

# Indirect Estimation

Our last topic is indirect estimation, a subject covered in Chapter 11 of the textbook and in much greater detail in the United Nation’s Manual X, recently updated in *Tools for Demographic Estimation*. We focus on some of the seminal contributions of Bill Brass.

## Fertility and P/F

Let  $f(a)$  denote fertility at age  $a$  and  $F(a) = \int_{15}^a f(x)dx$  cumulative fertility up to age  $a$ . If fertility has been relatively constant in the recent past one could estimate  $f(a)$  from age-specific fertility rates for the last year and  $F(a)$  from questions on children ever born by age of mother, and the two sets of estimates should be consistent.

Brass postulated, however, that these two sources of data are subject to different sources of error. Specifically, children ever born is subject to recall errors that increase with age, so  $F(a)$  may be considered reliable only for younger ages. On the other hand, reports of births in the past year are subject to time scale errors, referring to periods longer or shorter than a year; if these errors are independent of age then  $f(a)$  will have the wrong level but the right shape.

The basic idea of the procedure is to get the level from parity and the shape from fertility, hence the name P/F, usually read as “P over F”. Specifically, the method starts with mean children ever born at a young age, typically 20-24, as the estimate of P. It then accumulates age-specific fertility rates up to the same age, for example 5 times the 15-19 rate plus 2.5 times the 20-24 rate, as the estimate of F. The third step is to calculate the P/F ratio and use this to inflate the age-specific fertility rates, effectively correcting the level while preserving the shape.

In countries with good data, such as England and Wales in 1951, Brass finds P/F ratios very close to one. In Africa, however, he finds more dispersion, for example 0.8 in Guinea and 1.13 in Uganda, suggesting reference period errors.

The table on the right shows an illustrative calculation using Brass’s interpolation factors to accumulate ASFRs (for example for the age group 20-24 we use 2.695 instead of 2.5). The P/F ratio at 20-24 suggests that fertility is about 33% *higher* than reported, and leads to a revised TFR of 5.1 instead of 3.9.

Age	f	k	F	P	P/F	f*
15-19	0.021	1.345	0.028	0.038	1.345	0.028
20-24	0.170	2.695	0.563	0.747	1.326	0.225
25-29	0.195	2.865	1.514	1.892	1.250	0.259
30-34	0.172	3.085	2.461	2.884	1.172	0.228
35-39	0.124	3.200	3.187	3.560	1.117	0.164
40-44	0.067	3.405	3.638	3.868	1.063	0.089
45-49	0.022	4.020	3.833	3.868	1.009	0.029
<b>TFR</b>	<b>3.855</b>		<b>3.855</b>			<b>5.114</b>

There are better interpolation factors to accumulate fertility up to the middle of an interval, depending on the ratio of the rates at 20-24 and 15-19. A variant uses P/F ratios for first births to derive a correction factor. An adjustment is also needed when fertility rates are based on births last year by current age of woman, rather than true event-exposure rates. Models may play a role here. Manual X uses the Coale-Trussell model, and *Tools for Demographic Estimation* emphasizes the use of relational Gompertz models, and has a worked example with average parity and period fertility rates from the Malawi 2008 Census.

Schmertmann and collaborators have a nice 2013 paper in *Population Studies* on “Bayes plus Brass” to estimate total fertility for many small areas using sparse census data, with an application to 2000 Brazilian Census data for over five thousand municipalities. Their algorithm first uses Bayesian techniques to smooth local age-specific rates, and then applies a variant of Brass’s P/F method that is robust under conditions of rapid fertility decline.

### Child Mortality from Reports of Children Surviving

Brass proposed a method for estimating child mortality from mother’s reports of children ever born and children surviving, that quickly became the main source of child mortality estimates in the developing world. The basic idea is to ask a mother how many children she has given birth to, and how many are still alive. These questions have been added in many censuses, and are usually known as “the Brass questions”.

