

Population and Health Series

---

No. 123, February 2011

**Multivariate Life Table  
Methodology for Estimating the  
Effect of Child Mortality on the  
Total Fertility Rate and Its  
Components**

Hassan Eini-Zinab



EAST-WEST CENTER  
COLLABORATION • EXPERTISE • LEADERSHIP



---

The East-West Center promotes better relations and understanding among the people and nations of the United States, Asia, and the Pacific through cooperative study, research, and dialogue. Established by the U.S. Congress in 1960, the Center serves as a resource for information and analysis on critical issues of common concern, bringing people together to exchange views, build expertise, and develop policy options.

The Center's 21-acre Honolulu campus, adjacent to the University of Hawai'i at Mānoa, is located midway between Asia and the U.S. mainland and features research, residential, and international conference facilities. The Center's Washington, D.C., office focuses on preparing the United States for an era of growing Asia Pacific prominence.

The Center is an independent, public, nonprofit organization with funding from the U.S. government, and additional support provided by private agencies, individuals, foundations, corporations, and governments in the region.

*East-West Center Working Papers* are circulated for comment and to inform interested colleagues about work in progress at the Center.

For more information about the Center or to order publications, contact:

Publication Sales Office  
East-West Center  
1601 East-West Road  
Honolulu, Hawai'i 96848-1601

Telephone: 808.944.7145

Facsimile: 808.944.7376

Email: [EWCCBooks@EastWestCenter.org](mailto:EWCCBooks@EastWestCenter.org)

Website: [EastWestCenter.org](http://EastWestCenter.org)



Population and Health Series

---

No. 123, February 2011

## Multivariate Life Table Methodology for Estimating the Effect of Child Mortality on the Total Fertility Rate and Its Components

Hassan Eini-Zinab

Hassan Eini-Zinab is a project assistant in Population and Health Studies, Research Program, East-West Center, Honolulu, Hawai'i. Dr. Eini-Zinab can be reached at [eini@hawaii.edu](mailto:eini@hawaii.edu) or [h\\_eini\\_z@yahoo.com](mailto:h_eini_z@yahoo.com).

This working paper is an updated version of a previous working paper, which is based on the author's PhD dissertation.

**East-West Center Working Papers: Population and Health Series** is an unreviewed and unedited prepublication series reporting on research in progress. The views expressed are those of the author and not necessarily those of the Center. Please direct orders and requests to the East-West Center's Publication Sales Office. The price for Working Papers is \$3.00 each plus shipping and handling.

## Table of Contents

Abstract .....	3
Notation Used in this Paper .....	4
Introduction.....	6
Methodology.....	7
Background.....	7
Incorporating Child Mortality into the Underlying CLL Models.....	9
Constructing the Global Life Tables from the Birth Probabilities $P_{aitmn}$ .....	15
Partitioning Factor .....	22
“Overall Probability Approach” to Calculating the Partitioning Factor .....	23
“Logit Model Approach” to Calculating the Partitioning Factor.....	25
Estimating Child Mortality Rates from the Global Life Table.....	25
Estimating the Extent of Replacement of Dead Children from the Global Life Table.....	26
Validation and Application of the $P_{aitmn}$ Method.....	27
Validation of the $P_{aitmn}$ Method .....	31
Validation of the Partitioning Factor .....	50
Application of the Methodology to All Three NFHS Surveys .....	51
Conclusion .....	53
Acknowledgments.....	56
References.....	57

## ABSTRACT

*This paper develops, validates, and applies new multivariate methodology to assess the effect of child mortality on both period and cohort measures of fertility. The methodology, which can be applied to period data as well as cohort data, is based on discrete-time survival models of parity progression that enable construction of a multivariate life table of fertility covering all parity transitions. The five dimensions of this life table are woman's age, parity, duration in parity, and two dimensions representing lagged child mortality (number of dead children at the beginning and end of the previous year when the woman was one year younger). Additional socioeconomic predictor variables are also included in the underlying survival models. The life table is multivariate in the sense that it can be specified for values or categories of one socioeconomic predictor variable while holding other socioeconomic predictors constant. The life table yields a number of measures of both the quantum and the tempo of fertility and child mortality. It also yields a replacement rate, which measures the extent to which child deaths are replaced by additional births. Because the life table is multivariate, all measures calculated from it are also multivariate. By way of illustration, the methodology is applied to three Indian National Family Health Surveys conducted in 1992–93 (NFHS-1), 1998–99 (NFHS-2), and 2005–06 (NFHS-3). Major findings are that dead children are incompletely replaced, and that the replacement rate rises as the total fertility rate falls over the three surveys, reflecting women's increasing ability to achieve their wanted number of surviving children.*

