



**USAID**  
FROM THE AMERICAN PEOPLE

# DHS WORKING PAPERS

## **New Multivariate Methodology for Assessing the Effect of Child Mortality on the Total Fertility Rate and its Components**

Hassan Eini-Zinab

2010 No. 70

September 2010

This document was produced for review by the United States Agency for International Development.

*DEMOGRAPHIC  
AND  
HEALTH  
RESEARCH*

The *DHS Working Papers* series is a prepublication series of papers reporting on research in progress that is based on Demographic and Health Surveys (DHS) data. This research is carried out with support provided by the United States Agency for International Development (USAID) through the MEASURE DHS project (#GPO-C-00-08-00008-00). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

MEASURE DHS assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. Additional information about the MEASURE DHS project can be obtained by contacting MEASURE DHS, ICF Macro, 11785 Beltsville Drive, Suite 300, Calverton, MD 20705 (telephone: 301-572-0200; fax: 301-572-0999; e-mail: [reports@measuredhs.com](mailto:reports@measuredhs.com); internet: [www.measuredhs.com](http://www.measuredhs.com)).

**New Multivariate Methodology for Assessing the Effect of Child Mortality  
on the Total Fertility Rate and Its Components**

Hassan Eini-Zinab

ICF Macro  
Calverton, Maryland, USA

September 2010

*Corresponding author:* Hassan Eini-Zinab, Population and Health Studies, Research Program, East-West Center,  
1601 East West Rd. Honolulu, HI 96848, USA; E-mail: [eini@hawaii.edu](mailto:eini@hawaii.edu) and [h\\_eini\\_z@yahoo.com](mailto:h_eini_z@yahoo.com)

**Suggested citation:**

Eini-Zinab, Hassan. 2010. *New Multivariate Methodology for Assessing the Effect of Child Mortality on the Total Fertility Rate and Its Components*. DHS Working Papers No. 70. Calverton, Maryland, USA: ICF Macro.

## **ABSTRACT**

This paper develops and validates new multivariate methodology to assess the effect of child mortality on period and cohort fertility. Methodology is based on discrete-time survival models that are used to construct multivariate life tables of parity progression. Parity transitions are transitions from the woman's own birth to her first birth (0-1), from first to second birth (1-2), from second to third birth (2-3), and so on. The basic dimensions of these life tables are age, parity, duration in parity, and child mortality. By multivariate is meant that a life table can be constructed by values or categories of one socioeconomic predictor variable while holding the other socioeconomic predictors constant (socioeconomic predictors may include residence, education, wealth index, religion, etc). These multivariate life tables yield various measures of the quantum and tempo of fertility and child mortality. The methodology also yields a replacement rate, which measures the extent to which parents replace lost children.

The validity of the methodology is tested by applying it to India's first National Family Health Survey, conducted in 1992/93 (NFHS-1). The application is to both cohort and period data. The cohort analysis pertains to all women age 45-49 at time of survey. The period analysis pertains to the 5-year period preceding the survey, spanning years 1988 to 1992. Tests indicate that basic results agree closely with results from the birth history method and other methods, insofar as comparisons can be made. The replacement rate in the cohort analysis is 0.32 which means about one-third of lost children by cohort of women age 45-49 were replaced by additional births. The replacement rate in the period analysis is 0.50.



## INTRODUCTION

Griffith Feeney (1983, 1986) developed a methodology that uses life tables of parity progression to calculate parity progression ratios (PPRs) and a total fertility rate (TFR) from the PPRs using either period or cohort birth history data. Feeney's work inspired the development of a multivariate version of his methodology by Retherford et al., based on multivariate life tables of parity progression (Retherford et al., 2009, 2010a, 2010b). The term "multivariate" means that a life table can be constructed by values or categories of socioeconomic predictors of interest, such as residence, education, wealth index, religion, etc. The advantage of this multivariate methodology compared with Feeney's original is that it yields fertility measure by categories or values of predictor variables, while holding other covariates constant.

The multivariate methodology approaches the calculation of parity progression ratios from the perspective and with the tools of discrete-time event history analysis (Retherford et al., 2009, 2010a). The event of interest is the transition from parity  $i$  to parity  $i+1$  ( $i=0,1, 2\dots$ ) as a function of the duration  $t$  in parity  $i$  and a number of socioeconomic predictors of interest. This multivariate methodology is referred to as the  $P_{it}$  method, where  $i$  denotes parity and  $t$  denotes duration in parity  $i$ . If there are no socioeconomic predictors, the transition probabilities  $P_{it}$  estimated by the multivariate methodology coincide with the standard ones calculated using Feeney's approach.

Adding the woman's age at starting parity  $i$  to the  $P_{it}$  method allows construction of multi-dimensional (or "global") life tables where the basic dimensions are age, parity, and duration in parity (Retherford et al., 2010b). This extended methodology is referred to as the  $P_{ait}$  method, where  $a$  denotes a woman's age, and  $t$  denotes duration in parity  $i$ . The multivariate life tables derived applying the  $P_{ait}$  method yield various measures of the quantum and tempo of fertility. Measures of the quantum of fertility are parity progression ratios (PPRs), age-specific fertility rates (ASFRs), and the total fertility rate (TFR). Measures of the tempo or timing of fertility are mean and median ages at childbearing (both all births and by birth order) and mean and median closed birth intervals. For ease of exposition, these related measures are collectively referred to as "components of the TFR".

This paper builds on the  $P_{ait}$  method to incorporate child mortality. Specifically, we develop and validate a new multivariate methodology to assess the effect of child mortality on fertility. To do so, for each parity  $i$ , we take into account of the number of child deaths (regardless of child's age at death) that occurred to women age  $a$  at each duration  $t$ . This methodology is labeled as the  $P_{aitmn}$ , where  $a$  denotes age,  $i$  denotes parity,  $t$  denotes duration in parity,  $m$  denotes child mortality state (number of child deaths, lagged by one year) at the start of a one-year age period, and  $n$  denotes child mortality state at the end of that period.

In addition to the quantum and tempo measures of fertility mentioned above, this new methodology yields measures of the quantum and tempo of child mortality. Mean number of child deaths per woman and child mortality rates by age of woman (not by age of child, who can be of any age at death) are the measures of the quantum of child mortality, and mean age of women at child death is a measure of the tempo or timing of child mortality. The methodology also yields a replacement rate, which is based on the effect of child mortality on the TFR. The replacement rate measures the extent to which parents replace lost children (regardless of child's age at death) by having more births. The new methodology is validated using data from India's first National Family Health Survey (NFHS-1), conducted in 1992/93.

## THE $P_{aitmn}$ METHOD

### The Global Life Table

The new methodology that we develop in this paper builds on the  $P_{it}$  and  $P_{ait}$  methods. The details of these methods are presented elsewhere (Retherford et al., 2009, 2010a, 2010b)<sup>1</sup>, and we summarize here only their key elements.

$P_{it}$  and  $P_{ait}$  refer to multivariate models of parity progression for women of reproductive age. The event of interest in both cases is the transition from parity  $i$  to parity  $i+1$  ( $i=0,1, 2\dots$ ), and the key outcomes to be estimated are the probabilities of transition from parity  $i$  to parity  $i+1$ .<sup>2</sup> These probabilities are estimated using discrete-time survival models. Specifically, the complementary log-log (CLL) model<sup>3</sup> is used to model parity progression (i.e. the probability of transition) from a woman's own birth to her first birth (0-1), from first birth to second birth (1-2), from second birth to third birth (2-3), and so on<sup>4</sup>. For each parity transition, duration in parity ( $t$ ) is the basic predictor variable in the CLL model. The set of predictor variables also includes additional characteristics of interest such as residence, education, wealth index, religion, etc.

In the  $P_{it}$  method, each CLL model for a particular parity transition yields estimates of the transition probability  $P_{it}$  (Retherford et al., 2009, 2010a). These transition probabilities can be used to construct a life table for each parity, from which a PPR, mean and median age at first birth, and mean and median closed birth intervals are calculated. The PPRs of different parity transitions are combined to estimate the TFR.

Adding the woman's age to the underlying CLL models in the  $P_{it}$  method allows construction of multi-dimensional life tables where the basic dimensions are woman's age, parity, and duration in parity. This extended methodology is labeled as the  $P_{ait}$  method, where  $a$  denotes the woman's age,  $i$  denotes parity, and  $t$  denotes duration in parity (Retherford et al.,

---

<sup>1</sup> The procedure for data construction of this method is similar to the previous methods.

<sup>2</sup> It is important to clarify that  $P_{it}$  and  $P_{ait}$  represent the names of the methodologies as well as their outcomes (the transition probabilities between parities  $i$  and  $i+1$  as a function of duration in parity  $t$  and woman's age  $a$ ).

<sup>3</sup> For explanation of why complementary log-log is used instead of other survival models, see Retherford et al., 2009, 2010a).

<sup>4</sup> The 0-1 transition can be disaggregated into separate parity transitions from woman's birth to her first marriage (B-M) and from first marriage to first birth (M-1), if so desired (see Retherford et al., 2009, 2010a). This research uses a combined 0-1 transition, since child mortality becomes a determinant of fertility only after a woman has at least one birth, so that marriage is largely irrelevant.

2010b). A separate model is run for each parity  $i$  to estimate the probabilities of transition  $P_{ait}$ . Collectively, the models for the various parity transitions yield model-predicted estimates of probabilities of failure (i.e., first birth or next birth) by age, parity, and duration in parity, and by socioeconomic characteristics. The  $P_{ait}$  are conditional probabilities of failure between  $a$  and  $a+1$ ; i.e., they are conditional on “survival” to age  $a$ , parity  $i$ , and duration  $t$ .

The  $P_{ait}$  for a particular set of values of the socioeconomic characteristics allow construction of a “global life table.” The term “global” means that the life table associated to the estimated  $P_{ait}$  spans all parity transitions, rather than just a single parity transition. From the global life table one can calculate not only a TFR but also all of the various components of the TFR mentioned above. Because the  $P_{ait}$  are multivariate, the global life table is itself multivariate, as are all of the measures derived from it.

### **Child Mortality as an Additional Dimension of the Global Life Table**

The  $P_{aitmn}$  method that we develop in this paper adds a new dimension (child mortality) to the global life tables constructed with the  $P_{ait}$  method. In the new methodology,  $a$  denotes the woman’s age,  $i$  denotes parity,  $t$  denotes duration in parity,  $m$  denotes child mortality state at the start of a one-year age interval, and  $n$  denotes child mortality state at the end of that interval. Child mortality state corresponds to the number of child deaths between the beginning and the end of a one-year interval of the woman’s age, irrespective of the child’s age at death. Both  $m$  and  $n$  are lagged by one year, to ensure that causation is from child mortality to fertility, not the other way around.<sup>5</sup> In other words,  $m$  and  $n$ , though denoted at the woman’s age  $a$ , actually pertain to the number of child deaths between the beginning and end of the  $(a-1)^{\text{th}}$  age interval. Child mortality state is represented by two subscripts in the  $P_{aitmn}$  methodology because, as will be explained in the next section, women are partitioned into two groups over each one-year age interval. One group consists of women who do not experience a child death over the interval (in which case  $n = m$ ), and the other group consists of women who do experience a child death over the interval (in which case  $n = m + 1$ ). In the case of women who experience more than one child

---

<sup>5</sup> A more complete way to control for reverse causation in this research is running separate models for each parity transition. Because each parity transition has its own CLL model, the causation is always from child mortality to fertility, not the other way around. For example, if a fifth child dies, parity transitions 0-1, 1-2, 3-4, and 4-5 will not be affected. But parity transitions 5-6 and higher could be affected after one year of the child’s death.

death during the year (a relatively rare event), the additional child deaths are spread forward in time, so that no more than one child death occurs in any given year.

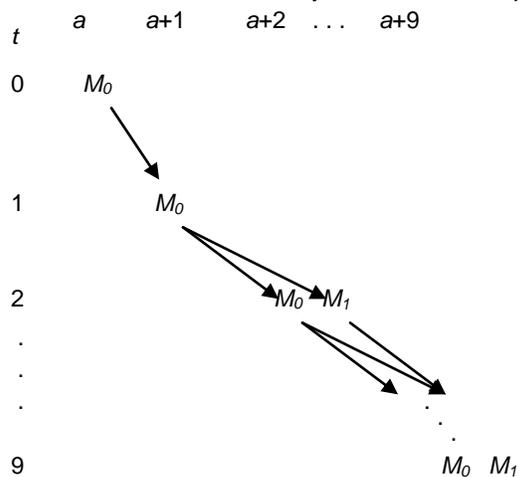
To illustrate the mechanisms of the new  $P_{aitmn}$  methodology, Figure 2.1 depicts transitions of women along the new global life table's multiple dimensions, for each of the first three transitions 1-2, 2-3, and 3-4 starting from any particular age. In the figure, duration in parity  $t$  starts at 0.<sup>6</sup>  $M_m$  denotes child mortality state (irrespective of child's age at death) for a woman age  $a$  and duration in parity  $t$ . The subscript  $m$  refers to the number of dead children at the start of the previous age interval ( $a-1$ ).

---

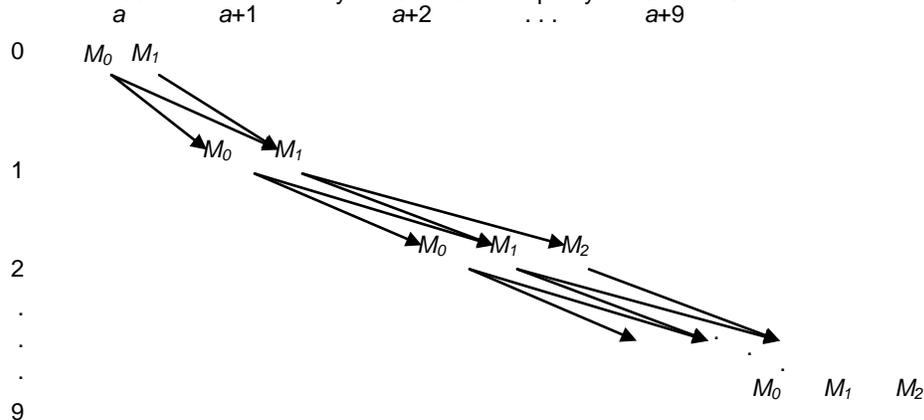
<sup>6</sup> In a CLL model for a particular parity transition,  $t$  starts at 1, which is later translated to zero when computing the global life tables.

**Figure 2.1: Possible child mortality transitions over duration in parity during parity transitions 1-2, 2-3, and 3-4**

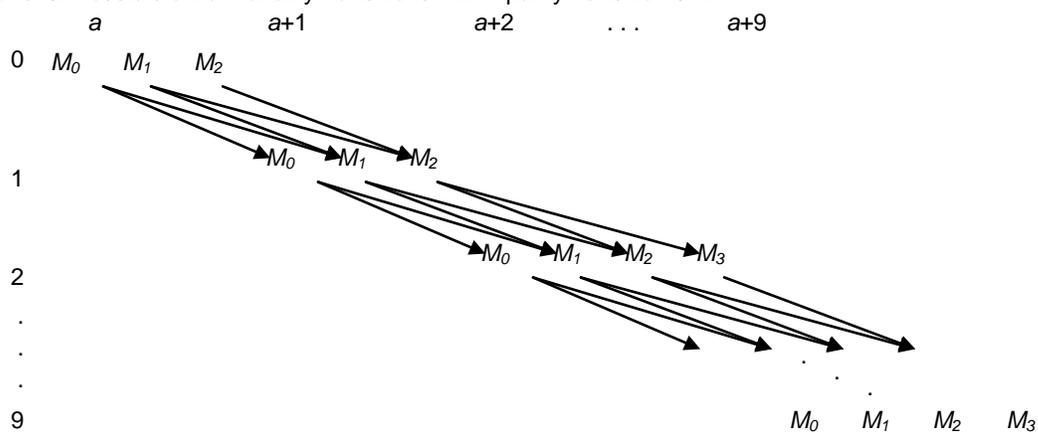
Panel 1: Possible child mortality transitions within parity transition 1-2:



Panel 2: Possible child mortality transitions within parity transition 2-3:



Panel 3: Possible child mortality transitions within parity transition 3-4:



Note:  $M_0, M_1, M_2, \dots$  denote one-year-lagged child mortality states<sup>7</sup>.