Let  $f(a)$  denote fertility at age  $x$  and  $p(a)$  denote the probability of surviving from birth to age  $a$ . If a mother is now age  $a$ , she is expected to have had  $F(a) = \int_{15}^a f(x)dx$  children. A child born when the mother was age  $x$  was born  $a - x$  years ago and has a probability  $p(a - x)$  of being alive today. The expected number of children surviving is then

$$S(a) = \int_{15}^a f(x)p(a - x)dx.$$

By the mean value theorem, we should be able to approximate the last integral by

$$S(a) = p(a - x^*) \int_{15}^a f(x)dx$$

where  $x^*$  is some value between 0 and  $a$ . Under these assumptions, the ratio of children surviving to children ever born is

$$\frac{S(a)}{F(a)} = p(a - x^*)$$

and is a direct estimate of the life table probability of surviving to some age  $a - x^*$ , that depends on the age of the mother and the average time since her children were born. (Compare this with equation 11.3 in the text, noting that I used  $a$  for the age of the mother in what I think is a simpler explanation.)

Obviously young women must have had their children relatively recently. Brass noted that women aged 15-19 had their children on average a year ago, so the ratio estimates  $p(1)$ , the probability of surviving to age one. For women 20-24 it estimates  $p(2)$ , and for women 30-34 it estimates  $p(5)$ , see the table on page 228 of the textbook for more details. These values, however, need to be adjusted depending on the age pattern of fertility.

Brass developed a set of correction factors using simulation. These were later revised by Sullivan and then by Trussell. The factors are in the form of regression coefficients that take as inputs the ratios of mean parities 15-19/20-24 and 20-24/25-29. The result is a correction factor that is multiplied by the proportion dead to yield an estimate of the appropriate  ${}_nq_0$  for each age of mother. (See Table 11.1 in the textbook. Note that  $b_i$  should be  $-.5381$  at age 20-24.)

A second problem is that the ratio of children surviving to children ever born depends on mortality conditions in the past. If mortality has not been constant, then estimates for older ages refer to periods further in the past.

Coale and Trussell developed formulas for estimating the period to which a set of estimates refer, based on an assumption of linearly declining mortality. These are also in the form of regression coefficients that take as input the ratio of mean CEB at ages 15-19/20-24 and 20-24/25-29. The result is an estimate of the period to which the estimates apply. (See Table 11.2 in the textbook.)

The table below shows calculations for the data from Zimbabwe in 1994 found in Box 11.1 in the textbook. The basic inputs are the mean parities and the proportions of children dead by age of mother.

Age	Parity	Prop dead	Child age	Mort adj	q(x)	Ref period
15-19	0.170	0.0560	1	1.080	0.0605	1.0
20-24	1.100	0.0817	2	1.050	0.0858	2.3
25-29	2.360	0.0760	3	1.001	0.0760	4.2
30-34	3.890	0.0847	5	1.009	0.0855	6.5
35-39	5.130	0.0935	10	1.027	0.0960	9.0

The relevant ratios of mean parities are  $0.17/1.1=0.155$  and  $1.1/2.365=0.465$ . Using these as inputs in the regression equation for adjusting the data for women aged 15-19 we get a correction factor of 1.08. Multiplying this by the proportion dead for women 15-19 we get an estimate of  ${}_1q_0 = 0.0605$ . Plugging the same two ratios in the regression equation for the reference period we get 1.0, so we estimate that the probability of infant death was about 60 per thousand, approximately one year before the survey. Calculations for the other age groups proceed along the same lines. The age group 30-34 leads to an estimate of  ${}_5q_0 = 0.0855$ , so under five mortality was about 86 per thousand, and we time this estimate around 6.5 years before the survey.

*Tools for Demographic Estimation* uses data on children ever born and children surviving from the 2008 Census of Malawi to illustrate the method, and relies on relational logits to convert estimates to  ${}_5q_0$  at various times in the past.

## Adult Mortality from Data on Orphanhood

Essentially the same logic used to estimate child mortality can be used to estimate adult mortality from reports of orphanhood. Let  $B(t)$  denote the birth density at time  $t$  and  $p(a)$  the probability of surviving to age  $a$ . The number of people age  $a$  at time  $t$  is

$$N(a, t) = B(t - a)p(a),$$

the number of births  $t - a$  years ago times the probability of surviving  $a$  years.