## NOTATION USED IN THIS PAPER

*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)

$A_d$ : Age of woman at child death in the global life table

### Parity transitions

**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 at starting parity and duration in parity

**$P_{aitmn}$  method**: method in which the basic predictor variables in the CLL model for each parity transition are age at starting parity, 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 child 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*

$\Delta D$ :  $D_2 - D_1$

*D*: Mean number of child deaths per woman at age 50, calculated from the global life table

*R*: Replacement rate, calculated as the model-predicted increment in the TFR generated by a one-child increase in the average number of child deaths experienced by women

by the end of their reproductive age spans (i.e., by age 50; women's mortality during the reproductive age span is assumed to be zero in this calculation)

$P_t$ : 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.

$Q_{a,i,t,m}$ : 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

**GLT**: Global life table

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

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

**$\Delta$ TFR**:  $TFR_{WCM} - TFR_{NCM}$

## INTRODUCTION<sup>1</sup>

This paper develops, validates, and applies new multivariate methodology to assess the effect of child mortality on fertility. Methodology is based on discrete-time survival models of parity progression, with one model for each parity transition. A woman's parity is defined as the number of children that she has ever borne. Parity transitions are from a woman's own birth to her first birth, from first to second birth, from second to third birth, and so on. Each model pertaining to a particular parity transition yields a set of model-predicted birth probabilities by woman's age, starting parity, duration in parity, and child mortality (two lagged child mortality variables representing number of child deaths at the beginning and end of the previous year when the woman was one year younger), as well as socioeconomic variables of interest. Together, over all parity transitions, the probabilities by age, parity, duration in parity, and child mortality are used to construct, for specified values of the socioeconomic variables, a five-dimensional life table of fertility covering all parity transitions. This life table is referred to as a "global life table" (GLT) in order to distinguish it from life tables pertaining to each parity transition separately.

The five dimensions of the global life table are age, parity, duration in parity, and the two dimensions representing child mortality. The global life table is multivariate in the sense that it can be constructed by values or categories of one socioeconomic predictor variable while holding the other socioeconomic predictors constant. The multivariate life tables yield various measures of the quantum and tempo of fertility and child mortality. Measures of the quantum of fertility include parity progression ratios (PPRs), age-specific fertility rates (ASFRs), and the total fertility rate (TFR). Measures of the tempo of fertility include mean and median ages at childbearing (by birth order and for all births combined) and mean and median closed birth intervals. For ease of exposition, these quantum and tempo measures are referred to collectively as "components of the TFR." Quantum measures of child mortality, also calculated from the global life table, include 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). Tempo measures of child mortality include mean and median ages of woman at time of child death. These quantum and tempo measures of child mortality are also multivariate. Most important, the multivariate global life table yields a multivariate replacement rate, which measures, on average, the extent to which parents replace a dead child.

The validity of the methodology is tested by applying it to India's first National Family Health Survey (NFHS-1), which was conducted in 1992–93. The application is to both cohort and period data. The cohort analysis is based on 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, and is based on all women age 10–49 at time of survey. Following validation, the methodology is applied in more detail to not only NFHS-1 but also NFHS-2 and NFHS-3, which were conducted in 1998–99 and 2005–06, respectively.

---

<sup>1</sup> This working paper is an updated version of a previous working paper (Eini-Zinab 2010b), which is based on the author's PhD dissertation (Eini-Zinab 2010a).