<sup>7</sup> Subscript  $m$  in  $M_m$  denotes woman's number of previous child deaths at the start of the previous age interval. Child deaths pertain to children of any age. The logic of the figure requires that a woman cannot experience more than one

Child mortality does not come into play in the 0-1 transition, because during this transition there are as yet no children born. Thus, Panel 1 of Figure 2.1 shows child mortality transitions from one age interval to the next within the 1-2 parity transition. Consider a woman who starts in state  $M_0$  at age  $a$  and duration 0 (i.e.,  $t = 0$ ). At age  $a+1$  and duration 1 (i.e.,  $t = 1$ ), there is still only one child mortality state,  $M_0$ , because child mortality state is lagged by one year of age. A woman in the  $M_0$  state at age  $a+1$  and duration 1 can continue in that state until age  $a+2$  and duration 2 if she does not experience a child death between age  $a$  and duration 0 and age  $a+1$  and duration 1 (i.e., between exact age  $a$  and exact duration 0 up to but not including exact age  $a+1$  and duration 1). Alternatively, the woman can experience a child death between age  $a$  and duration 0 and age  $a+1$  and duration 1, in which case she moves from  $M_0$  to  $M_1$ . Once a woman attains child mortality state  $M_1$ , she remains in that state until one year after she has a second birth (in which case she is no longer in parity transition 1-2) or reaches ten years of duration in parity without having yet had a second birth (which is the situation depicted in the figure). The logic of Panels 2 and 3 is similar, the difference being that the number of possible child mortality transitions increases as parity increases.

In the next section, we illustrate how we use the CLL model to estimate the transition probabilities  $P_{aitmn}$  and the associated global life tables that take child mortality into account.

### **Incorporating Child Mortality into the CLL Models**

As in the earlier methodologies, the new probabilities  $P_{aitmn}$  are also estimated by a set of CLL models, one for each parity transition. The basic predictor variables in these CLL models are similar to those in the  $P_{ait}$  method. They are woman's age  $A$  at starting parity  $i$  (where  $A=a+t$ ) and woman's duration in parity  $i$  (denoted by  $t$  or a set of dummy variables  $T_1, T_2\dots$ ). The difference is that, for the 1-2 and higher-order parity transitions, the  $P_{aitmn}$  model also includes the predictor variables  $D_1$  (number of previous child deaths at the start of the  $(a-1)^{\text{th}}$  age interval) and  $D_2$  (number of previous child deaths at the end of the  $(a-1)^{\text{th}}$  age interval), representing the values of  $m$  and  $n$  at age  $a$ , parity  $i$ , and duration  $t$ . The variables  $D_1$  and  $D_2$  are defined in the

---

child death in any given year. But some actually do. This problem is handled in the following way: If, for example, a woman had two child deaths in a year (a rare occurrence), one of the two deaths is randomly moved to the following year in the woman's birth history, before the expanded data set is constructed. If a woman had three child deaths in the same year, two of them, randomly selected, are moved to the following two years, with one child death assigned to each of those two years.

same way as the subscripts  $m$  and  $n$  in  $P_{aitmn}$ ; i.e.,  $D_1 = m$  and  $D_2 = n$ , where  $m$  and  $n$  can take on different values at different values of  $a$ . The more elaborate notation  $D_1$  and  $D_2$  is used for variable names in the model equations, in order to make clear that the number pair  $(m, n)$  is variable from one age interval to the next and also to conform to the convention of using upper-case letters to represent variable names. Additional predictor variables are socioeconomic characteristics of interest.

In the CLL models,  $D_1$  and  $D_2$  are also interacted with  $t$  and  $t^2$ . This is because it is assumed that the effect of a child death on the probability of having a next birth is nonlinear, and that a quadratic specification of duration  $t$  adequately captures the nonlinearity. Both  $D_1$  and  $D_2$  are treated as continuous variables, the ranges of which increase as parity increases. But it is still the case that the difference between  $D_1$  and  $D_2$  is either zero or one.

The new CLL model for parity transitions higher than 0-1, with place of residence and education as socioeconomic predictors of interest, is:

$$\begin{aligned}
 P = 1 - \exp\{-\exp[a + b_1T_1 + b_2T_2 + \dots + b_9T_9 + A(c_1+c_2t+c_3t^2) + A^2(d_1+d_2t+d_3t^2) + \\
 D_1(e_1+e_2t+e_3t^2) + D_2(f_1+f_2t+f_3t^2) + U(g_1+ g_2t+ g_3t^2) + L(j_1+ j_2t+ j_3t^2) + \\
 H(k_1+ k_2t+ k_3t^2) + mU D_1 + nL D_1 + oH D_1]\} \quad (2.1)
 \end{aligned}$$

where  $A$  denotes the woman's age at the start of the parity transition;  $U$  denotes urban/rural residence; and  $L$  and  $H$  denote low and high education (with illiterate as the reference category). Only  $D_1$  is interacted with the socioeconomic variables, because  $D_1$  and socioeconomic predictors pertain to the same age interval. Values of  $P_{aitmn}$  are calculated from this equation in essentially the same way that values of  $P_{it}$  and  $P_{ait}$  were calculated in Retherford et al., 2009, 2010a, 2010b.

At higher parities, where sample sizes get smaller and non-convergence may occur, a quadratic specification of duration is used in the CLL model equation:

$$\begin{aligned}
P = I - \exp\{-\exp[a + b_1t + b_2t^2 + A(c_1+c_2t+c_3t^2) + A^2(d_1+d_2t+d_3t^2) + \\
D_1(e_1+e_2t+e_3t^2) + D_2(f_1+f_2t+f_3t^2) + U(g_1+ g_2t+ g_3t^2) + L(j_1+ j_2t+ j_3t^2) + \\
H(k_1+ k_2t+ k_3t^2) + mU D_1 + nL D_1 + oH D_1]\} \quad (2.2)
\end{aligned}$$

Since a woman's age increases by parity, person-year contributions from the lower-order parity transitions in an open-ended parity transition are smaller than those from higher-order parity transitions. Person-year observations contributed from lower-order parity transitions to the open-ended parity transition are also greater than person-years contributed from higher-order parity transitions, because not all women continue childbearing to higher-order parities. Therefore, age of women at the start of parity is not included in the CLL model for open-ended parity transition. However, the global life table itself will include age for open-ended parities (Retherford et al., 2010b). The CLL model for open-ended parity transition is:

$$\begin{aligned}
P = I - \exp\{-\exp[a + b_1t + b_2t^2 + D_1(c_1+c_2t+c_3t^2) + D_2(e_1+e_2t+e_3t^2) + \\
U(f_1+ f_2t+ f_3t^2) + L(g_1+ g_2t+ g_3t^2) + H(j_1+ j_2t+ j_3t^2) + kU D_1 + \\
mL D_1 + nH D_1]\} \quad (2.3)
\end{aligned}$$

When calculating the global life table, it is assumed in the case of open parity intervals that  $P_{aitmn} = P_{itmn}$  for ages within the open parity interval, where  $i$  denotes the open parity interval (e.g., 6+) and  $P_{itmn}$  is calculated from equation (2.3).

Collectively, the CLL models for the various parity transitions yield model-predicted failure (i.e., birth) probabilities  $P_{aitmn}$  for specified values (usually representing categories) of the socioeconomic characteristics. The  $P_{aitmn}$  are conditional probabilities of failure (next birth) between age  $a$  and  $a+1$ ; i.e., they are conditional on "survival" to age  $a$ , parity  $i$ , duration  $t$ , and child mortality states  $m$  at age  $a$  and  $n$  at age  $a+1$ . (Because child mortality variables are lagged by one year, the actual ages are  $a-1$  and  $a$ .) The  $P_{aitmn}$  for a particular set of values of the socioeconomic characteristics allow calculation of a global life table (illustrated in the next section), from which one can calculate not only TFR and its components, but also child mortality measures and a replacement rate indicating the extent to which parents replace children who have died.

## Constructing the Global Life Table from the Birth Probabilities $P_{aitmn}$

The predicted estimates of  $P_{aitmn}$  from the CLL model are the building blocks for constructing the global life table that incorporates child mortality, referred to here as the  $GLT_{WCM}$ . Because the  $P_{aitmn}$  are multivariate, the global life table is itself multivariate, as are all of the measures derived from it.

The two other building blocks of the  $GLT_{WCM}$  are the functions  $S_{aitmn}$  and  $f_{aitmn}$ .  $S_{aitmn}$  shows number of “survivors” (women) at age  $a$ , parity  $i$ , duration in parity  $t$ , child mortality state  $m$  at exact age  $a$ , and child mortality state  $n$  at exact age  $a+1$ . (Recall that, because the child mortality variables are lagged by one year, the actual ages are  $a-1$  and  $a$ .) When the difference between  $m$  and  $n$  is 0, there is no change in child mortality state over the interval  $a$  to  $a+1$  and duration  $t$  to  $t+1$ . When the difference is 1, it indicates that, among the women of parity  $i$  who attained duration  $t$  at age  $a$ , child mortality state changes from  $M_m$  to  $M_{m+1}$  over the interval  $a$  to  $a+1$  and duration  $t$  to  $t+1$ . (A detailed presentation of the formulae to construct the global life tables that include child mortality is included in Appendix A.1.)

Because a woman may or may not experience a child death over a one-year interval in the global life table, it is necessary to partition the  $S_{aitmn}$  women into two groups, according to whether or not they experienced a child death. This requires estimation of a partitioning factor, denoted by  $Q_{aitm}$ , which represents the proportion who experience a child death over the interval. The proportion of women who do not experience a child death is then  $1-Q_{aitm}$  (Appendix A.2 elaborates the procedure to calculate the partitioning factors.)

## Estimating Child Mortality Rates from the Global Life Table

The fact that women are partitioned into two groups at any given age, parity, duration in parity, and child mortality state in the  $GLT_{WCM}$  allows estimation of the number of changes in child mortality state (i.e., the number of child deaths) that occur in the global life table as a whole, as women progress from age 10 to age 50. The number of changes in child mortality state is the same as the number of child deaths lagged by one year. Over each one-year age interval, only one change in child mortality state (i.e., one lagged child death) can occur. So, if one calculates the number of women who change their starting child mortality state between ages  $a$

and  $a+1$  in the global life table, one can also calculate an age-specific child mortality rate, where age refers to woman's age, and child deaths can occur at any age of the child.

In the global life table, women who change their starting child mortality state are located at  $S_{aitmn}$ , where  $n = m + 1$ . The following formulae calculate the child mortality rate by age of woman (lagged by one year) and the overall mean number of child deaths per woman:

$$\bar{D}_a = (\sum S_{aitmn})/1,000 \quad (2.4)$$

where  $\bar{D}_a$  is the child mortality rate by woman's age, with child deaths lagged by one year. The summation is over  $i$ ,  $t$ ,  $m$ , and  $n$ , but only for cases where  $n = m + 1$ .

If the summation in equation (2.4) also includes age  $a$ , the result is overall mean number of child deaths per woman:

$$\bar{D} = (\sum S_{aitmn})/1,000 \quad (2.5)$$

This child mortality measure is different than conventional child mortality measures. It reflects fertility as well as conventionally measured child mortality. In the case of conventionally measured child mortality, children rather than women are the units of analysis. But women are units of analysis in the case of  $\bar{D}$ . This measure reflects fertility, since the higher the fertility, the higher the level of  $\bar{D}$ .

As mentioned by Retherford et al. (2010b) in the discussion of the  $P_{ait}$  method, a seemingly illogical property of the global life table is that the number of women at any given age,  $S_a$ , does not equal the starting radix of 1,000 women, due to the global life table property that women can experience more than one event; i.e., next birth; in any given one-year age interval. However, this property does not affect the calculation of the lagged child mortality rates. The reason is that the global life table does not allow women to experience more than one child death in a single year of age.

## Estimating the Effect of Child Mortality on Fertility Using the Global Life Table

There are different ways to calculate the effect of child mortality on fertility using the  $P_{aimn}$  method. In all cases, the starting point is to construct a child-mortality-effect-free global life table, which one can do using the  $P_{aimn}$  approach and setting the partitioning factor equal to zero. Setting the partitioning factor to zero means that no one in the global life table experiences a change in child mortality state; i.e., there are no child deaths. In other words, when the partitioning factor is zero, it is always true that  $S_{a,i,t,m,n} = S_{a,i,t,0,0}$ . We denote the child-mortality-effect-free global life table as  $GLT_{NCM}$ .

Since all measures derived from the  $GLT_{NCM}$  are also child-mortality-effect-free, one can estimate the effect of child mortality on fertility by comparing these measures with corresponding measures calculated from the standard global life table with child mortality ( $GLT_{WCM}$ ). For example, the difference between  $TFR_{WCM}$  (the child-mortality-effect-present TFR) and  $TFR_{NCM}$  (the child-mortality-effect-free TFR) is attributable to the effect of child mortality on TFR:

$$\Delta TFR = TFR_{WCM} - TFR_{NCM} \quad (2.6)$$

$\Delta TFR$  depends on the magnitude of child mortality (i.e., mean number of child deaths per woman). The higher the mean number of child deaths per woman, the higher  $\Delta TFR$ . A more accurate way to measure the effect of child mortality on TFR, however, is to use the replacement rate.

The replacement rate is defined as the difference between the child-mortality-effect-present TFR and the child-mortality-effect-free TFR, divided by the overall mean number of child deaths per woman,  $\bar{D}$ :

$$\text{Replacement Rate} = \frac{\Delta TFR}{\bar{D}} \quad (2.7)$$

The replacement rate indicates the extent to which women replace dead children with additional births.

## VALIDATION OF THE METHODOLOGY

This section tests the validity of the  $P_{aitmn}$  method by comparing the fertility and child mortality measures derived from it with estimates derived from the other methods (birth history method,  $P_{it}$ , and  $P_{ait}$ ), to the extent possible. The methodology is considered valid if the estimates derived by the various methods reasonably agree.

The validation procedure includes the application of the four methods (birth history method,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$ ) to both cohort and period data from India's first National Family Health Survey (NFHS-1), conducted in 1992/93. The cohort analysis pertains to all women age 45-49 at the time of the survey. The period analysis pertains to the five-year period preceding the survey (1988 to 1992)<sup>8</sup>. The difference between cohort and period analysis is in how one constructs the expanded data set. In cohort analysis the expanded data set includes all person-year observations for women from the age at which the life table starts, e.g., age 10, to their current age. In contrast, in period analysis the expanded data set includes only those person-year observations located in the period of interest. (Each person-year observation is located in a particular calendar year.)

One expects estimates derived by the  $P_{aitmn}$  method to agree closely with estimates derived by the other methods when cohort data are used. However, when using period data, one deals with a synthetic cohort instead of an actual cohort, which makes the comparisons between the birth history method and the life table methods more complicated. In period analysis the biggest difference in results among the various methods is expected to be between the birth history method and the global life table methods (both  $P_{ait}$  and  $P_{aitmn}$ ), because in a global life table the distribution of women at higher-order parities depends on the distribution of women at lower-order parities. In the actual population the distribution of women at higher-order parities does not depend on the distribution of women at lower-order parities, because the synthetic cohort splices person-year observations from different cohorts.

---

<sup>8</sup> Calendar years in this analysis do not exactly match actual calendar years. Calendar years in this research are calculated by counting by increments of 12 months backward from the first complete survey month. For example, 1992 refers to the first 12 months preceding the 1992/93 NFHS survey. The second 12 months preceding the survey are labeled 1991, and so on. The survey month is not included in the analysis because it is an incomplete month for most women.

