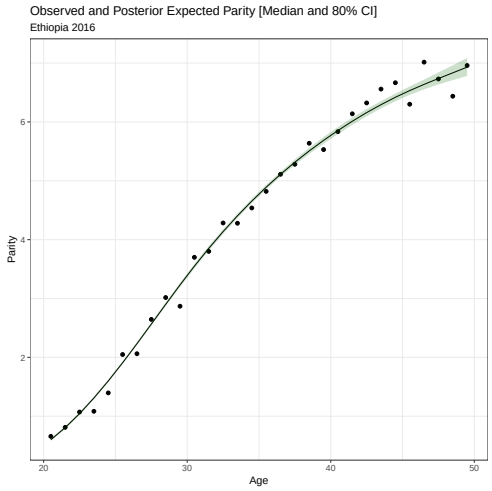


Figure 3: Average Parity by Age of Mother (Data + Fit), Ethiopia 2016 DHS

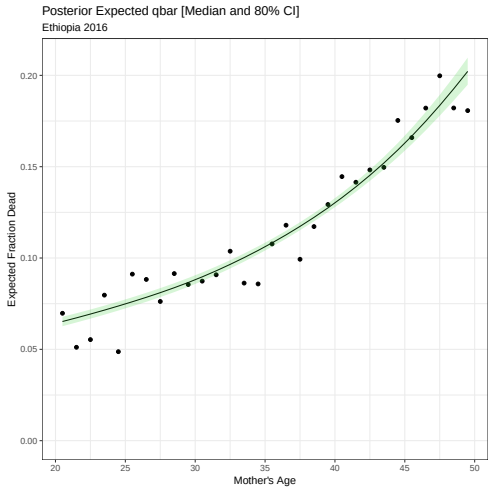


Figure 4: Average Fraction of Children Dead by Age of Mother (Data + Fit), Ethiopia 2016 DHS

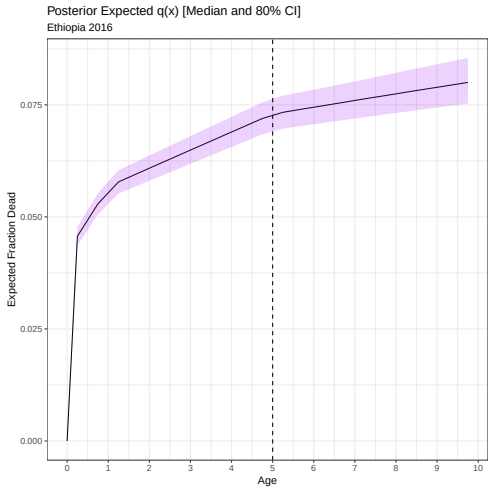


Figure 5: Posterior Estimate of q(x) function, Ethiopia 2016 DHS

5.1.2 Comparison to Direct Estimates

Because the Ethiopia 2016 collected full birth histories, we can compare the Bayesian indirect estimates to direct estimates of period mortality from event/exposure data. Figure 6 makes this comparison, showing the direct estimates reported by UN-IGME and the implied time series of period q_5 from the Bayesian model. For Bayesian estimates, sampled (q_5^*, s) pairs from the joint posterior are translated into period q_5 values at times 0,5,10, and 15 years before the average interview date; the purple band shows 80% pointwise intervals for all of the sampled time series.