## METHODOLOGY

### Background

The methodology builds on earlier work, starting with a fertility estimation method developed by Feeney (1983, 1986). Feeney's method, which is not multivariate, is also based on life tables of parity progression, with one life table for each parity transition. Each life table yields a parity progression ratio (PPR). A total fertility rate (TFR) is calculated from the PPRs. The life tables for the various parity transitions also yield estimates of mean and median closed birth intervals by birth order.

Using discrete-time survival models of parity progression (one such model for each parity transition  $i$  to  $i+1$ ), Retherford et al. (2009, 2010a) extended Feeney's method by making it multivariate. For any given value of  $i$ , the survival model for transition  $i$  to  $i+1$  yields a set of model-predicted values of the probabilities  $P_{it}$ , where  $i$  denotes parity,  $t$  denotes duration in parity (measured in years), and  $P_{it}$  denotes the probability of a next birth between durations  $t$  and  $t+1$  for a woman of starting parity  $i$ . The model-predicted values of  $P_{it}$  are multivariate in the already-mentioned sense that they can be tabulated by values or categories of one socioeconomic predictor variable while holding constant the other socioeconomic predictor variables that are included in the model. The probabilities  $P_{it}$  enable construction of a life table of parity progression for the particular parity transition under consideration. Because the  $P_{it}$  are multivariate, the parity-specific life table as a whole is also multivariate, as are all fertility measures calculated from it (e.g., PPR and mean and median closed birth intervals). A multivariate TFR is calculated from the multivariate PPRs pertaining to all the parity transitions. If there are no socioeconomic predictor variables in the underlying survival models, the only remaining predictor variable in each model is duration in parity  $t$ , in which case the methodology yields results that are (to a close approximation) identical to those yielded by Feeney's original method, which is more exact. Retherford et al. (2009, 2010a) refer to this extension of Feeney's method as the  $P_{it}$  method.

Retherford et al. (2010b) subsequently extended the methodology further by adding woman's age  $A$  at starting parity to the set of predictor variables in each of the underlying discrete-time survival models of parity progression. Collectively these models yield multivariate estimates of the birth probabilities  $P_{ait}$ , where  $a$  denotes woman's age at duration  $t$ . The multivariate probabilities  $P_{ait}$  are then used to construct the multivariate global life table (which is actually a set of global life tables, depending on the values of the socioeconomic variables) that spans all parity transitions, in contrast to the separate life table for each parity transition in the  $P_{it}$  method. This global life table yields multivariate estimates of not only PPRs, mean and median closed birth intervals, and TFR, but also ASFRs and mean and median ages at childbirth (both overall and by child's birth order). Retherford et al. (2010b) refer to this extension of the  $P_{it}$  method as the  $P_{ait}$  method.<sup>2</sup>

---

<sup>2</sup> The earlier papers on the  $P_{it}$  and  $P_{ait}$  methods subdivide the parity transition 0-1 into two transitions: woman's own birth to her first marriage, and first marriage to first birth. This is not done in this paper, because child deaths cannot occur during the 0-1 transition, because no birth has yet occurred.

Retherford et al. (2010b) have summarized some general features of the discrete-time survival models that underlie the  $P_{it}$  and  $P_{ait}$  methods. These features, which also underlie the models that underlie the further-extended method developed in the present paper, are the following:

As in the case of the conventional TFR, the global life table, which also yields a TFR, ignores mortality of women. I.e., the calculation of the global life table assumes zero mortality of women during the reproductive age span. (Of course, the global life table that incorporates child mortality does allow that the children of these women may die.)

The type of discrete-time survival model used in these methods is the complementary log–log (CLL) model. The CLL model is preferred over the more commonly used discrete-time logit model for reasons explained by Retherford et al. (2009, 2010a). Regardless of which type of discrete-time survival is used, the CLL model for a particular parity transition is applied not to the original “person sample” but instead to an “expanded sample” of person-year observations created from the original person observations (Allison 1982; 1995). A separate expanded data set is created for each parity transition. For each woman in the original sample who at some time in the past made it to the parity transition’s starting parity, a person-year observation is created for each single-year value of duration in parity  $t$  up to the year of failure or censoring. A person-year observation is created for the year in which failure occurred if a failure did occur, but person-year observations are not created for censored years.