The probability that a person age  $a$  at time  $t$  will not be a maternal orphan is  $p_M(a)$ , the probability that a woman would survive  $a$  years after giving birth. The density of people age  $a$  at time  $t$  whose mother is alive is then

$$NO(a, t) = B(t - a)p(a)p_M(a)$$

and the ratio of non-orphans to the total population age  $a$  at time  $t$  is

$$\frac{NO(a, t)}{N(a, t)} = p_M(a)$$

a direct estimate of the probability that a mother would survive  $a$  years after giving birth.

The next question is how to relate this ratio to life table survival probabilities. The answer depends on the average age of mothers given the age of their offspring. To a first approximation  $p_M(a)$  is the probability of surviving  $a$  years from the mean age of childbearing  $M^*$  to age  $M^* + a$ , or  $l_{M^*+a}/l_{M^*}$ . If mean age of childbearing was 27.5, then the proportion non-orphan among respondents 15-19 would estimate the probability of surviving from age 27.5 to 45 (or 27.5+17.5).

Just as was done for the children surviving method, Hill and Trussell developed a set of regression equations for converting the proportions non-orphaned by age into survival probabilities using simulation, based on model schedules of fertility and mortality. The equations take as input the mean age of mothers at childbirth, and the proportion of people in the age group whose mothers are alive. The survival ratios estimated range from  $l_{45}/l_{25}$  for the age group 15-19 to  $l_{60}/l_{25}$  for 30-34.

Unfortunately, the method depends on the assumption of constant mortality and there is no reliable procedure to date the estimates if mortality has been declining. A potential source of bias is selectivity, induced by the fact that only surviving children can report their orphanhood status. For younger respondents there may also be an "adoption effect", where children report the survivorship of their adopted rather than their biological mother. Alternative methods rely on the survival of siblings or spouses, but they tend to be less accurate for adult mortality.

*Tools for Demographic Estimation* has an application of the orphanhood method to data from Iraq. They also have an extensive discussion of the impact of the HIV/AIDS on mortality estimation, with an illustration from Kenya.

## The Sisterhood Estimate of Maternal Mortality

The last indirect method we will mention estimates maternal mortality from reports of sisters. The following brief description borrows heavily from a note I wrote with Trussell. The textbook describes the procedure in more detail in section 11.3 and has an example using data from Gambia in 1987.

The basic idea of the method is to take a sample of women and ask how many sisters have ever married and, of these, how many (if any) have died during pregnancy, childbirth or puerperium. In populations where sexual relations outside marriage are common, or where marriage itself is not well defined, inquiries can be made about sisters past menarche or past age 15; the basic idea remains the same.

If the sample consists of women aged 60 or over, the simple fraction of sisters who died of maternal causes turns out to be an estimator of the lifetime risk of maternal mortality in the presence of other causes of death. If the sample includes women under 60, however, some of their sisters are still at risk of maternal mortality, so an inflation factor must be used to convert the fraction dead to a lifetime risk. Appropriate adjustment factors have been computed using standard fertility and mortality schedules. Estimates typically refer to mortality conditions about 12 years before the survey.

A key assumption of the method is independence between the number of siblings and their survival probabilities, as well as independence of the mortality experiences of adult sisters. An interesting feature of the method is the fact that the sampling frame appears to count the experience of some women multiple times. In fact, an early DHS survey restricted the sampling frame so only one sister per household was allowed to answer the maternal mortality note. It turns out, however, that restricting the sampling frame introduces biases; multiple reporting is not only simpler, by not requiring linking sisters who may live in different households, but essential for the success of the technique.

*Tools for Demographic Estimation* has a worked example using data from the Malawi 2004 DHS to estimate pregnancy-related mortality. They note that sampling uncertainty is very large compared to estimates of under-5 mortality, so while estimates of levels may be useful, interpretation of differentials is hazardous, and any conclusions about trends should be based on estimates from two or more surveys.