Both quantum and tempo measures of fertility and child mortality are calculated with all four methods, insofar as possible. The birth history method directly uses the data set to calculate these measures<sup>9</sup>, while the  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods rely on discrete-time survival models to estimate the measures.

## **Cohort Analysis**

This section applies the  $P_{aitmn}$  method to the cohort of women age 45-49 in India's NFHS-1. The fertility and child mortality measures derived by the global life table are then compared with the fertility and child mortality measures estimated by the birth history,  $P_{it}$ , and  $P_{ait}$  methods. The estimates of quantum and tempo measures of fertility and child mortality are considered separately.

### ***Quantum measures of fertility***

Table 3.1 compares, to the extent possible, the quantum measures of fertility—TFRs, PPRs, and ASFRs—derived by the four methods.

---

<sup>9</sup> To make the comparisons between the birth history method and the life table methods accurate, the data set for birth history method is constructed the same way as the data set for the life table methods. For example, all first births after age 40 and next births after 10 years of duration in parity are eliminated from both data sets.

**Table 3.1: Comparison of quantum measures of fertility derived by the birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: Cohort analysis, women age 45-49, India's NFHS-1**

| PPR        | Birth History | $P_{it}$ Method | $P_{ait}$ Method | $P_{aitmn}$ Method | ASFR         | Birth History | $P_{ait}$ Method | $P_{aitmn}$ Method |
|------------|---------------|-----------------|------------------|--------------------|--------------|---------------|------------------|--------------------|
| $p_0$      | 0.96          | 0.96            | 0.96             | 0.96               | <b>10-14</b> | 15            | 9                | 9                  |
| $p_1$      | 0.95          | 0.94            | 0.94             | 0.94               | <b>15-19</b> | 172           | 156              | 154                |
| $p_2$      | 0.92          | 0.91            | 0.91             | 0.91               | <b>20-24</b> | 294           | 291              | 286                |
| $p_3$      | 0.86          | 0.85            | 0.85             | 0.85               | <b>25-29</b> | 262           | 265              | 260                |
| $p_4$      | 0.79          | 0.78            | 0.79             | 0.78               | <b>30-34</b> | 168           | 170              | 167                |
| $p_5$      | 0.73          | 0.73            | 0.73             | 0.72               | <b>35-39</b> | 81            | 77               | 76                 |
| $p_6$      | 0.68          | 0.67            | 0.68             | 0.66               | <b>40-44</b> | 21            | 22               | 22                 |
| $p_7$      | 0.62          | 0.62            | 0.62             | 0.60               | <b>45-49</b> | 6             | 5                | 5                  |
| $p_8$      | 0.56          | 0.57            | 0.57             | 0.54               |              |               |                  |                    |
| $p_9$      | 0.51          | 0.53            | 0.48             | 0.47               |              |               |                  |                    |
| $p_{10}$   | 0.42          | 0.45            | 0.47             | 0.46               |              |               |                  |                    |
| $p_{11}$   | 0.37          | 0.40            | 0.45             | 0.44               |              |               |                  |                    |
| $p_{12}$   | 0.34          | 0.39            | 0.42             | 0.41               |              |               |                  |                    |
| $p_{13}$   | 0.44          | 0.46            | 0.38             | 0.37               |              |               |                  |                    |
| $p_{14}$   | 0.54          | 0.71            | 0.35             | 0.33               |              |               |                  |                    |
| $p_{15}$   | 0.30          |                 |                  |                    |              |               |                  |                    |
| <b>TFR</b> | 5.03          | 4.93            | 4.97             | 4.89               | <b>TFR</b>   | 5.09          | 4.97             | 4.89               |
| <b>CEB</b> | 5.03          |                 |                  |                    |              |               |                  |                    |

Notes: In the  $P_{it}$  method, the underlying CLL model for each parity transition includes duration in parity as the sole basic predictor variable. In the  $P_{ait}$  method, the underlying CLL model for each parity transition includes age and duration in parity as the basic predictor variables. In the  $P_{aitmn}$  method, the underlying CLL model for each parity transition includes age, duration in parity, and child mortality as the basic predictor variables.

The  $P_{it}$  method does not yield ASFRs.

The open-ended parity for both the  $P_{ait}$  and  $P_{aitmn}$  methods is 9+-10+.

ASFRs are per 1000 women.

TFR<sub>PPR</sub> and TFR<sub>ASFR</sub>, in birth history method are not equal because TFR<sub>PPR</sub> takes into account censoring at age 45-49 while TFR<sub>ASFR</sub> does not.

As the table shows, the TFR calculated with the birth history method equals the mean children ever born (CEB) of the cohort of analysis, as it was expected. Taking the birth history as the base of the comparison, the table shows that all three life table methods ( $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$ ) underestimate the TFR. PPRs at lower-order parities are higher for the birth history method than the life table methods, but at higher-order parity transitions the PPR estimated by life table methods are higher than the birth history method. The TFRs calculated from different methods, however, are at acceptable levels. Experience shows the life table approach usually gives TFRs that are as much as 0.5 of a child too low or too high compared with the birth history method (Retherford et al., 2009; National Bureau of Statistics of China and East-West Center, 2007).

Part of this underestimation or overestimation could be attributed to the underlying regression models of the life table approach. Since the fit of the models to the data set is not perfect, some discrepancy may occur. The small discrepancy between the different methods is acceptable, as the life table approaches are usually used to study the effect of socioeconomic variables on fertility, not to measure the fertility rates in the general population.

As the table shows, the pattern of ASFRs in the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods is similar, but the level of fertility at each age group differs by method. The  $P_{aitmn}$  method has the lowest level of fertility for almost all age groups. The  $P_{ait}$  method has lower fertility rates than the birth history method and a higher fertility level than the  $P_{aitmn}$  method for most age groups. However, the differences in ASFRs and TFR of the three methods are minimal.<sup>10</sup>

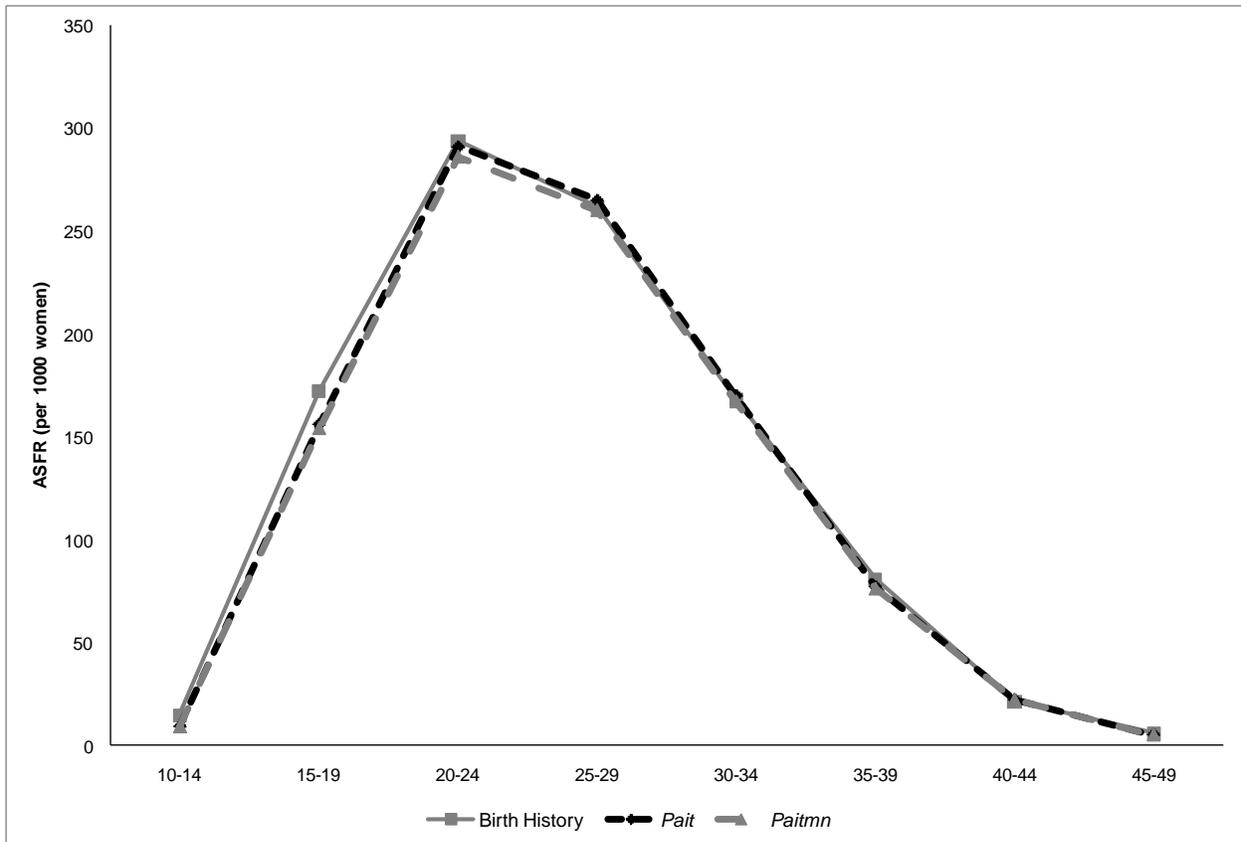
The ASFRs at 10-14 and 15-19 are underestimated for both  $P_{ait}$  and  $P_{aitmn}$  methods (see Figure 3.1)<sup>11</sup>. This underestimation may be related to the quadratic specification of the effect of age on probabilities of failure. It seems that the quadratic specification of age does not completely capture the effect of age on fertility. It is worth noting that for both the  $P_{ait}$  and  $P_{aitmn}$  methods the effect of age on the probabilities of failure is specified as quadratic by means of  $A$  and  $A^2$ .

---

<sup>10</sup> The TFRs calculated from birth history method are not identical because  $TFR_{PPR}$  takes into account the censoring at last age groups, i.e., 45-49, but  $TFR_{ASFR}$  does not take into account censoring at this age group. Censoring occurs because not all women are fully exposed to this age group.

<sup>11</sup> The  $P_{ii}$  method does not yield ASFRs because age is not included in its underlying CLL models.

**Figure 3.1: Comparison of ASFRs derived by the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods: Cohort analysis, women age 45-49, India's NFHS-1**

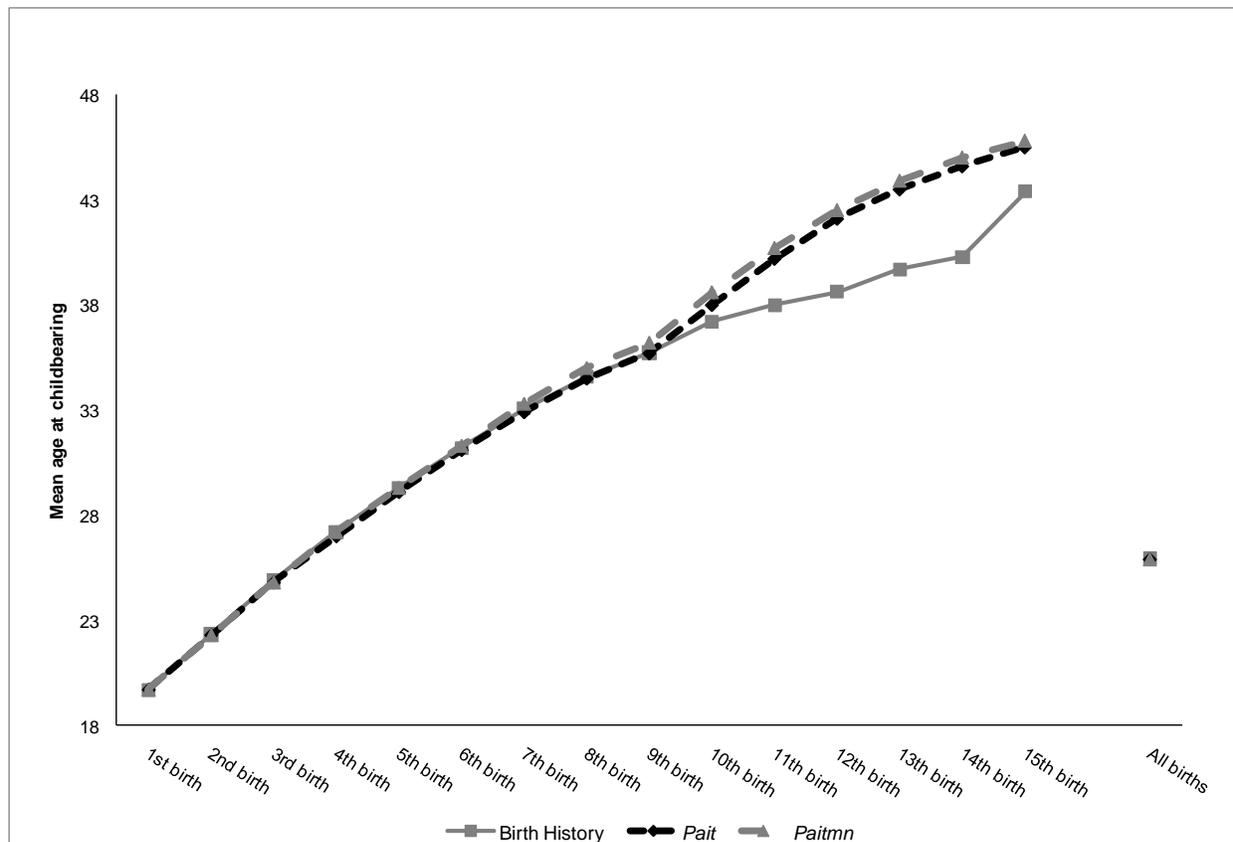


### *Tempo measures of fertility*

The tempo measures of fertility include mean and median age at childbearing (both all births and by birth order) and mean and median closed birth intervals<sup>12</sup>. Figure 3.2 compares the mean ages at childbearing by birth order and all births for the cohort of women age 45-49 in India's NFHS-1.

<sup>12</sup> Only mean age at childbearing and mean closed birth intervals are presented in this paper.