Cases where two births from two different pregnancies occur in the same year and cases where multiple births from the same pregnancy occur during the year are included in the expanded data sets. This is possible because each parity transition is modeled separately, so that two or more successive parity transitions (involving two or more successive failures) for a particular woman can occur in the same year of age—but not in the same year of duration in parity  $t$ , which reverts back to zero immediately after the first in a set of multiple events because of the increase in the woman’s parity. In the case of multiple births from the same pregnancy, births are assumed to occur sequentially, and birth orders are randomly assigned. This way of constructing the expanded data sets guarantees that the estimates of  $P_{it}$  or  $P_{ait}$  calculated from the expanded data sets incorporate all events that occurred, even when more than one event occurred in the same year of age. An assumption underlying this way of modeling multiple events in a one-year age interval is that all events, whether a first birth or a next birth, occur at the start of the interval.

The expanded person-year data set for a particular parity transition makes it easy to include time-varying socioeconomic predictor variables in the CLL model for that transition. For example, if a person moves from rural to urban, some of the person-year observations created for that person are coded as rural and some are coded as urban (which can be done if migration data are both included in the data set and sufficiently

detailed). The CLL model can also handle time-varying effects of predictor variables, by interacting predictor variables with time or some function of time (e.g.,  $t$  and  $t^2$ ).

Multivariate estimates of TFR and its components can be derived not only from cohort data but also from period data. Because the CLL model handles left-censoring as well as right-censoring, one simply bases period estimates of TFR and its components on expanded data sets that treat person-year observations before and after the period of interest as censored. Otherwise the methodology is the same for period data (pertaining to synthetic cohorts) and cohort data (pertaining to real cohorts). The only difference is how the expanded person-year data set is constructed.

The methodology allows calculation of model-predicted estimates of TFR and its components by categories of one socioeconomic predictor variable while controlling for (i.e., holding constant) the other socioeconomic predictor variables that are included in the underlying CLL models. These estimates are referred to as “adjusted estimates.” Estimates of TFR or one of its components by categories of a predictor variable without controlling for other predictor variables, which are omitted from the underlying CLL models, are referred to as “unadjusted estimates.” Regardless of the degree of complexity of the underlying models, both unadjusted and adjusted estimates can be presented in simple bivariate tables or graphs, the essential meaning of which is readily understood by non-statisticians. This is a useful feature of the methodology, especially when trying to communicate results to policymakers.

Standard errors of the model-predicted estimates of TFR and its components can be calculated by the jackknife method, following the approach used in DHS surveys. This was done and significance tests were conducted in the earlier papers using the  $P_{it}$  method (Retherford et al. 2009, 2010a). Application of the jackknife method, which is a brute-force method, was found to be very computer-intensive, however, requiring weeks of two fast desktop computers operating around the clock. Application of the jackknife method promises to be much more computer-intensive in the case of the  $P_{ait}$  method, to which it has never been applied. It is also not applied in this paper.

Further details about these common features may be found in the earlier papers (Retherford et al. 2009, 2010a, 2010b).

### **Incorporating Child Mortality into the Underlying CLL models**

The present paper builds on the  $P_{ait}$  method by incorporating two measures of child mortality as an additional dimensions of the global life table, and then by using this expanded global life table to estimate the effect of child mortality on fertility. As already mentioned, the expanded global life table has five dimensions: age, parity, duration in parity, and two dimensions representing child mortality. The child mortality dimension is indicated by two indices,  $m$  and  $n$ , where  $m$  denotes number of child deaths at the start of the interval  $a-1$  to  $a$ , and  $n$  denotes child mortality state at the end of that interval (i.e., at age  $a$ ). These definitions of  $m$  and  $n$  effectively lag child mortality by one year of age in the CLL model for any given parity transition. This one-year lag in child mortality

minimizes the problem of reverse-causality from fertility to child mortality, inasmuch as a birth to a woman of parity  $i$  between ages  $a$  and  $a+1$  cannot have much of an effect on whether any of her previous children died one or more years ago when she was one or more years younger. (The fact that she was pregnant during part of the previous year may have had some effect on her care of previous children and therefore on their risk of mortality, however. In this paper this pregnancy-related risk is ignored.) This extension of the  $P_{ait}$  method is accordingly referred to here as the  $P_{aitmn}$  method.