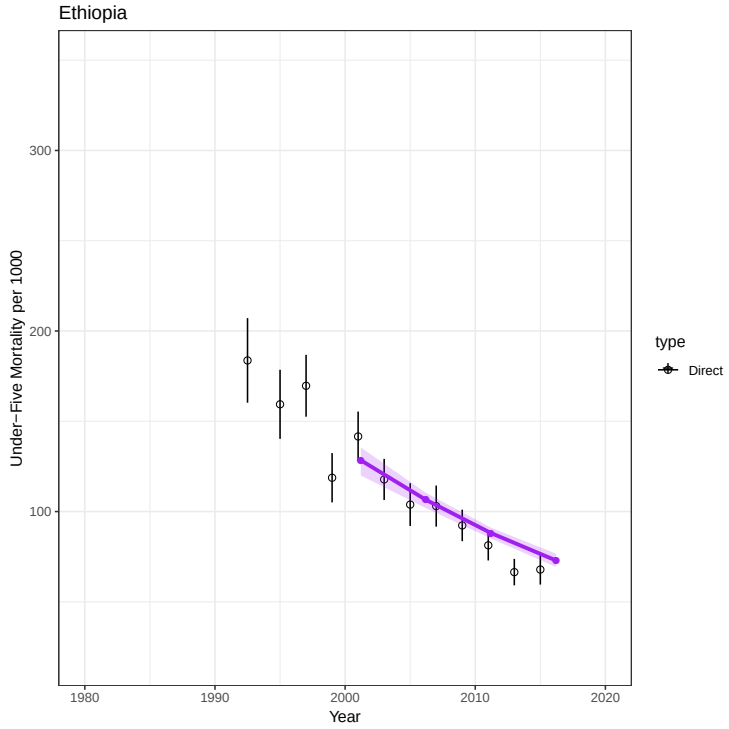


Figure 6: Direct Event-Exposure Estimates [+90% intervals = black bars] vs. Bayesian Indirect Estimates [+80% intervals = purple bands], Ethiopia 2016 DHS

The main result from Figure 6 is clear. In this case (and many others not shown) there is a close correspondence between indirect time series estimates from the Bayesian model applied to summary data, and direct estimates from detailed data. For Ethiopia 2016 data a model based on age patterns of B_a and D_a yields good estimates of recent period q_5 values, as well as the supplementary information in Figures 2, 3, 4 and 5.

5.2 Example 2: Lesotho 2009 DHS

UN-IMGE (n.d.) data includes both direct and indirect (Brass-Feeney) estimates for period q_5. Figure 7 illustrates.

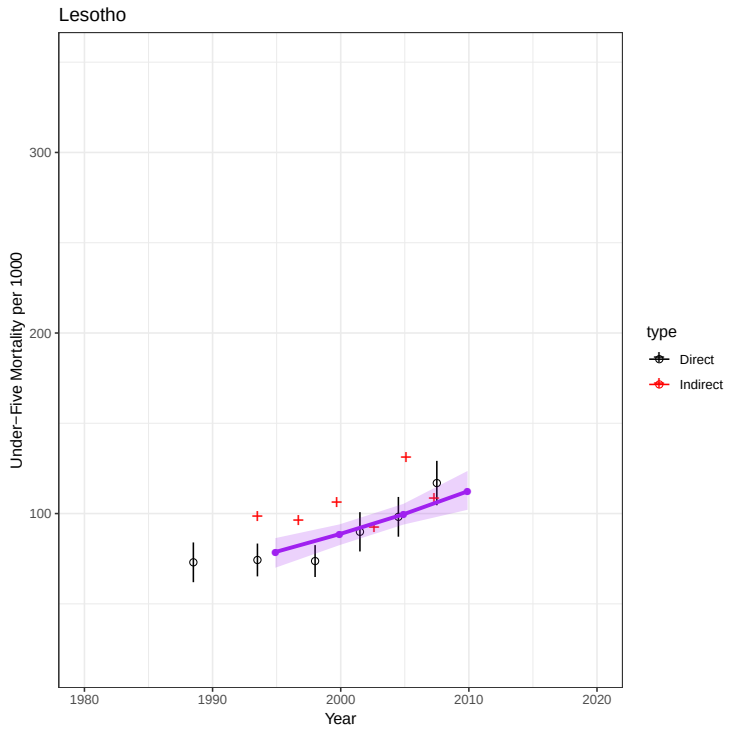


Figure 7: Direct Event-Exposure Estimates [+90% intervals = black bars] vs. Bayesian Indirect Estimates [+80% intervals = purple bands], Lesotho 2009 DHS

In the Lesotho 2009 case there is a smaller sample (5784 women 20–49), and the recent trend in mortality is upward. The Bayesian model finds a median q_5^* of 113 per 1000, with an 80% probability range of [103,124] per 1000. In contrast to the falling mortality rates in Ethiopia, for Lesotho 2009 the most plausible rate of mortality change is positive: +2.5% per year, with an 80% range of [+1.4%, +3.8%].