Figure 3.2: Comparison of mean age at childbearing derived by the birth history,  $P_{ait}$  and  $P_{aitmn}$  methods: Cohort analysis, women age 45-49, India's NFHS-1

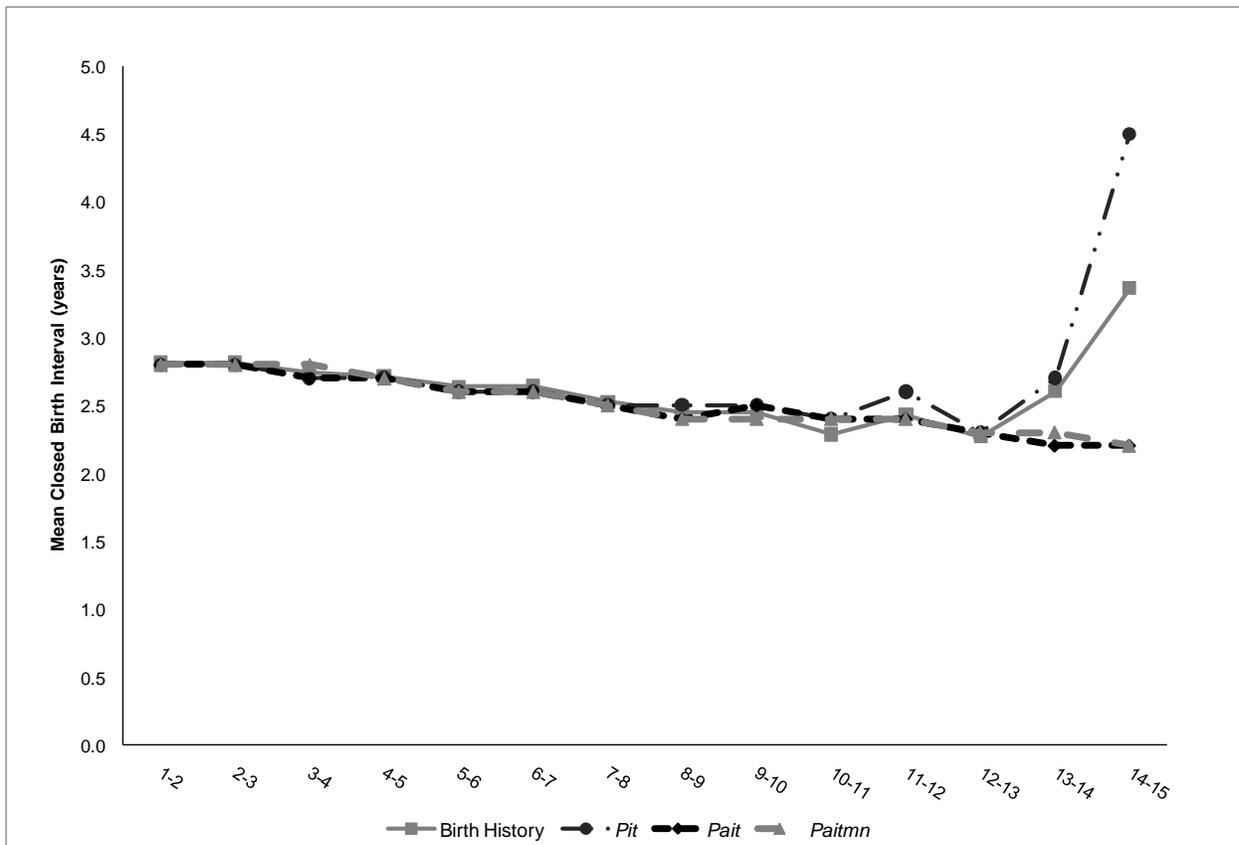


As the figure shows, mean ages at childbearing of all three methods are close to each other in parities one to nine, where each parity transition has its own CLL model in the life table approaches, i.e., the  $P_{ait}$  and  $P_{aitmn}$  methods. At higher-order parity transitions mean ages of childbearing calculated from life table approaches are different than those of the birth history method, due to pooling of all higher-order transitions in a single open-ended parity transition. However, the two life table methods have similar mean ages at childbearing at higher-order parity transitions. The overall mean ages at childbearing calculated from different methods are quite similar.

Figure 3.3 presents the other tempo measures of fertility—that is, mean closed birth intervals. Mean closed birth intervals correspond to the interval between the two births. For instance, the “2-3” shows the birth interval between the second and the third births. The interval between a woman’s own birth and her first birth is not shown here, as it is equivalent to her age

at first birth and thus is shown in the previous figure. As figure 3.3 shows, all four methods yield similar results. At lower parity transitions the means of all methods agree well, while at higher-order transitions the results of different methods do not agree well, due to the pooling of data at open-ended parity transition. The  $P_{it}$  method does not use open-ended parity transition, while the  $P_{ait}$  and  $P_{aitmn}$  methods use 9+-10+ as open-ended parity transition.

**Figure 3.3: Comparison of mean closed birth intervals derived by the birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: Cohort analysis, women age 45-49, India's NFHS-1**



### *Measures of child mortality*

The child mortality measures considered here are the overall mean number of child deaths per woman, child mortality rates by women's age (not child's age, which can be any age at death), and mean and median age of women at child death (all child deaths, so that if a woman experiences two child deaths, two ages for her are included in the calculation of mean and median). Only measures derived by the birth history method can be compared with the child mortality measures derived by  $P_{aitmn}$  method, since the other life table methods do not yield child mortality measures. Because child mortality is a one-year lagged variable, we apply the birth history method also by lagging child mortality by one year.<sup>13</sup>

Table 3.2, compares the child mortality measures derived by the birth history method and the  $P_{aitmn}$  method for cohort of women age 45-49 in India's NFHS-1. As the table indicates, the two methods yield overall mean numbers of child deaths per woman that are close together. The overall mean number of child deaths per woman estimated from the  $P_{aitmn}$  method is slightly lower than the equivalent birth history method value, but it is acceptable. In terms of child mortality rates by women's age, as figure 3.4 shows, the two methods yield quite similar patterns. However, the level of the measures is lower for the  $P_{aitmn}$  method than for the birth history method. As the figure shows, most of the underestimation of child mortality by the  $P_{aitmn}$  method occurs at middle age groups (20-29).

---

<sup>13</sup> The procedure to calculate the child mortality (lagged by one year) by age of woman from the original data set is as follows: first the original dataset should be expanded by current age of woman (one person-year observation for each age). Each person-year observation then gets a value for age that shows age of woman at each person-year observation. A counter variable counts number of child mortality experiences at each person-year observation as of one-year ago. Using the expanded dataset and child mortality variable, one can calculate child mortality rates by age of woman (not child's age that can be any age at death) just like the calculation of ASFRs. The child mortality rates could be reported by single ages of woman or averaged by five-year age groups. The overall mean number of child deaths per woman is calculated by either summing the rates by single ages or summing the five-year rates and multiplying the sum by five. The overall mean number of child deaths per woman could also be calculated directly from the dataset, without expansion, by counting number of child mortality experiences for each woman as of one year ago and taking the mean of the variable.

**Table 3.2: Comparison of child mortality measures derived by the birth history and  $P_{aitmn}$  methods: Cohort analysis, women age 45-49, India's NFHS-1**

| <b>Age Group</b>         | <b>Birth History</b> | <b><math>P_{aitmn}</math> Method</b> |
|--------------------------|----------------------|--------------------------------------|
| 10-14                    | 0                    | 0                                    |
| 15-19                    | 16                   | 13                                   |
| 20-24                    | 48                   | 39                                   |
| 25-29                    | 53                   | 48                                   |
| 30-34                    | 42                   | 41                                   |
| 35-39                    | 24                   | 27                                   |
| 40-44                    | 14                   | 14                                   |
| 45-49                    | 8                    | 6                                    |
| <b>Overall</b>           | <b>1.03</b>          | <b>0.94</b>                          |
| <b>Mean</b>              | 29.2                 | 29.5                                 |
| <b>Median</b>            | 28.7                 | 29.3                                 |
| <b>TFR<sub>WCM</sub></b> |                      | 4.89                                 |
| <b>TFR<sub>NCM</sub></b> |                      | 4.58                                 |
| <b>Replacement Rate</b>  |                      | 0.33                                 |

Note: age specific rates are per 1000 women.

“Overall” refers to mean number of child deaths per woman lagged by one year.

Mean and median refer to mean and median age of women at child death.

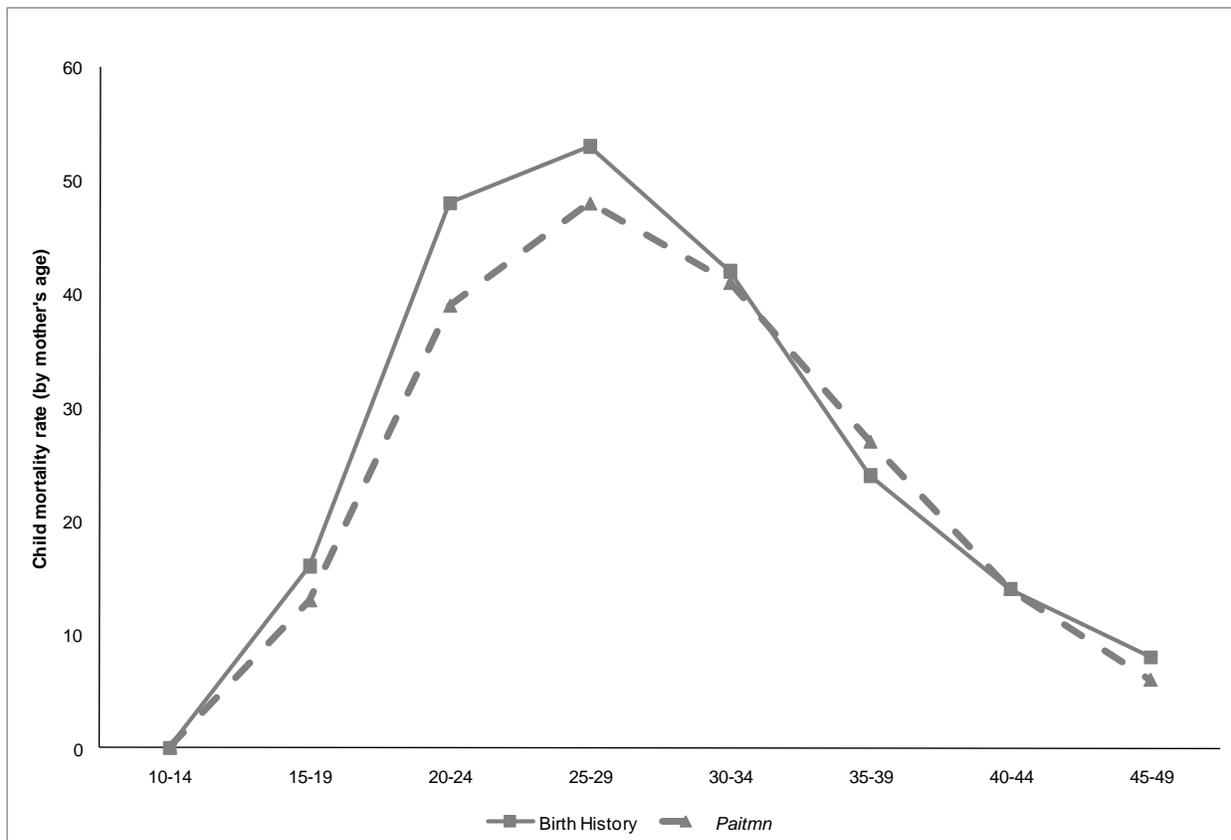
TFR<sub>WCM</sub> refers to child-mortality-effect-present TFR. TFR<sub>NCM</sub> refers to child-mortality-effect-free TFR.

Replacement rate is  $(TFR_{WCM} - TFR_{NCM})/\text{mean number of child deaths per woman}$ .

Mean age of women at child death (lagged by one year), calculated from the two methods agrees well. The two medians also agree but the median calculated from the birth history method is slightly younger than the median estimated by the  $P_{aitmn}$  method.

For the  $P_{aitmn}$  method, in the table we present also the replacement rate, which cannot be calculated by applying the birth history method. The replacement rate is about 0.33, which means that in the cohort of study every 100 children who die are replaced by 33 additional births.

**Figure 3.4: Comparison of child mortality rates (lagged by one year) by mother's age derived by the birth history and  $P_{aitmn}$  methods: Cohort analysis, women age 45-49, India's NFHS-1**



### Period Analysis

This section applies the  $P_{aitmn}$  method to period data for the five years preceding India's NFHS-1 (1988 to 1992). The fertility and child mortality measures derived by the global life table are then compared with fertility and child mortality measures estimated by the birth history,  $P_{it}$ , and  $P_{ait}$  methods insofar as possible. The comparisons are made, once again, by quantum and tempo measures of fertility and child mortality.

### *Quantum measures of fertility*

Table 3.3 compares quantum measures of fertility derived by all four methods insofar as possible.

**Table 3.3: Comparison of quantum measures of fertility derived by the  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: 5-year period spanning 1988-92, India's NFHS-1**

| PPR        | $P_{it}$<br>Method | $P_{ait}$<br>Method | $P_{aitmn}$<br>Method | ASFR         | Birth<br>History | $P_{ait}$<br>Method | $P_{aitmn}$<br>Method |
|------------|--------------------|---------------------|-----------------------|--------------|------------------|---------------------|-----------------------|
| $p_0$      | 0.92               | 0.92                | 0.92                  | <b>10-14</b> | 0                | 3                   | 3                     |
| $p_1$      | 0.93               | 0.91                | 0.91                  | <b>15-19</b> | 121              | 99                  | 98                    |
| $p_2$      | 0.78               | 0.73                | 0.73                  | <b>20-24</b> | 234              | 225                 | 222                   |
| $p_3$      | 0.66               | 0.62                | 0.61                  | <b>25-29</b> | 172              | 181                 | 178                   |
| $p_4$      | 0.62               | 0.58                | 0.56                  | <b>30-34</b> | 97               | 90                  | 87                    |
| $p_5$      | 0.58               | 0.53                | 0.52                  | <b>35-39</b> | 43               | 32                  | 30                    |
| $p_6$      | 0.59               | 0.55                | 0.53                  | <b>40-44</b> | 15               | 8                   | 7                     |
| $p_7$      | 0.57               | 0.52                | 0.49                  | <b>45-49</b> | 6                | 2                   | 2                     |
| $p_8$      | 0.50               | 0.44                | 0.42                  |              |                  |                     |                       |
| $p_9$      | 0.50               | 0.47                | 0.43                  |              |                  |                     |                       |
| $p_{10}$   | 0.46               | 0.44                | 0.41                  |              |                  |                     |                       |
| $p_{11}$   | 0.46               | 0.41                | 0.38                  |              |                  |                     |                       |
| $p_{12}$   | 0.30               | 0.36                | 0.34                  |              |                  |                     |                       |
| $p_{13}$   | 0.41               | 0.31                | 0.29                  |              |                  |                     |                       |
| $p_{14}$   | 0.67               | 0.26                | 0.25                  |              |                  |                     |                       |
| <b>TFR</b> | 3.51               | 3.20                | 3.13                  | <b>TFR</b>   | 3.44             | 3.20                | 3.13                  |

Note:  $P_{it}$  method refers to life tables in which the underlying CLL model for each parity transition includes duration in parity as the sole basic predictor variable.  $P_{ait}$  method refers to life tables in which the underlying CLL model for each parity transition includes age and duration in parity as the basic predictor variables.  $P_{aitmn}$  method refers to life tables in which the underlying CLL model for each parity transition includes age, duration in parity, and child mortality as the basic predictor variables.

$P_{it}$  method does not yield ASFRs.

The open-ended parity for both  $P_{ait}$  and  $P_{aitmn}$  methods is 9+-10+.

ASFRs are per 1000 women.