As in the case of the  $P_{ait}$  method, the  $P_{aitmn}$  method yields PPRs, ASFRs, TFR, mean and median ages at childbirth by child's birth order (or by woman's parity just before she had the birth, when woman's parity was one less than the child's birth order) and for all births combined, and mean and median closed birth intervals. It additionally yields mean number of child deaths per woman, child mortality rates by age of woman (not by age of child, who can be of any age at death), and mean and median ages of woman at child death. The methodology also yields a replacement rate  $R$ , which measures the extent to which parents replace dead children. More specifically,  $R$  is calculated as the model-predicted increment in the TFR generated by a one-child increase in the average number of child deaths experienced by women by the end of the reproductive age span (i.e., by age 50). Since the replacement rate is calculated from a global life table that is multivariate, the replacement rate is also multivariate.

The  $P_{aitmn}$  method, like the  $P_{it}$  and  $P_{ait}$  methods, is based on a set of underlying discrete-time survival models, one for each parity transition. The difference is that, for any specified value of  $i$ , the dependent variable is  $P_{aitmn}$  instead of  $P_{it}$  or  $P_{ait}$ , and the set of predictor variables additionally includes the two child mortality variables mentioned earlier. As in the earlier papers on the  $P_{it}$  and  $P_{ait}$  methods, the discrete-time survival model that is used is the complementary log-log, CLL, model.

Figure 1 depicts graphically the transitions of women along the new global life table's multiple dimensions, for each of the first three parity transitions 1–2, 2–3, and 3–4, starting from any particular age  $a$ . (Child mortality does not come into play in the 0–1 transition, because there are as yet no previous births.) Duration in parity  $t$  starts at 0.<sup>3</sup>  $M_m$  denotes child mortality state (irrespective of child's age at death) for a woman age  $a$  and duration  $t$ . Subscript  $m$  refers to number of dead children at the start of the previous age interval ( $a-1$  to  $a$ ).

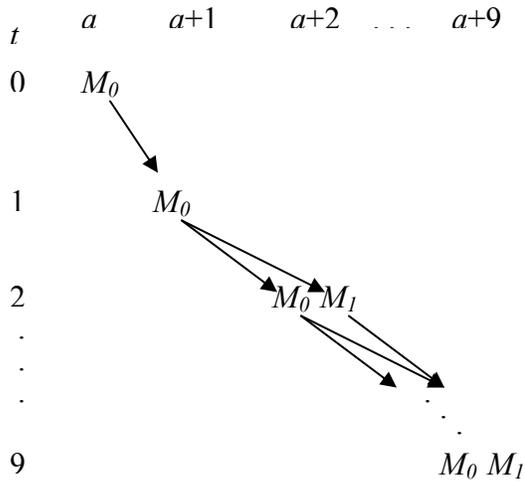
Panel 1 of Figure 1 portrays the 1–2 parity transition. The logic of this panel and of the  $P_{aitmn}$  method in general requires that a woman cannot experience more than one child death in any given year. But some actually do. This problem is handled in the construction of the data set in the following 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

---

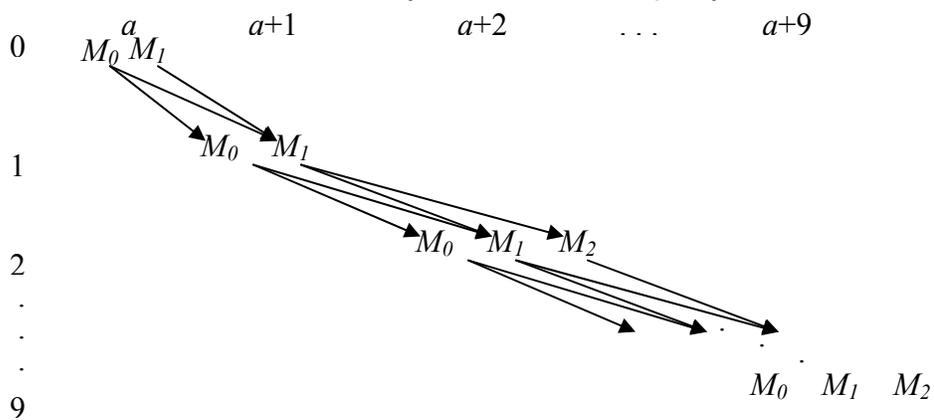
<sup>3</sup> In a CLL model for a particular parity transition,  $t$  starts at 1, which is later translated to zero when computing the global life tables.