As seen in Figure 7, a Bayesian model for indirect estimation with flexible time trends not only matches the direct estimates well, but it also outperforms the usual Brass approach to indirect estimation.

5.3 Example 3: Niger 2006 DHS (problematic)

Figure 8 below illustrates the Bayesian model fit against direct estimates from the Niger 2006 DHS. Here the model's performance is obviously very poor.

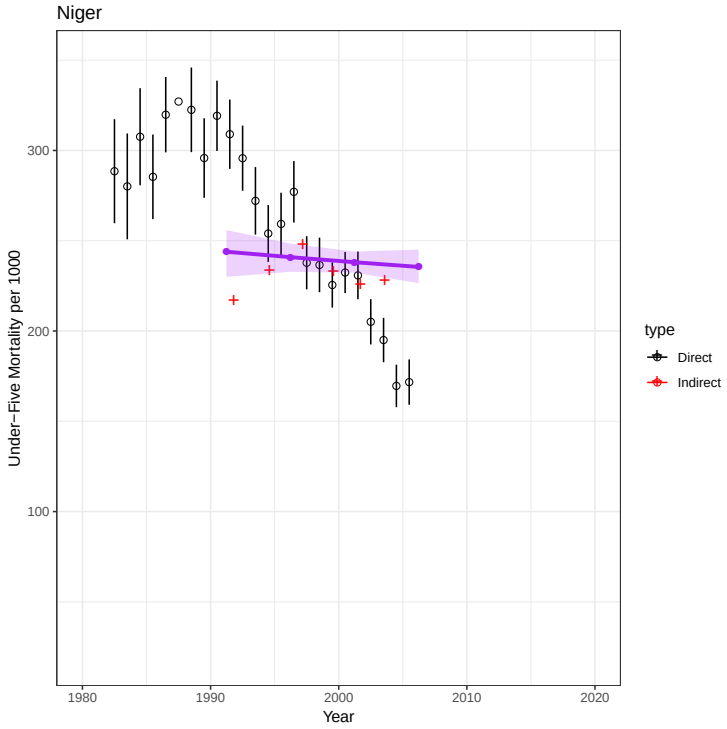


Figure 8: Direct Event-Exposure Estimates [+90% intervals = black bars] vs. Bayesian Indirect Estimates [+80% intervals = purple bands], Niger 2006 DHS

These problematic results remain a puzzle (as of this draft, Sep 2019). Although Niger is not logically different from other places covered by DHS samples, it is an outlier in terms of fertility levels (mean parity at age 49 was above 8) and mortality levels (in most maternal age groups 20-30% of children ever born had died). One clue to pursue is the similarity of Bayesian estimates to the trajectory of indirect estimates (red crosses in the Figure).

6 Discussion

A Bayesian approach to indirect mortality estimation appears very promising. It also suggests that many other indirect estimation techniques could be re-cast as probabilistic models with demographic priors derived from large data collections.