As the table shows, the  $P_{it}$  method produces the highest TFR that is closest, although slightly high, to the  $TFR_{asfr}$  calculated by birth history method<sup>14</sup>. Once age is introduced into the life tables derived by the  $P_{ait}$  and  $P_{aitmn}$  methods, the TFR drops substantially. This drop could be attributed to the fact that the distribution of the actual population and the life table population are different and that the actual population has undergone a substantial change in fertility and marriage patterns and levels during the previous decade(s). The change in fertility patterns and levels is one of the reasons that the two populations have different distributions. As discussed earlier, the fact that the global life tables use the synthetic cohort approach makes the distribution

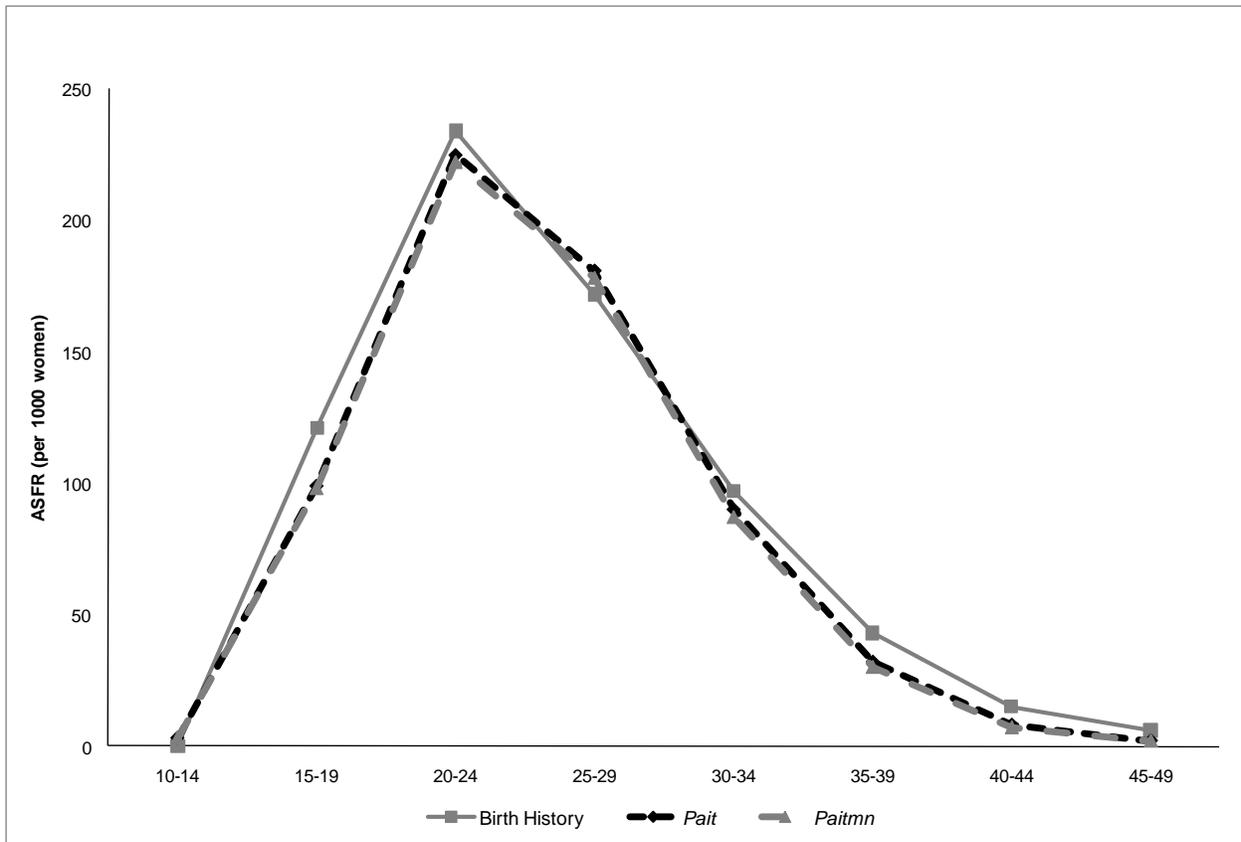
<sup>14</sup> The  $TFR_{PPR}$  is not calculated for period analysis, as it requires use of Feeney's method.

of women by age, parity, and duration in parity in the global life table different than in the actual population.

The birth history method to calculate the PPRs of period data is to apply Feeney’s period parity progression ratios (PPPR) methodology, but it is not applied here.

With regard to the age-specific fertility rates, as the table shows the birth history method produces the highest ASFRs and TFRs, while the ASFRs and TFRs of the  $P_{ait}$  and  $P_{aitmn}$  methods are lower, but close to each other. Figure 3.5 shows the ASFRs of all methods. As the figure indicates, the ASFRs estimated by the birth history method are almost always higher than the ASFRs estimated by the  $P_{ait}$  and  $P_{aitmn}$  methods. Although the levels of ASFRs of the three methods are different, all three methods have a similar pattern. This shows that the  $P_{ait}$  and  $P_{aitmn}$  methods underestimate fertility at all age groups.

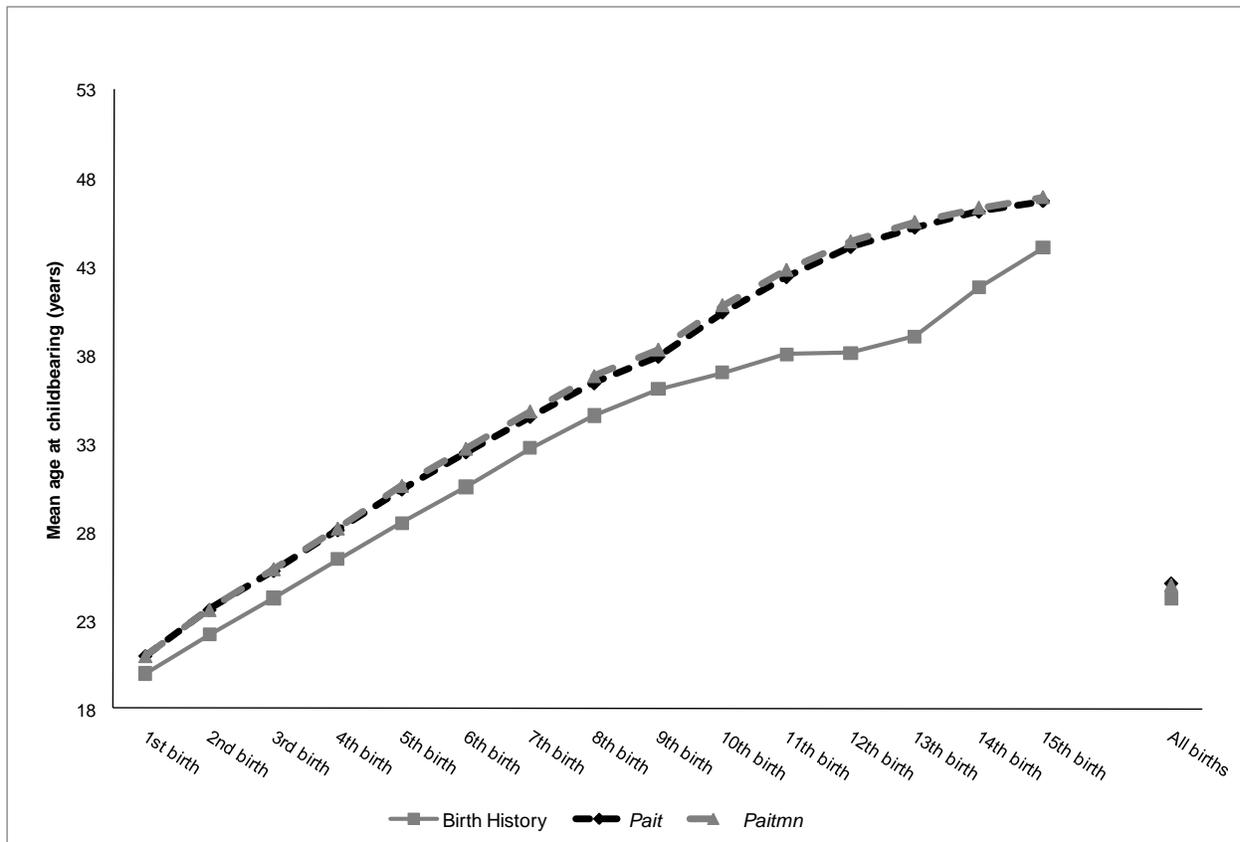
**Figure 3.5: Comparison of ASFRs derived by the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods: 5-year period spanning 1988-92, India's NFHS-1**



### Tempo measures of fertility

The tempo measures of fertility include mean and median age at childbearing (both all births and by birth order) and mean and median closed birth intervals. Figure 3.6 compares mean ages at childbearing, derived alternatively by the birth history, the  $P_{ait}$ , and the  $P_{aitmn}$  methods, for the period of interest. The birth history method to calculate mean age at childbearing uses the ASFRs. In the case of mean ages at childbearing by birth order, the ASFRs at each birth order are used.

**Figure 3.6: Comparison of mean age at childbearing derived by the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods: 5-year period spanning 1988-92, India's NFHS-1**



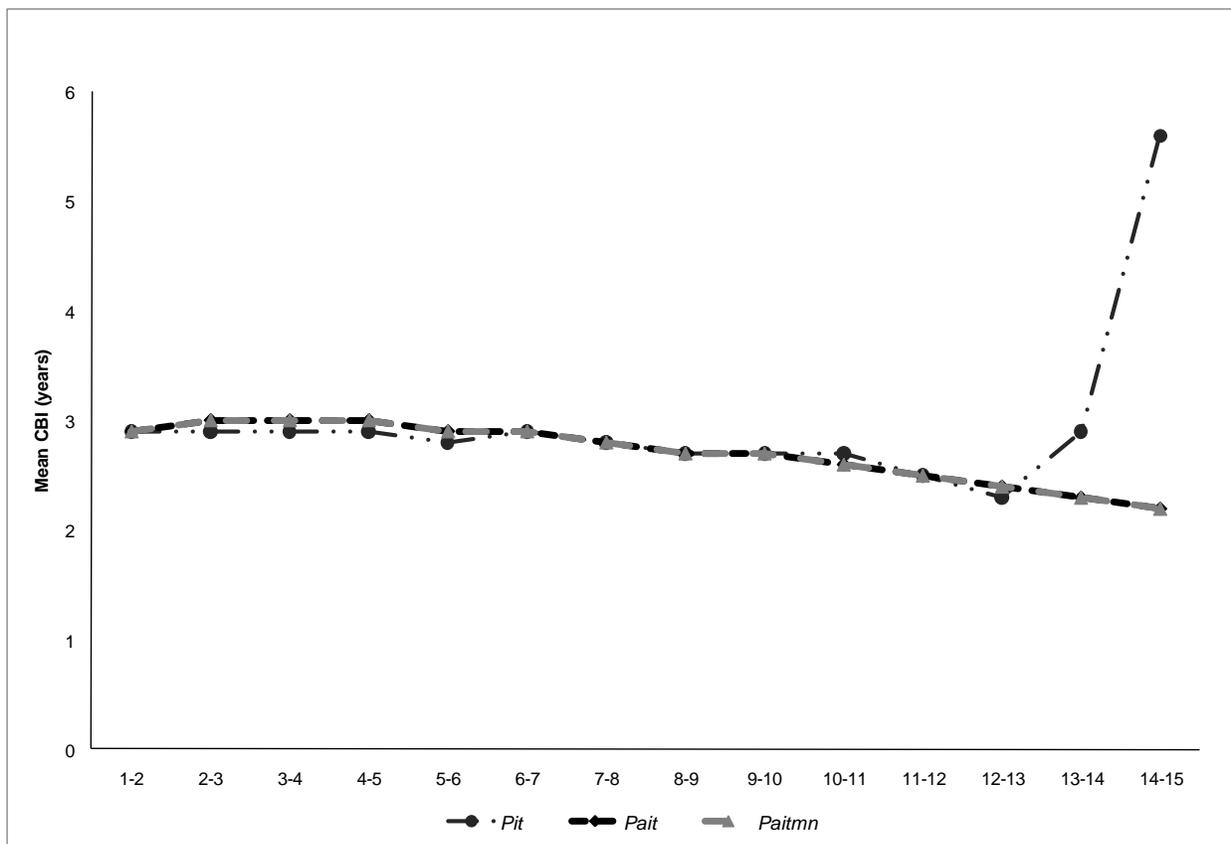
Since the age distribution of the population of the global life table may differ from the age distribution of the actual population, the period estimates of mean age at childbearing by birth order calculated from the birth history method and the global life table methods may not agree well. However, mean ages at childbearing by birth order should agree closely when the

two life table methods (the  $P_{ait}$  and  $P_{aitmn}$  methods) are compared. As the table shows, the means directly calculated from the data set, using the birth history method, are lower than the means calculated by the life table methods. But the estimates of the  $P_{ait}$  and  $P_{aitmn}$  methods are almost identical, showing that the two methods yield similar results. These higher means show older population for life table populations than for the actual population. The overall age at childbearing in the actual population is about one year younger than in the global life table populations. This difference could be attributed to the differences in distribution of actual and global life table populations. The close proximity of the results of the  $P_{ait}$  and  $P_{aitmn}$  methods shows that the two methods have similar population distributions, and introduction of child mortality to the life table does not change the population distribution of the global life table. As previously mentioned, the means calculated from the global life table methods at higher-order transitions are not comparable to the birth history method, because life table approaches use open-ended parity transition and age is not included in the underlying CLL model.

Figure 3.7 presents the next tempo measure of fertility, mean closed birth intervals. The figure shows mean closed birth intervals calculated from the  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods, for transitions 1-2 and higher. Mean closed birth intervals correspond to the mean interval between the two births. The mean closed birth intervals are not calculated by the birth history method, as one would need to apply Feeney's approach.

As the table shows, the three methods yield almost similar results. The results of the  $P_{it}$  method at higher-order births (10<sup>th</sup> birth and higher) are different than the  $P_{ait}$ , and  $P_{aitmn}$  results. This could be related to the open-ended parity transition in the  $P_{ait}$ , and  $P_{aitmn}$  methods. The  $P_{it}$  method does not use open-ended parity transition, while the  $P_{ait}$ , and  $P_{aitmn}$  methods use 9+-10+ as open-ended parity transition.

**Figure 3.7: Comparison of mean closed birth intervals derived by the  $P_{ib}$ ,  $P_{aib}$  and  $P_{aitmn}$  methods: 5-year period spanning 1988-92, India's NFHS-1**



### Measures of child mortality

The child mortality measures that could be calculated from a global life table that includes child mortality are the overall mean number of child deaths per woman, child mortality rates by woman's age (not child's age, which can be any age at death), and mean and median age of women at the child's death. Only the birth history method could be compared with the child mortality measures derived by the  $P_{aitmn}$  method, as the other life table methods do not yield child mortality measures.

Due to lag function, the number of child deaths per woman in the  $P_{aitmn}$  method actually shows the child mortality measure as of one-year ago. In order to make the measures comparable, the birth history method should also lag child mortality by one year.

Table 3.4 compares the child mortality measures derived by both the birth history method and the  $P_{aitmn}$  method for the five-year period 1988-1992 in India's NFHS-1.

**Table 3.4: Comparison of child mortality measures derived by the birth history and  $P_{aitmn}$  methods: 5-year period spanning 1988-92, India's NFHS-1**

| Age Group                | Birth History | $P_{aitmn}$ Method |
|--------------------------|---------------|--------------------|
| 10-14                    | 0             | 0                  |
| 15-19                    | 7             | 4                  |
| 20-24                    | 24            | 15                 |
| 25-29                    | 27            | 17                 |
| 30-34                    | 21            | 12                 |
| 35-39                    | 15            | 7                  |
| 40-44                    | 10            | 3                  |
| 45-49                    | 8             | 1                  |
| <b>Overall</b>           | 0.56          | 0.29               |
| <b>Mean</b>              | 29.0          | 28.4               |
| <b>Median</b>            | 27.9          | 28.0               |
| <b>TFR<sub>WCM</sub></b> |               | 3.13               |
| <b>TFR<sub>NCM</sub></b> |               | 2.99               |
| <b>Replacement Rate</b>  |               | 0.50               |

Note: age specific rates are per 1000 women.

"Overall" refers to mean number of child deaths per woman lagged by one year.

Mean and median refer to mean and median age of women at child death.

TFR<sub>WCM</sub> refers to child-mortality-effect-present TFR. TFR<sub>NCM</sub> refers to child-mortality-effect-free TFR.

Replacement rate is  $(TFR_{WCM} - TFR_{NCM})/\text{mean number of child deaths per woman}$ .

As Table 3.4 indicates, the  $P_{aitmn}$  method underestimates the child mortality measures. The underestimation is much worse in the period case in Table 3.4 than in the cohort case in Table 3.2. The overall mean number of child deaths per woman estimated by this method is about half the number that the birth history method generates. Also the  $P_{aitmn}$  method underestimates child mortality rates by mother's age at all ages, and especially at ages 20 and above. The main reason for discrepancies in child mortality estimates could be the difference in age distributions of the actual period population and the period global life table population. Within a particular age group, the distribution of women by parity and duration in parity evidently differs substantially between the actual population and the global life table population. When child mortality declines dramatically, the period global life table underestimates child mortality. The measures estimated by the global life table actually show that these measures will be achieved by the hypothetical cohort if the rates stay constant and the synthetic cohort follows those rates over the reproductive span.