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

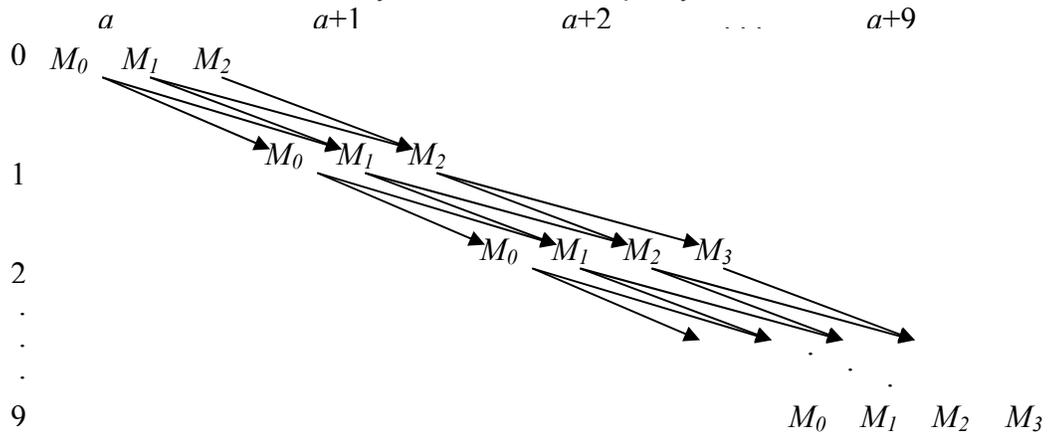
Panel 1: Possible child mortality transitions within parity transition 1-2



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>4</sup>.

<sup>4</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.

moved to the following two years, with one child death assigned to each of those two years.

Consider a woman in Panel 1 who starts in state  $M_0$  at age  $a$  duration 0 (i.e.,  $t = 0$ ). At age  $a+1$  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, so that  $M_0$  at age  $a+1$  corresponds to number of previous child deaths at age  $a$ . A woman in the  $M_0$  state at age  $a+1$  duration 1 can continue in that state until age  $a+2$  duration 2 if she does not experience a child death between age  $a$  duration 0 and age  $a+1$  duration 1. Alternatively, the woman can experience a child death between age  $a$  duration 0 and age  $a+1$  duration 1, in which case she moves from  $M_0$  to  $M_1$  between age  $a+1$  duration 1 and age  $a+2$  duration 2. 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 sequence is truncated at duration 10 because it is assumed that if the woman has not had a next birth by duration 10, she will never have a next birth. The logic of Panels 2 and 3 is similar, the difference being that the number of possible child mortality transitions increases as parity increases.

The basic predictor variables in the underlying CLL model for a particular parity transition  $i$  to  $i+1$  are woman's age  $A$  at starting parity and woman's duration in parity, where duration in parity is specified sometimes by  $t$  and sometimes by a set of dummy variables  $T_1, T_2, \dots, T_9$  (with the tenth interval as the reference category), indicating particular one-year age intervals. 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 as of one year ago (at age  $a-1$ ) when the woman was one year younger) and  $D_2$  (number of previous child deaths at age  $a$ ). The difference  $D_2 - D_1$  is either 0 or 1. Additional predictor variables are socioeconomic characteristics of interest.