7 Between now and EPC

The agenda between Oct 2019 and the EPC meeting includes

- development of summary measures to evaluate model performance over the whole set of DHS samples
- investigation of problematic cases such as Niger 2006, in order to improve the model, or to better understand its limitations
- experiments with more flexible time series models for mortality. Expected proportions dead by age $\frac{D_a}{B_a}$ are quite sensitive to the rate of mortality change s. Given the occasional time series failure like the Niger 2006 case, this suggests that it may be valuable to model mortality time series more flexibly
- more detailed comparisons to the United Nations "B3" estimates for time series of U5MR

A Appendix: Approximations

As illustrated in Fig 9, I approximate expected fertility histories by assuming that women at integer age a (exact age $\in [a, a+1)$) are all $a + 0.5$ years old, and that their births happened when they were exact ages $m = 12, 13, \ldots a$, at rates

$$f^*(12)\, e^{-r\cdot(a-11.5)},\ f^*(13)\, e^{-r\cdot(a-10.5)},\ f^*(14)\, e^{-r\cdot(a-9.5)} \ldots f^*(a) e^{-r\cdot(0.5)}$$

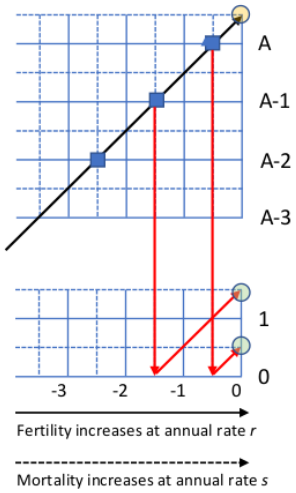
so that Eq (3) becomes

$$Q^*(a) \ \approx \ \frac{\sum_{m=12}^{a} f^*(m)\, e^{-r\,(a-m+0.5)}\, q(a-m+0.5)}{\sum_{m=12}^{a} f^*(m)\, e^{-r\,(a-m+0.5)}} \tag{11}$$

where the denominator in Eq (11) represents th expected parity of women at integer age a

$$P_a = \sum_{m=12}^{a} f^*(m)\, e^{-r\,(a-m+0.5)}$$

Figure 9: Discrete Approximation for fertility and mortality histories



B Appendix: Discretizing the mortality model's $H_c^*(x)$ terms

Cumulative mortality to the cohort born exactly age x years earlier is

$$H_c^*(x) = \int_0^x \mu^*(z)\, e^{-s(x-z)}\, dz$$

In a model like Wilmoth et al. (2012) with piecewise-constant mortality hazards $m_1 \ldots m_{10}$ over age intervals $A_1 \ldots A_{10}$, each defined by an lower and higher limit $A_j = [L_j, H_j)$ this is

$$
\begin{aligned}
H_c^*(x) &= \int_0^x \left[\sum_j I(z \in A_j) \cdot m_j \right] e^{-s(x-z)}\, dz \\
&= e^{-sx} \sum_j m_j \left[\int_0^x I(z \in A_j)\, e^{+sz} dz \right] \\
&= e^{-sx} \sum_j m_j \left[\int_{L_j}^{MIN(H_j, MAX(L_j, x))} e^{+sz} dz \right] \\
&= e^{-sx} \sum_j m_j \left[\int_{L_j}^{U_j(x)} e^{+sz} dz \right] \\
&\approx e^{-sx} \sum_j m_j \left[\int_{L_j}^{U_j(x)} 1 + sz + \frac{1}{2} s^2 z^2 dz \right] \\
&= e^{-sx} \sum_j m_j \left\{ (U_j(x) - L_j) + s \cdot \left(\frac{U_j^2(x) - L_j^2}{2} \right) + s^2 \cdot \left(\frac{U_j^3(x) - L_j^3}{6} \right) \right\}
\end{aligned}
$$

For the fixed grid of x values $x = 0.5, 1.5, \ldots, 37.5$ used by the approximation in Eq (11) the $U_j(x)$ and L_j terms are constants, so that the approximation can be simplified to

$$
H_c^*(x) \approx e^{-sx}\, m' \begin{bmatrix} Z_{x1} & Z_{x2} & Z_{x3} \end{bmatrix} \begin{bmatrix} 1 \\ s \\ s^2 \end{bmatrix}
$$

where $m \in \mathbb{R}^{10}$ is the vector of piecewise rates from the extended Wilmoth et al. model, and $Z_x = \begin{bmatrix} Z_{x1} & Z_{x2} & Z_{x3} \end{bmatrix}$ is a 10×3 matrix of pre-calculated constants.