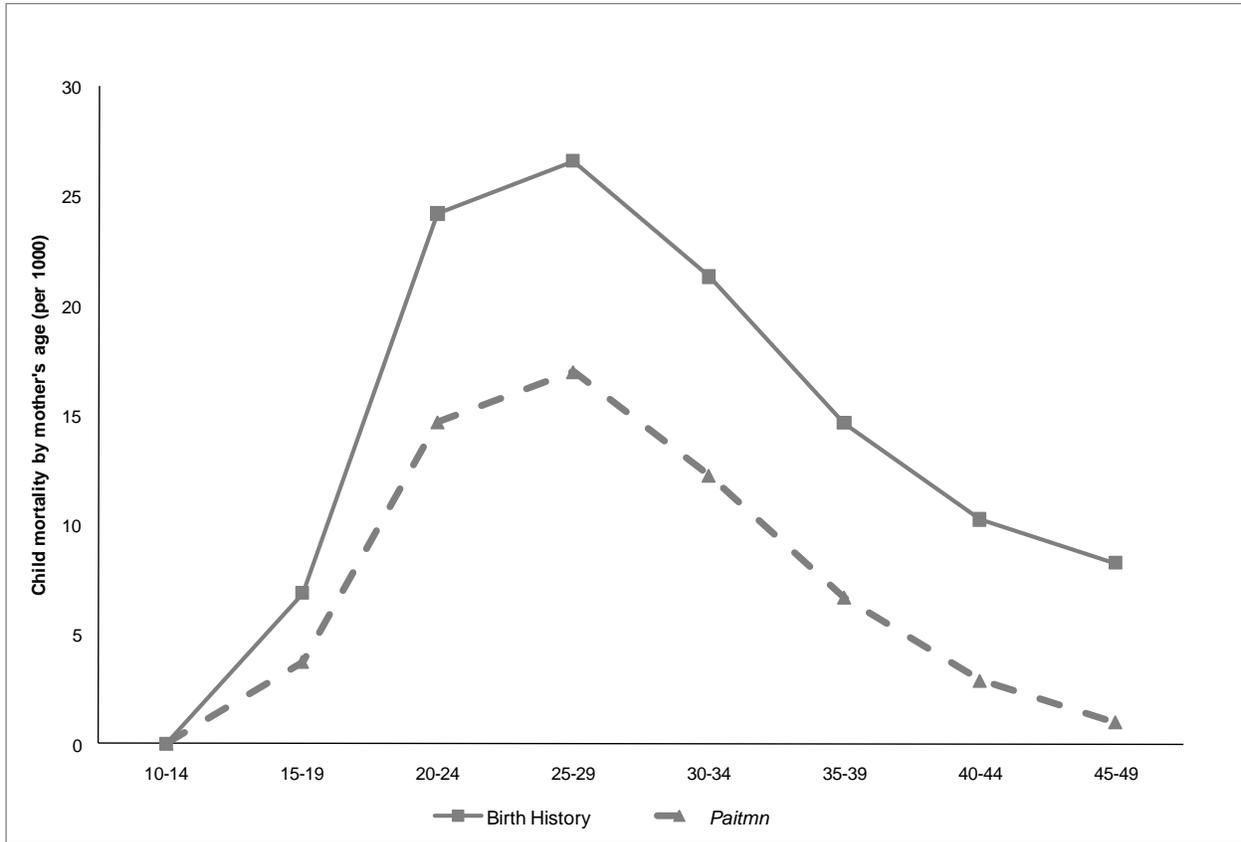
The mean age of women at time of child death (lagged by one year) is slightly higher for the actual population than the global life table population. However, the two populations have almost identical median age of women at time of child death<sup>15</sup>.

The replacement rate estimated by the  $P_{aitmn}$  method is about 0.50 for the period global life table population constructed by using data for the five-year period 1988-1992 from India's NFHS-1. The child replacement rate 0.50 shows that, of each two lost children in the period GLT population, one is replaced by an additional birth.

---

<sup>15</sup> The mean and median ages of woman at child death by the birth history method are also calculated by using the age-specific (woman's age) child mortality rates.

**Figure 3.8: Comparison of child mortality rates (lagged by one year) by mother's age derived by the birth history and  $P_{aitmn}$  methods: 5-year period spanning 1988-92, India's NFHS-1**



As explained earlier, the child mortality measure used in this research is different than conventional child mortality measures, as it contains elements of fertility as well as child mortality. For example, in Figure 3.8 the age-specific child mortality rate is higher than the age-specific fertility rate at age 45-49. At first it appears that there must be some kind of error. But this result could occur inasmuch as there are very few births at age 45-49 but there could be even more child deaths at age 45-49, because the child deaths pertain not only to children born to mothers age 45-49 but also to all of the woman's previous births.

## CONCLUSIONS

This paper develops and validates new multivariate methodology to assess the effect of child mortality on period and cohort fertility. The methodology is based on discrete-time survival models that are used to construct multivariate life tables of parity progression. This new methodology is built on previous research on developing multivariate methods of estimating the TFR and its components. The methodology is important because it yields comprehensive results on the effect of child mortality on fertility. The key measure of the effect is the child replacement rate, which shows the extent to which parents replace lost children by having additional births. The method is also important due to its applicability to period data and its ability to control for reverse causality between child mortality and fertility.

The research tested the validity of the methodology by applying it to India's first National Family Health Survey (NFHS-1), conducted in 1992/93. The application is to both cohort and period data. The results of the analysis show that the new estimates of child mortality and fertility reasonably agree with the estimates derived from other methods.

Although this methodology overcomes the deficiencies of the previous methods used in the literature, it has its own limitation as well. The biggest limitation is the calculation of standard errors of estimates in order to conduct significance tests. The jackknife method is usually used to estimate the standard error of estimates in this method. The jackknife procedure requires repeatedly running the whole process of estimating fertility and child mortality measures by number of clusters in each sample. One drops one cluster from the sample each time the process is repeated until all clusters have dropped once. The process of estimating fertility and child mortality measures include expanding datasets, running CLL models for each parity transition, constructing the global life tables, and estimating the measures from the global life tables. This process should be repeated for each category of predictor variables on each survey. In the case of global life tables that include child mortality, calculation of standard errors for a study similar to this research with few additional socioeconomic variables requires months of super fast computer time.

## NOTATION AND ABBREVIATIONS

*i*: woman's parity (number of children ever born)

*t*: duration in parity (in years)

$T_1, T_2, \dots, T_{k-1}$ : dummy variables representing the first  $k-1$  of  $k$  one-year life table time intervals  
(alternative representation of duration in parity)

*A*: woman's starting age (in years) in any given parity transition

*a*: woman's age (in years) at a particular duration in parity  $t$ , where  $a = A + t$  (in the global life table, but not in the CLL equations, age  $a$  is translated to age  $a-10$ , so that  $a=0$  at age 10 at the beginning of the global life table)

### Parity transitions

**B-M**: transition from a woman's own birth to her first marriage

**M-1**: transition from first marriage to first birth

**0-1**: transition from a woman's own birth to her first birth

**1-2**: transition from first to second birth

**2-3**: transition from second to third birth

**3-4**: transition from third to fourth birth

**4-5**: transition from fourth to fifth birth

**5-6**: transition from fifth to sixth birth

**6+ - 7+**: transition from sixth or higher-order birth to next higher-order birth

**CLL model**: complementary log-log model (one type of discrete-time survival model)

**$P_{it}$  method:** method in which the basic predictor variable in the CLL model for each parity transition is duration in parity

**$P_{ait}$  method:** method in which the basic predictor variables in the CLL model for each parity transition are age and duration in parity

**$P_{aitmn}$  method:** method in which the basic predictor variables in the CLL model for each parity transition are age, duration in parity, and two indicators of child mortality

**$m$ :** child mortality state at the beginning of age  $a-1$  (number of previous child deaths)

**$n$ :** child mortality state at the beginning of age  $a$  (number of previous child deaths)

**$\bar{n}$ :** average mortality state at the beginning of age  $a$  (average number of previous child deaths to women at the beginning of age  $a$ , as calculated for person-year observations with specified characteristics in the expanded data set)

**$M_m$ :** Alternative notation for  $m$

**$M_n$ :** Alternative notation for  $n$

**$D_1$ :** Alternative notation for  $m$

**$D_2$ :** Alternative notation for  $n$

**$\bar{D}$ :** Mean number of child deaths per woman, calculated from the entire global life table

**$P_t$ :** for women of any given parity,  $P_t$  is the conditional probability of failure (next birth) between durations  $t$  and  $t+1$

**$P_{ait}$ :** for women age  $a$ , parity  $i$ , and duration  $t$ ,  $P_{ait}$  is the conditional probability of failure (next birth) between ages  $a$  and  $a+1$  and therefore also between durations  $t$  and  $t+1$

**$P_{aitmn}$ :** for women age  $a$ , parity  $i$ , duration  $t$ , and child mortality states  $m$  and  $n$ ,  $P_{aitmn}$  is the conditional probability of failure (next birth) between ages  $a$  and  $a+1$  and therefore also between durations  $t$  and  $t+1$

$P_{a,i,t,m,\bar{n}}^-$ : for women age  $a$ , parity  $i$ , duration  $t$ , and child mortality states  $m$  and  $\bar{n}$ ,  $P_{a,i,t,m,\bar{n}}^-$  is the conditional probability of failure (next birth) between ages  $a$  and  $a+1$  and therefore also between durations  $t$  and  $t+1$  ( $\bar{n}$  is substituted for  $D_2$  when computing  $P_{a,i,t,m,\bar{n}}^-$  from the underlying CLL model equation)

$f_{aitmn}$ : number of failures (births) between ages  $a$  and  $a+1$  (and therefore also between durations  $t$  and  $t+1$ ) in the global life table to women of parity  $i$  and child mortality states  $m$  and  $n$

$F_{aitmn}$ : failure rates by age, parity, duration in parity, and mortality states  $m$  and  $n$  are obtained by dividing  $f_{aitmn}$  by 1,000.

$Q_{a,i,t,m}$ : partitioning factor, the proportion of women at age  $a$ , parity  $i$ , duration  $t$ , and child mortality state  $m$  for whom child mortality state changes during the transition.

$S_{aitmn}$ : number of women at age  $a$ , parity  $i$ , duration in parity  $t$ , and child mortality states  $m$  and  $n$  in the global life table

**TFR<sub>N<sub>CM</sub></sub>**: child-mortality-effect-free TFR that is calculated from a GLT that incorporates child mortality, which is set to zero

**TFR<sub>W<sub>CM</sub></sub>**: child-mortality-effect-present TFR calculated from a GLT that incorporates child mortality

**CM+**: denotes women with child mortality experience in a GLT that incorporates child mortality

**CM-**: denotes women with no child mortality experience in a GLT that incorporates child mortality

## REFERENCES

- Allison, P. 1982. *Discrete-time methods for the analysis of event histories*. In S. Leinhardt (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.
- Allison, P. 1995. *Survival Analysis Using SAS: A Practical Guide*. Cary, N. C.: SAS Institute Inc.
- Feeney, G. 1983. Population dynamics based on birth intervals and parity progression. *Population Studies* 37: 75-89.
- Feeney, G. 1986. Period Parity Progression Measures of Fertility in Japan. NUPRI Research Papers Series No. 35. Tokyo: Nihon University Population Research Institute.
- Li, X., Y. Hu, J. Chen, R. D. Retherford, and M. K. Choe. 2007. *Fertility Estimates for Provinces of China, 1975-2000*. Beijing: National Bureau of Statistics. Honolulu: East-West Center.
- Retherford, R. D., and M. K. Choe. 1993. *Statistical Models for Causal Analysis*. New York: John Wiley.
- Retherford, R. D., N. Ogawa, R. Matsukura, and H. Eini-Zinab. 2009. Multivariate Analysis of Parity Progression-based Measures of the Total Fertility Rate and its Components Using Individual-level Data. East-West Center Working Paper No. 119. Honolulu: East-West Center. <http://www.eastwestcenter.org/fileadmin/stored/pdfs/POPwp119.pdf>
- Retherford, R. D., N. Ogawa, R. Matsukura, and H. Eini-Zinab. 2010a. Multivariate analysis of parity progression-based measures of the total fertility rate and its components. *Demography* 47: 97-124.
- Retherford, R. D., H. Eini-Zinab, M.K. Choe, N. Ogawa, and R. Matsukura. 2010b. Further Development of Methodology for Multivariate Analysis of the Total Fertility Rate and its Components Based on Birth-history Data. Paper presented at the Population Association of America Annual Meeting, April 15-17, Dallas, Texas.

**APPENDIX A.1: CONSTRUCTING THE GLOBAL LIFE TABLE FROM THE BIRTH PROBABILITIES  $P_{aitmn}$**

As in the development of the  $P_{ait}$  method, the formulae are built up one parity transition at a time. The formulae for the 0-1 transition in the new GLT are basically the same as in the GLT derived by the  $P_{ait}$  method, since there is no previous child mortality in the case of this transition. The formulae for the 0-1 ( $i = 0$ ) transition are:

$$S_{0,0,0,0,0} = 1,000 \tag{A1.1}$$

$$S_{a,0,t,0,0} = S_{a,0,a,0,0} = S_{a-1,0,a-1,0,0} (1 - P_{a-1,0,a-1,0,0}) \quad \text{for } a > 0 \tag{A1.2}$$

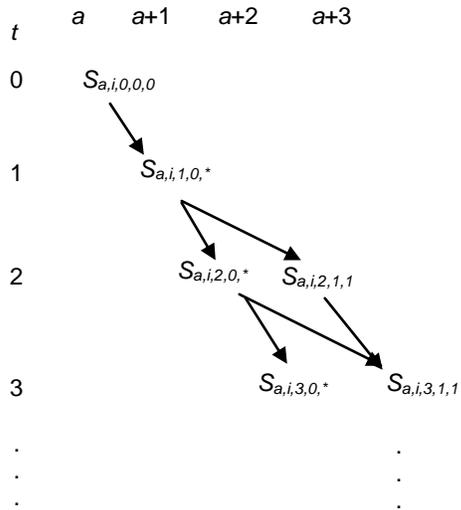
$$f_{a,i,t,0,0} = S_{a,i,t,0,0} P_{a,i,t,0,0} \tag{A1.3}$$

$S_{a,i,t,0,0}$  is number of survivors and  $f_{a,i,t,0,0}$  is number of failures (i.e., births) by age, parity, and duration in parity, respectively. Because there is no one-year lagged child mortality in the case of the 0-1 transition,  $m$  and  $n$  are both zero for  $S_{aitmn}$  and  $f_{aitmn}$ .

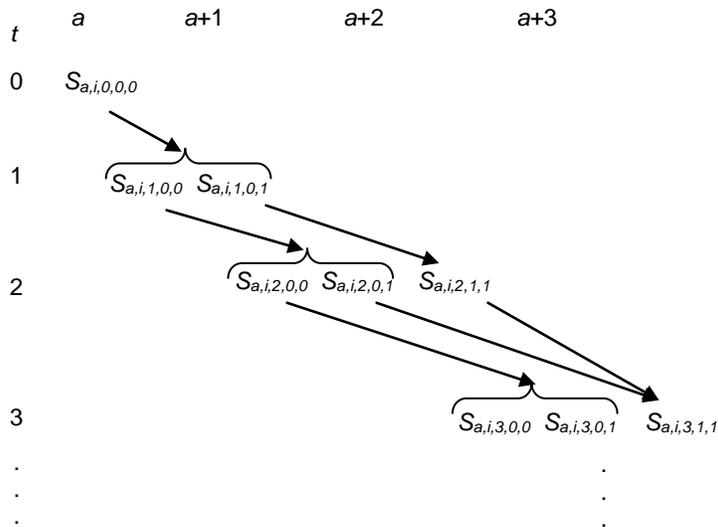
The formulae for higher-order transitions use the same basic logic as that used in the  $P_{ait}$  method, but the formulae are more elaborate because they incorporate non-zero child mortality. As discussed in the text, Panel 1 of Figure 2.1 for the 1-2 parity transition shows how women progress by age, duration in parity, and child mortality state. Panel 1 of Figure A.1 shows the corresponding progression of survivors, starting with  $S_{aitm^*}$ , where the asterisk indicates that the value of  $n$  can be either  $m$  or  $m+1$ . Panel 2 elaborates Panel 1 by partitioning  $S_{aitm^*}$  into two groups, one consisting of women who do not experience a child death during a one-year interval and the other consisting of women who do experience a child death during the one-year interval.