The CLL model for parity transitions higher than 0–1, with residence and education as examples of socioeconomic predictors (the full set of socioeconomic predictors is shown later), is:

$$\begin{aligned}
P = 1 - \exp\{ & -\exp[a + b_1T_1 + b_2T_2 + \dots + b_9T_9 + A(c_1+c_2t+c_3t^2) + A^2(d_1+d_2t+d_3t^2) \\
& + D_1(e_1+e_2t+e_3t^2) + D_2(f_1+f_2t+f_3t^2) + U(g_1+g_2t+g_3t^2) + L(j_1+j_2t+j_3t^2) \\
& + H(k_1+k_2t+k_3t^2) + D_1(m_1A + m_2A^2) + D_2(n_1A + n_2A^2) + U(o_1A + o_2A^2) \\
& + L(q_1A + q_2A^2) + H(r_1A + r_2A^2) + D_1(s_1U + s_2L + s_3H) + vUL + wUH]\} \quad (1)
\end{aligned}$$

where  $P$  is the predicted probability of a next birth (also called the discrete hazard of next birth) in a one-year interval;  $T_1, T_2, \dots, T_9$  are dummy variables representing the first nine life table time intervals (with the tenth interval as the reference category);  $t$  is a counter variable (equal to 1, 2, ..., 10) that also denotes duration interval;  $A$  is age at starting parity;  $D_1$  and  $D_2$  represent child mortality as defined earlier;  $U$  is urban/rural residence (1 if urban, 0 if rural);  $L$  and  $H$  denote low and high education, with illiterate as the

reference category; and  $b_i, c_i, d_i, e_i, f_i, g_i, j_i, k_i, m_i, n_i, o_i, q_i, r_i, s_i, v,$  and  $w$  are coefficients, along with the intercept  $a$ , to be fitted to the data. Regarding the variables  $T_i$ , duration interval 1 (corresponding to  $t=1$ ) is specified by  $T_1 = 1$  and  $T_2 = T_3 = \dots = T_9 = 0$ . Duration interval 2 (corresponding to  $t=2$ ) is specified by  $T_1 = 0, T_2 = 1,$  and  $T_3 = \dots = T_9 = 0$ . And so on, up to duration interval 10 (the last interval), which is specified by  $T_1 = T_2 = \dots = T_9 = 0$ . In this equation,  $A$  ranges from 10 to 49 and  $t$  ranges from 1 to 10. Births occurring at a duration of more than 10 years are negligible in number and are ignored.

The model equation for the 0–1 transition is similar, except that  $t$  ranges from 1 to 30 (corresponding to ages 10–40), the set of  $T_i$  variables is expanded to  $T_1, T_2, \dots, T_{29}$ , and the age variable  $A$  and the child mortality variables  $D_1$  and  $D_2$  are dropped. First births below age 10 (if any) are re-coded to age 10 (the age at which the global life table begins, in the case of our Indian application). First births after age 40 are negligible in number and are ignored.<sup>5</sup>

The rationale for the specification of the predictor variables is much the same as in the earlier papers (Retherford et al. 2009, 2010a, 2010b). An  $A^2$  term is included as well as an  $A$  term in the set of predictor variables, because the rise and fall of fecundability as age increases suggest that the effect of starting age on parity progression will be non-linear, and that a quadratic specification of starting age will adequately capture this non-linearity. The effects of both  $A$  and  $A^2$  are specified as time-varying (i.e.,  $t$ -varying) because the effects of duration in parity on parity progression change as starting age increases, due not only to biological influences (changing fecundability) but also to behavioral influences. An example of a behavioral influence is that couples are more likely to settle into a life style with few or no children the longer they delay marriage and childbearing (Retherford et al. 2010b).

Except for the variable  $D_2$ , the set of predictor variables in equation (1) includes all two-way interactions among the predictor variables in order to achieve a more flexible fit to the data. In the case of  $D_1$  and  $D_2$ , we want to allow that the effects of both  $D_1$  and  $D_2$  can be non-linear over age and duration within a particular parity transition. Therefore, both  $D_1$  and  $D_2$  are both interacted with  $t, t^2, A,$  and  $A^2$ . In the case of the socioeconomic predictors, however, it is necessary to interact only  $D_1$  with the socioeconomic predictors. The reason is that  $D_1$  and  $D_2$  together measure a woman's child mortality trajectory over a one-year age or duration interval, and the aim is to interact this trajectory with each of the socioeconomic predictors. Only one of the variables  $D_1$  and  $D_2$  is needed to do this. We choose  $D_1$  over  $D_2$  because, for any particular woman, the value of  $D_1$  at the start of the interval is the same regardless of whether she experiences a child death during the interval.