The probability of death by any age on the grid of x values is then $q_c^*(x) = 1 - e^{-H_c^*(x)}$.

References

Alkema, L. & New, J. R. (2014), 'Global estimation of child mortality using a Bayesian B-spline Bias-reduction model', *The Annals of Applied Statistics* **8**(4), 2122–2149.
URL: *https://projecteuclid.org/euclid.aoas/1419001737*

Boyle, E. H., King, M. & Sobek, M. (2019), 'IPUMS-Demographic and Health Surveys: Version 7 [dataset]. Minnesota Population Center and ICF International'.
URL: *https://doi.org/10.18128/D080.V7*

Brass, W. & Coale, A. (1968), *Methods of analysis and estimation*, Princeton University Press, Princeton, pp. 88–150.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017), 'Stan: A probabilistic programming language', *Journal of Statistical Software* **76**(1), 1–32.
URL: *https://www.jstatsoft.org/index.php/jss/article/view/v076i01*

Cutler, A. & Breiman, L. (1994), 'Archetypal analysis', *Technometrics* **36**(4), 338–347.
URL: *http://www.jstor.org/stable/1269949*

Eugster, M. J. A. & Leisch, F. (2009), 'From Spider-Man to Hero – archetypal analysis in R', *Journal of Statistical Software* **30**(8), 1–23.
URL: *http://www.jstatsoft.org/v30/i08/*

Feeney, G. (1980), 'Estimating infant mortality trends from child survivorship data', *Population Studies* **34**(1), 109–128. PMID: 22077937.
URL: *https://www.tandfonline.com/doi/abs/10.1080/00324728.1980.10412839*

R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Rajaratnam, J. K., Marcus, J. R., Flaxman, A. D., Wang, H., Levin-Rector, A., Dwyer, L., Costa, M., Lopez, A. D. & Murray, C. J. (2010), 'Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 19702010: a systematic analysis of progress towards Millennium Development Goal 4', *The Lancet* **375**(9730), 1988–2008.

Schmertmann, C. P., Cavenaghi, S. M., Assuno, R. M. & Potter, J. E. (2013), 'Bayes plus Brass: Estimating total fertility for many small areas from sparse census data', *Population Studies* **67**(3), 255–273.
URL: *http://dx.doi.org/10.1080/00324728.2013.795602*

Stan Development Team (2016), 'RStan: the R interface to Stan'. R package version 2.14.1.
URL: *http://mc-stan.org/*

Trussell, T. J. (1975), 'A re-estimation of the multiplying factors for the brass technique for determining childhood survivorship rates', *Population Studies* **29**(1), 97–107. PMID: 22091807.
URL: *https://www.tandfonline.com/doi/abs/10.1080/00324728.1975.10410187*

UN Inter-agency Group for Child Mortality Estimation (n.d.), 'CME Info - Child Mortality Estimates'.
URL: *https://childmortality.org/data*

United Nations (1983), *Manual X. Indirect techniques for demographic estimation*, Vol. ST/ESA/SER.A/81, United Nations Publications.

Verhulst, A. (2016), 'Child mortality estimation: An assessment of summary birth history methods using microsimulation', *Demographic Research* **34**(39), 1075–1128.
URL: *https://www.demographic-research.org/volumes/vol34/39/*

Wilmoth, J., Zureick, S., Canudas-Romo, V., Inoue, M. & Sawyer, C. (2012), 'A flexible two-dimensional mortality model for use in indirect estimation', *Population Studies* **66**(1), 1–28.
URL: *http://www.tandfonline.com/doi/abs/10.1080/00324728.2011.611411*

Wilson, K. & Wakefield, J. (2018), 'Child mortality estimation incorporating summary birth history data [arxiv eprint 1810.04140]'.