**Figure A.1: Elaboration of panel 1 of Figure 2.1:**

Panel 1: Parity transition 1-2 ( $S_{a,i,t,m^*}$ )



Panel 2: Parity transition 1-2 ( $S_{a,i,t,m,n}$ )



Note: For  $t > 0$ , the  $S_{a,i,t,m^*}$  in panel 1 are partitioned into  $S_{aitmm}$  and  $S_{aitmn}$  (where  $n = m + 1$ ) in panel 2. The asterisk in  $S_{a,i,t,m^*}$  indicates that the child mortality state at the end of the interval (equivalently, at the start of the next interval) could be either  $m$  or  $m+1$

Formulae corresponding to Panel 2 of Figure A.1 for the 1-2 transition are as follows:

If  $t=0$ :

$$S_{a,1,0,0,0} = \sum (S_{a,0,t,0,0} P_{a,0,t,0,0}) = \sum f_{a,0,t,0,0} \quad \text{where the summation is over } t, \quad (\text{A1.4})$$

$$S_{a,1,0,0,1} = 0 \quad (\text{A1.5})$$

$$S_{a,1,0,1,1} = 0 \quad (\text{A1.6})$$

If  $t = 1$ :

$$S_{a,1,1,0,0} = S_{a-1,1,0,0,0} (1 - P_{a,1,1,0,0}) (1 - Q_{a,1,1,0}) \quad \text{for } a > 0 \quad (\text{A1.7})$$

$$S_{a,1,1,0,1} = S_{a-1,1,0,0,0} (1 - P_{a-1,1,0,0,0}) Q_{a,1,1,0} \quad \text{for } a > 0 \quad (\text{A1.8})$$

$$S_{a,1,1,1,1} = 0 \quad (\text{A1.9})$$

If  $t = 2$ :

$$S_{a,1,2,0,0} = S_{a-1,1,1,0,0} (1 - P_{a-1,1,1,0,0}) (1 - Q_{a,1,2,0}) \quad \text{for } a > 0 \quad (\text{A1.10})$$

$$S_{a,1,2,0,1} = S_{a-1,1,1,0,0} (1 - P_{a-1,1,1,0,0}) Q_{a,1,2,0} \quad \text{for } a > 0 \quad (\text{A1.11})$$

$$S_{a,1,2,1,1} = S_{a-1,1,1,0,1} (1 - P_{a-1,1,1,0,1}) \quad \text{for } a > 0 \quad (\text{A1.12})$$

If  $t > 2$ :

$$S_{a,1,t,0,0} = S_{a-1,1,t-1,0,0} (1 - P_{a-1,1,t-1,0,0}) (1 - Q_{a,1,t,0}) \quad \text{for } a > 0 \quad (\text{A1.13})$$

$$S_{a,1,t,0,1} = S_{a-1,1,t-1,0,0} (1 - P_{a-1,1,t-1,0,0}) Q_{a,1,t,0} \quad \text{for } a > 0 \quad (\text{A1.14})$$

$$S_{a,1,t,1,1} = (S_{a-1,1,t-1,0,1}) (1 - P_{a-1,1,t-1,0,1}) + (S_{a-1,1,t-1,1,1}) (1 - P_{a-1,1,t-1,1,1}) \quad \text{for } a > 0 \quad (\text{A1.15})$$

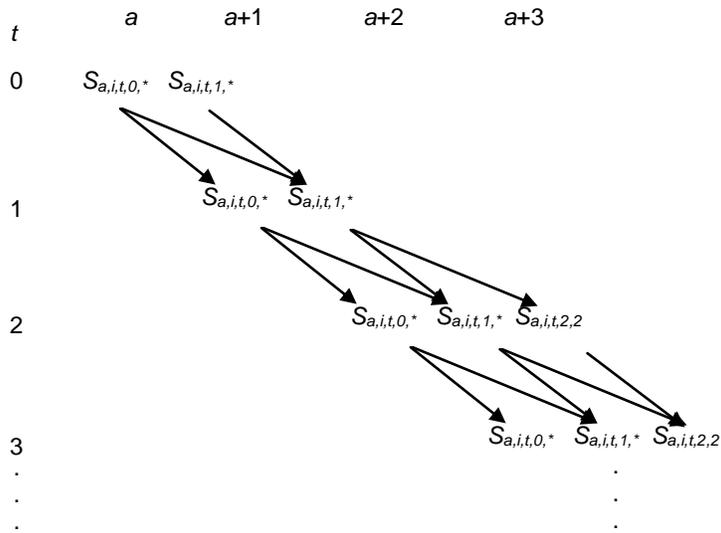
$$f_{aitmn} = S_{aitmn} P_{aitmn} \quad (\text{A1.16})$$

The formulae for higher-order transitions follow the same logic, with the following exceptions: (1) When  $t = 0$ , one needs to add number of failures (i.e., number of births) from the previous parity transition in order to determine how many women start the next parity transition in each mortality state. (2) More child mortality states are added as parity increases.

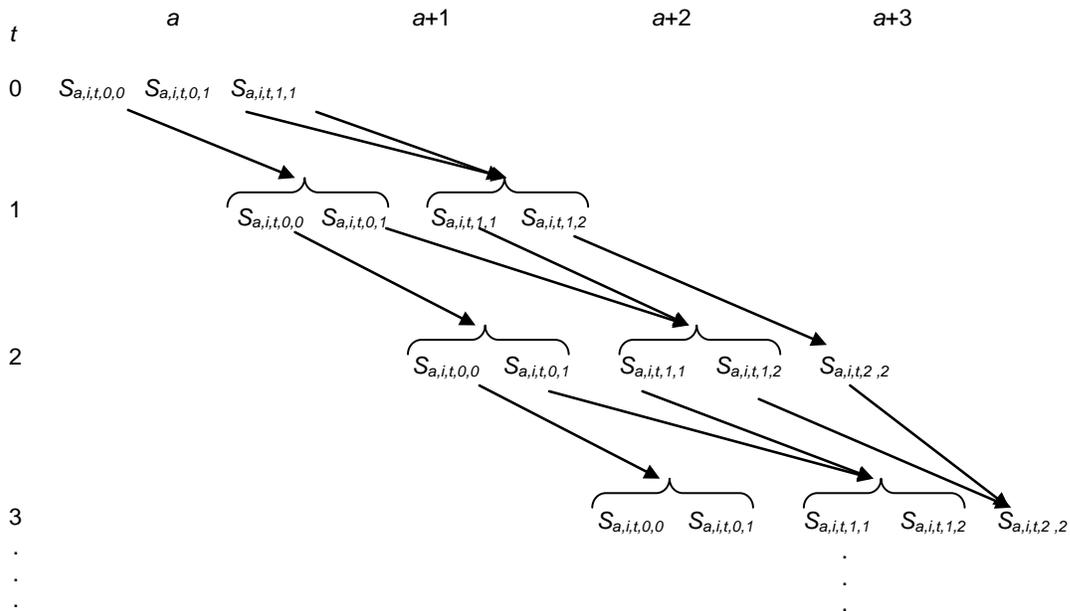
As an example, the formulae for parity transition 2-3 ( $i = 2$ ) are presented next. In order to clarify the formulae for the 2-3 transition, Panel 2 of Figure 2.1 is elaborated in a Figure A.2 to show how women are partitioned into two groups at each age, parity, duration in parity, and child mortality state.

**Figure A.2: Elaboration of panel 2 of Figure 2.1:**

Panel 1: Parity transition 2-3 ( $S_{a,i,t,m^*}$ )



Panel 2: Parity transition 2-3 ( $S_{a,i,t,m,n}$ )



Note:  $S_{a,i,t,m^*}$ 's of panel 1 are partitioned to  $S_{aitmm}$  and  $S_{aitmn}$  (where  $n = m + 1$ ) in panel 2. The asterisk in  $S_{a,i,t,m^*}$  shows that the ending child mortality state could be equal to  $m$  or  $m+1$ .

If  $t = 0$ :

$$S_{a,2,0,0,0} = \left( \sum (S_{a,1,t,0,0} P_{a,1,t,0,0}) \right) (1 - Q_{a,2,0,0}) \quad \text{where the summation is over } t \quad (\text{A1.17})$$

$$S_{a,2,0,0,1} = \left( \sum (S_{a,1,t,0,0} P_{a,1,t,0,0}) \right) Q_{a,2,0,0} \quad \text{where the summation is over } t \quad (\text{A1.18})$$

$$S_{a,2,0,1,1} = \sum (S_{a,1,t,0,1} P_{a,1,t,0,1}) + \sum (S_{a,1,t,1,1} P_{a,1,t,1,1}) \quad \text{where the summation is over } t \quad (\text{A1.19})$$

$$S_{a,2,0,1,2} = 0 \quad (\text{A1.20})$$

$$S_{a,2,0,2,2} = 0 \quad (\text{A1.21})$$

If  $t = 1$ :

$$S_{a,2,1,0,0} = S_{a-1,2,0,0,0} (1 - P_{a-1,2,0,0,0}) (1 - Q_{a,2,1,0}) \quad \text{for } a > 0 \quad (\text{A1.22})$$

$$S_{a,2,1,0,1} = S_{a-1,2,0,0,0} (1 - P_{a-1,2,0,0,0}) Q_{a,2,1,0} \quad \text{for } a > 0 \quad (\text{A1.23})$$

$$S_{a,2,1,1,1} = (S_{a-1,2,0,0,1} (1 - P_{a-1,2,0,0,1}) + S_{a-1,2,0,1,1} (1 - P_{a-1,2,0,1,1})) (1 - Q_{a,2,1,1}) \quad \text{for } a > 0 \quad (\text{A1.24})$$

$$S_{a,2,1,1,2} = (S_{a-1,2,0,0,1} (1 - P_{a-1,2,0,0,1}) + S_{a-1,2,0,1,1} (1 - P_{a-1,2,0,1,1})) Q_{a,2,1,1} \quad \text{for } a > 0 \quad (\text{A1.25})$$

$$S_{a,2,1,2,2} = 0 \quad (\text{A1.26})$$

If  $t = 2$ :

$$S_{a,2,2,0,0} = S_{a-1,2,1,0,0} (1 - P_{a-1,2,1,0,0}) (1 - Q_{a,2,2,0}) \quad \text{for } a > 0 \quad (\text{A1.27})$$

$$S_{a,2,2,0,1} = S_{a-1,2,1,0,0} (1 - P_{a-1,2,1,0,0}) Q_{a,2,2,0} \quad \text{for } a > 0 \quad (\text{A1.28})$$

$$S_{a,2,2,1,1} = (S_{a-1,2,1,0,1} (1 - P_{a-1,2,1,0,1}) + S_{a-1,2,1,1,1} (1 - P_{a-1,2,1,1,1})) (1 - Q_{a,2,2,1}) \quad \text{for } a > 0 \quad (\text{A1.29})$$

$$S_{a,2,2,1,2} = (S_{a-1,2,1,0,1} (1 - P_{a-1,2,1,0,1}) + S_{a-1,2,1,1,1} (1 - P_{a-1,2,1,1,1})) Q_{a,2,2,1} \quad \text{for } a > 0 \quad (\text{A1.30})$$

$$S_{a,2,2,2,2} = S_{a-1,2,1,1,2} (1 - P_{a-1,2,1,1,2}) \quad \text{for } a > 0 \quad (\text{A1.31})$$

If  $t > 2$ :

$$S_{a,2,t,0,0} = S_{a-1,2,t-1,0,0}(1 - P_{a-1,2,t-1,0,0})(1 - Q_{a,2,t,0}) \quad \text{for } a > 0 \quad (\text{A1.32})$$

$$S_{a,2,t,0,1} = S_{a-1,2,t-1,0,0}(1 - P_{a-1,2,t-1,0,0})Q_{a,2,t,0} \quad \text{for } a > 0 \quad (\text{A1.33})$$

$$S_{a,2,t,1,1} = (S_{a-1,2,t-1,0,1}(1 - P_{a-1,2,t-1,0,1}) + S_{a-1,2,t-1,1,1}(1 - P_{a-1,2,t-1,1,1}))(1 - Q_{a,2,t,1})$$

$$\text{for } a > 0 \quad (\text{A1.34})$$

$$S_{a,2,t,1,2} = (S_{a-1,2,t-1,0,1}(1 - P_{a-1,2,t-1,0,1}) + S_{a-1,2,t-1,1,1}(1 - P_{a-1,2,t-1,1,1}))Q_{a,2,t,1} \quad \text{for } a > 0 \quad (\text{A1.35})$$

$$S_{a,2,t,2,2} = (S_{a-1,2,t-1,1,2}(1 - P_{a-1,2,t-1,1,2}) + (S_{a-1,2,t-1,2,2}(1 - P_{a-1,2,t-1,2,2}))) \quad \text{for } a > 0 \quad (\text{A1.36})$$

The rest of the global life table is constructed in a similar fashion. The new global life table with child mortality included has all the properties of the original global life table (Retherford et al., 2010b). Once the new global life table is constructed, it is straightforward to calculate the TFR and its components from it. Formulae are similar to those in the original  $P_{ait}$  method for calculating the original global life table in the previous section. The only difference is that summations run over  $m$  and  $n$  as well as  $a$ ,  $i$ , and  $t$ . For example, TFR is calculated from the new global life tables as

$$\text{TFR} = (\sum f_{aimn})/1,000 \quad (\text{A1.37})$$

where the summation is over  $a$ ,  $i$ ,  $t$ ,  $m$ , and  $n$ .

## APPENDIX A.2 THE PARTITIONING FACTOR

As mentioned in the text, the main difference between the original global life table and the global life table that includes child mortality is that women in some mortality transitions of the new global life table are partitioned into two groups, the first of which sustain the starting child mortality state and the second of which experience a change in the child mortality state. To assign women in each group, one needs a partitioning factor. The partitioning factor is defined as the proportion of women at age  $a$ , parity  $i$ , duration  $t$ , and child mortality  $M_m$  who change their child mortality state to  $M_{m+1}$  by age  $a+1$ . During a one-year transition, these women may survive having next birth and stay at parity  $i$  (in which case duration increases to  $t+1$ ), or they may experience the terminating event as well and change their parity to  $i+1$ , in which case their duration in parity reverts to  $t = 0$ .

Two approaches are introduced in this research to calculate the partitioning factor: the “overall probability” approach and the “logit model” approach. The former approach uses the overall probability of failure (i.e., next birth) at age  $a$ , parity  $i$ , duration  $t$ , and child mortality state  $m$ , to estimate the partitioning factor. The latter approach applies logit regression models to estimate the probabilities of change in child mortality state by age, parity, duration in parity, and child mortality state. The probabilities directly show the proportion of women who change their child mortality state by age, parity, duration in parity, and child mortality. The former approach is preferred to the latter (see next two sections for a discussion).

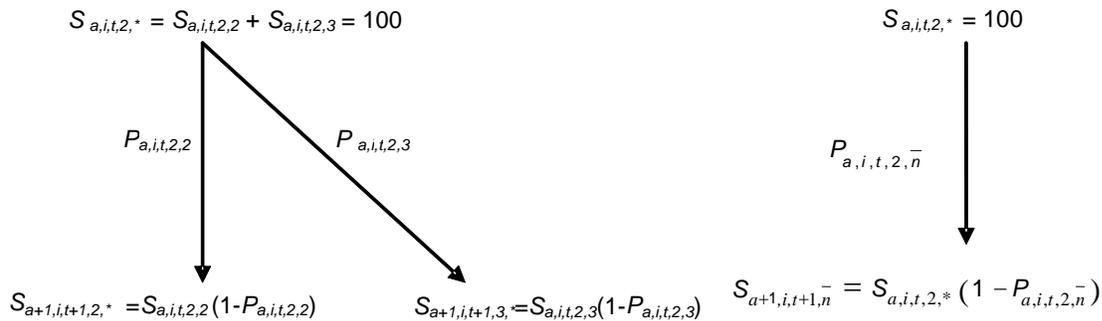
### A.2.1 “Overall Probability” Approach to Calculation of the Partitioning Factor

$S_{aim^*}$  indicates number of women at age  $a$ , parity  $i$ , duration in parity  $t$ , and child mortality state  $m$  in the new global life table. The asterisk shows that child mortality state at the end of a one-year transition could be either  $m$  or  $m+1$ .  $S_{aim^*}$  is partitioned into two groups, the first of which remain in child mortality state  $m$  at the end of a one-year transition, and the second of which experiences a change in child mortality state to  $m+1$  at the end of the one-year transition.

By way of illustration, suppose there are 100 women at age  $a$ , duration  $t$ , of parity  $i$ , with child mortality state 2 ( $M_m = M_2$ ). I.e.,  $S_{aitm^*} = S_{a,i,t,2,*} = 100$ . During age  $a$  and duration  $t$  some of these 100 women will remain in the same child mortality state  $M_2$  and some will move to the next child mortality state  $M_{m+1} = M_3$ . The number of women in the first group is denoted as  $S_{a,i,t,2,2}$ . The probability of having a birth for this group is  $P_{a,i,t,2,2}$ . The number of women in the second group, which changes child mortality state (from  $M_2$  to  $M_3$ ) during age  $a$  and duration  $t$ , is denoted as  $S_{a,i,t,2,3}$  where  $n = m + 1$ . The probability of having a next birth for this group is  $P_{a,i,t,2,3}$ . Thus  $S_{a,i,t,2,*} = S_{a,i,t,2,2} + S_{a,i,t,2,3}$ . The partitioning factor shows the proportion of  $S_{a,i,t,2,*}$  who belong to  $S_{a,i,t,2,3}$ .

The main issue is how to divide the  $S_{a,i,t,2,*}$  women into the two groups  $S_{a,i,t,2,2}$  and  $S_{a,i,t,2,3}$ . One possible solution is to use the overall transition probability to calculate the number of women in each group. In other words, if one knows how many women move to age  $a+1$  and duration  $t+1$  overall, regardless of their subsequent child mortality state, one can calculate the number of women in each group. The overall probability of transition is denoted by  $P_{aitm\bar{n}}$ , where  $\bar{n}$  is the average child mortality state (mean number of previous child deaths as of one year ago) at age  $a-1$  and duration  $t-1$ .  $\bar{n}$  is always equal or greater than  $m$  and equal or less than  $n$ . (The procedure to estimate  $P_{aitm\bar{n}}$  is explained shortly.) In this example  $\bar{n}$  is a number between 2 and 3. Figure A2.1 simplifies this example.

**Figure A2.1: Depicting the “overall probability” approach to the partitioning factor**



The following equations are generated from the figure:

$$S_{a,i,t,2,*} = S_{a,i,t,2,2} + S_{a,i,t,2,3} \quad (\text{A2.1})$$

$$S_{a+1,i,t+1,2,*} = S_{a,i,t,2,2}(1-P_{a,i,t,2,2}) \quad (\text{A2.2})$$

$$S_{a+1,i,t+1,3,*} = S_{a,i,t,2,3}(1-P_{a,i,t,2,3}) \quad (\text{A2.3})$$

It is assumed that

$$S_{a+1,i,t+1,n}^- = S_{a+1,i,t+1,2,*} + S_{a+1,i,t+1,3,*} \quad (\text{A2.4})$$

Equation (A2.4) assumes that the number of women who move to the next age and duration in parity, regardless of their child mortality state, equals the sum of the number of women who remain in child mortality state  $m = 2$  and those who move to child mortality state  $m+1 = 3$ . When equations (A2.1) and (A2.4) are solved simultaneously, using elements from equations (A2.2) and (A2.3), the numbers of women  $S_{a,i,t,2,2}$  and  $S_{a,i,t,2,3}$  are estimated as

$$S_{a,i,t,2,2} = \frac{S_{a,i,t,2,*}(P_{a,i,t,2,3} - P_{a,i,t,2,n}^-)}{(P_{a,i,t,2,3} - P_{a,i,t,2,2})} \quad (\text{A2.5})$$

$$S_{a,i,t,2,3} = \frac{S_{a,i,t,2,*}(P_{a,i,t,2,2} - P_{a,i,t,2,n}^-)}{(P_{a,i,t,2,2} - P_{a,i,t,2,3})} \quad (\text{A2.6})$$

Using equation (A2.6), one can estimate the partitioning factor,  $Q_{aim}$  (the proportion of women who change mortality state) as

$$Q_{a,i,t,2} = \frac{P_{a,i,t,2,2} - P_{a,i,t,2,\bar{n}}}{P_{a,i,t,2,2} - P_{a,i,t,2,3}} \quad (\text{A2.7})$$

This equation could be generalized to any child mortality state as

$$Q_{a,i,t,m} = \frac{P_{a,i,t,m,m} - P_{a,i,t,m,\bar{n}}}{P_{a,i,t,m,m} - P_{a,i,t,m,n}} \quad (\text{A2.8})$$

where  $m$  denotes the starting child mortality state and  $n$  denotes the ending child mortality state if there is a change in child mortality state (i.e.,  $n = m+1$ ). If there is no change in child mortality state,  $m$  is used instead of  $n$ .

$P_{aitmn}^-$  is estimated by substituting values of age  $A$  (where  $A = a - t + 10$ , recall that age is translated to start from 0 instead of 10 (Retherford et al., 2010b)), duration  $t$ , starting child mortality state  $m$ , and mean child mortality state  $\bar{n}$  at the end of a one-year transition (the values of  $m$  and  $\bar{n}$  are substituted for  $D_1$  and  $D_2$ ) into the CLL model equation for each parity transition. Mean child mortality state  $\bar{n}$  at the end of a one-year transition is specific by age  $a$ , parity  $i$ , duration  $t$ , and starting child mortality state  $m$ . For example, to estimate  $P_{25,3,4,2,\bar{n}}^-$ , one substitutes  $A = 31$ ,  $i = 3$ ,  $t = 5$  (recall that duration in CLL models starts from 1 not zero, thus  $t = 5$ , instead of 4),  $D_1 = 2$ , and  $D_2 =$  the computed value of  $\bar{n}$  into equation (2.1), (2.2), or (2.3). Since the starting child mortality state in this example is 2, the mean of  $D_2$ , specified by age, parity, and starting child mortality state, is a number between 2 and 3. If no one changes her child mortality state, the mean is 2. If everyone changes child mortality state, the mean is 3. If some women change their child mortality state and others sustain their starting child mortality state, the mean is a number between 2 and 3. If the CLL model includes predictor variables, the mean of  $\bar{n}$  should be specific to the main predictor variable as well. For instance, to estimate the values of  $P_{aitmn}^-$  by urban/rural residence one should specify the  $\bar{n}$  by residence as well.

Formulae ((A2.7) and (A2.8)) are fine as long as the two sets of probabilities, i.e.,  $P_{a,i,t,2,2}$  and  $P_{a,i,t,2,3}$  (or  $P_{aitmm}$  and  $P_{aitmn}$ , where  $n = m + 1$ ) are not equal. When the probabilities are equal, the formulae are undefined and one needs to find another solution to partition women in two groups. In this research, this problem is solved with estimating the probabilities of changing child mortality state by age, parity, duration in parity, and the starting child mortality state.

### A.2.2 “Logit Model” Approach to Calculation of the Partitioning Factor

Another solution to partitioning women into two groups is to use a logit model to estimate the probabilities of change in child mortality state by age, parity, duration in parity, and starting child mortality state. In this method, for any given parity transition, one uses the same dataset that is used to estimate the CLL model for that transition. The response variable in the logit regression is a dummy variable representing change in child mortality state at each person-year observation, namely  $\Delta D = D_2 - D_1$ . Predictor variables in the logit model include age, duration in parity, starting child mortality, and socioeconomic variables, if any. The logit model is estimated for each parity transition higher than 0-1, where child mortality is involved.

The logit model is used instead of a CLL model because the response variable, change in child mortality state, can occur at any person-year observation in the data. The CLL model should be used only when the event could occur at the last person-year observation.

The probability form of the logit model is:

$$P_{\Delta D} = \frac{1}{1 + e^{-(a+bt+ct^2+dA+eA^2+fD_1+gSES)}} \quad (\text{A2.9})$$

where  $P_{\Delta D}$  is probability of change in child mortality state,  $a$  is the intercept,  $t$  is duration in parity,  $A$  is age at starting parity,  $D_1$  is child mortality state at the start of age  $a$  (not to be confused with the intercept) and duration  $t$ ,  $SES$  is a column vector of socioeconomic predictor which is included in the CLL model as well,  $\mathbf{g}$  is a row vector of coefficients, and  $a, b, c, d, e, f$ , and  $\mathbf{g}$  are the intercept and coefficients to be estimated.

Once the equation is fitted to the dataset, one can estimate the probabilities of change in child mortality state by substituting specific values of age, duration, child mortality state, and socioeconomic predictors, using the same logic explained for the CLL models.

The procedure to estimate the probabilities of change in child mortality state by age, parity, duration in parity, and child mortality state ( $Q_{aitm}$ ) is then similar to the estimation of  $P_{aitmn}$ .  $Q_{aitm}$  estimated this way can then be used as a partitioning factor.

The “logit model” approach complements the “overall probability” approach and should be used only when the probability of a next birth for women who remain in their initial child mortality state and the probability of a next birth for women who experience a child death in the interval are equal. In almost all cases, they will not be equal, in which case the overall probability approach should be used. The main reason for preferring the overall probability approach is that the same CLL models used to construct the global life table are also used to estimate the partitioning factor. Because all probabilities needed to construct the global life table and to estimate the partitioning factor are derived from similar CLL models, these probabilities are internally consistent. This internal consistency constrains the global life table to be internally consistent too, inasmuch as the overall probability of transition will always be a number between the probabilities of failure for women in the two groups (those who remain in their initial child mortality state and those who experience a child death during the one-year interval).

Another advantage of the “overall probability” approach over the logit model approach is that one only needs to estimate a single model for each parity transition in order to construct a global life table. By contrast, if the logit model approach is used, one needs to estimate both a CLL model and a logit model for each parity transition higher than 1-2 transition. Inasmuch as no model fits the data perfectly, the chance of overestimation or underestimation is likely to be higher when two models are used than when one model is used. The next section, on validation of methodology, compares results from these two approaches, when one or the other is used exclusively.

### A.2.3 Validation of the Partitioning Factor

The two approaches for calculating the partitioning factor are tested here. For reasons explained earlier, the “overall probability” approach is preferred over the “logit model” approach. In practice, the latter approach substitutes for the former approach when the equation for the former approach is undefined. This section tests the sensitivity of the global life tables to the choice of partitioning factor.

To test for the sensitivity of the global life table to the choice of partitioning factor, four different global life tables are constructed using four methods. The first method uses the “overall probability” approach to construct the global life table. When equation (A2.8) is undefined, i.e., when the probability of a birth during the interval is the same for the two groups, the partitioning factor is set to zero,  $Q_{aim} = 0$ . In this method, when the two groups have equal probabilities of next birth, all women are assigned to the first group, i.e., no one changes the child mortality state. The second method also uses the “overall probability” approach to construct the global life table. When equation (A2.8) is undefined, the partitioning factor is set to one,  $Q_{aim} = 1$ . In this method, when the two groups have equal probabilities of next birth, all women are assigned to the second group, i.e., everyone changes the child mortality state. The third method uses the “logit model” approach to calculate the partitioning factor throughout the global life table. Finally, the fourth method uses the “overall probability” approach if the probability of a birth during the interval differs between the two groups and uses the “logit model” approach if the probability of a birth during the interval is the same for the two groups. The global life tables are constructed using cohort data (women age 45-49) from India’s NFHS-1. All GLTs use similar CLL models that include no socioeconomic predictors other than variables representing the basic dimensions of the global life table.

The main purposes of constructing four different global life tables are: 1) to show how frequently the issue of similar probabilities of two groups of women in a one-year transition occurs; 2) to demonstrate the effect of this issue on final outputs of the GLT; and 3) to show that the preferred approach yields better results than other methods.

Results show that the fourth method is preferred. Table A.1 shows the results of all four methods. The table includes TFRs, mean number of child deaths per woman, and child replacement rates only.

**Table A.1: Sensitivity of selected global life table results to the choice of the partitioning factor  $Q_{aitm}$ ; Cohort analysis, women age 45-49, NFHS-1**

| Method | TFR  | CM   | RR   |
|--------|------|------|------|
| 1      | 4.89 | 0.94 | 0.33 |
| 2      | 4.89 | 1.00 | 0.31 |
| 3      | 5.04 | 1.27 | 0.36 |
| 4      | 4.89 | 0.94 | 0.33 |

Notes: Partitioning refers to how, over a one-year interval in the global life table, women are partitioned into two groups, the first of which does not experience a child death over the previous interval and the second of which experiences a child death over the previous interval.  $Q_{aitm}$  denotes the partitioning factor and shows the proportion of women assigned to the second group.

Method 1: The "overall probability" approach is used to estimate  $Q_{aitm}$  if the probability of a birth during the interval differs between the two groups.  $Q_{aitm}$  set to zero if the probability of a birth during the interval is the same for the two groups.

Method 2: The "overall probability" approach is used to estimate  $Q_{aitm}$  if the probability of a birth during the interval differs between the two groups.  $Q_{aitm}$  set to one if the probability of a birth during the interval is the same for the two groups.

Method 3: The "logit model" approach is used to estimate  $Q_{aitm}$ .

Method 4: The "overall probability" approach is used to estimate  $Q_{aitm}$  if the probability of a birth during the interval differs between the two groups. The "logit model" approach is used if the probability of a birth during the interval is the same for the two groups.

CM: Mean number of child deaths per woman in the global life table.

RR: Replacement rate.

As the table shows, the three methods that use the "overall probability" approach to assign women in the two groups yield similar TFRs, regardless of what partitioning strategy is used, when the two groups of women have equal probabilities of birth during the one-year interval. This indicates that having equal probabilities is not a frequent event and occurs so rarely that the two extreme solutions,  $Q_{aitm}$  equal to 1 or 0, yield similar TFRs, and also other fertility measures, not shown here. However, since the mean number of child deaths per woman is directly affected by the choice of partitioning factor, the three methods yield different child replacement rates. When all women are assigned to the first group (method 1), the mean number of child deaths per woman is the lowest, since all women who lose a child during one-year transitions where both groups have equal failure probabilities are assigned to a group with no child deaths. Similarly, when all women are assigned to the second group, the mean number of child deaths per woman is the highest. Similar results of methods 1 and 4 are due to the fact that a changing child mortality state is a rare event (about 4 percent in the overall population), as are transitions in the global life table where the two groups have equal failure probabilities. Using the "logit model" approach, when equation (A2.8) is undefined, yields more accurate and realistic results than assigning all women to one group.

Why not use the “logit model” approach only? As mentioned above, the main reason for not using the “logit model” approach per se is to avoid additional errors in construction of the global life tables. It is obvious that most models do not fit to the dataset perfectly. The imperfectly fitted models result in overestimation or underestimation of fertility measures derived by the global life table. When only one type of model (CLL models) is used to construct the global life tables, they have only one source of error. When the logit model added, the source of error is two-fold, which means a bigger discrepancy may occur in the final results. This is obvious from the results of the third method in Table A.1. As the table shows, the mean number of child deaths per woman in this method is 1.268, which is much higher than the actual number, 1.027. Although one might get similar child mortality measures with this method, the method may erroneously assign women in higher (or lower) child mortality groups, which would result in higher (or lower) TFRs (because women with high (low) child mortality experience also have high (low) fertility rates) and higher (or lower) child replacement rates.