---

<sup>5</sup> In the expanded data set for the 1-2 transition, however, first births after age 40 (up to a maximum of 49) are included in the set of starting events. Likewise, in the expanded data sets for the 2-3 and higher-order transitions, all next births at durations 0-9, regardless of woman's age, are included in the set of terminal events, and all births of the specified birth order, regardless of woman's age or duration in parity, are included in the set of starting events.

Model-predicted values of  $P_{Aitmn}$  for specified values of the socioeconomic predictors are obtained by indicating the value of  $i$  for the parity transition under consideration and substituting appropriate values of  $A$ ,  $t$ ,  $T_1$ , ...,  $T_9$ ,  $D_1$ , and  $D_2$  into fitted equation (1).

At higher parities, where sample sizes get smaller and non-convergence may occur, a quadratic specification of duration is used throughout:

$$\begin{aligned}
P = 1 - \exp\{ & -\exp[a + b_1t + b_2t^2 + A(c_1+c_2t+c_3t^2) + A^2(d_1+d_2t+d_3t^2) \\
& + D_1(e_1+e_2t+e_3t^2) + D_2(f_1+f_2t+f_3t^2) + U(g_1+ g_2t+ g_3t^2) + L(j_1+j_2t+j_3t^2) \\
& + H(k_1+ k_2t+ k_3t^2) + D_1(m_1A + m_2A^2) + D_2(n_1A + n_2A^2) + U(o_1A + o_2A^2) \\
& + L(q_1A + q_2A^2) + H(r_1A + r_2A^2) + D_1(s_1U + s_2L + s_3H) + vUL + wUH]\} \quad (2)
\end{aligned}$$

Non-convergence mainly occurs when one or more of the four cells in the 2x2 cross-classification of the individual-level dichotomous dependent variable FAILURE (1 if failure (i.e., the terminating event occurred), 0 otherwise) against any of the dichotomous predictor variables is empty (Allison, 1995).

Eventually, as one proceeds to higher parity transitions, even equation (2) will not converge to a solution. In this case, one needs to use an open-ended parity transition (e.g., 10+ to 11+), for which all person-year observations in the expanded data sets for the single-parity transitions comprising the open-ended parity transition (e.g., 10–11, 11–12, 12–13, etc.) are pooled. One first tries the following CLL model for this open-ended parity transition:

$$\begin{aligned}
P = 1 - \exp\{ & -\exp[a + b_1T_1 + b_2T_2 + \dots + b_9T_9 + A(c_1+c_2t+c_3t^2) + A^2(d_1+d_2t+d_3t^2) \\
& + D_1(e_1+e_2t+e_3t^2) + D_2(f_1+f_2t+f_3t^2) + U(g_1+ g_2t+ g_3t^2) + L(j_1+j_2t+j_3t^2) \\
& + H(k_1+ k_2t+ k_3t^2) + D_1(m_1A + m_2A^2) + D_2(n_1A + n_2A^2) + U(o_1A + o_2A^2) \\
& + L(q_1A + q_2A^2) + H(r_1A + r_2A^2) + D_1(s_1U + s_2L + s_3H) + vUL + wUH + x_1I \\
& + x_2I^2]\} \quad (3)
\end{aligned}$$

where  $I$  denotes a woman's parity at the start of the year within a person-year observation (before any birth that may have occurred within that person-year observation). If, because of convergence problems, this equation cannot be estimated, one then replaces the dummy variables  $T_1$ , ...,  $T_9$  with  $t$  and  $t^2$  in equation (3) and tries to fit that model. If that does not work, one has to broaden the open parity interval to include one more lower parity transition (e.g., 9–10, so that the open parity interval becomes 9+ to 10+).

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); 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  $a$  rather than  $a$  and  $a+1$ .) The probabilities  $P_{aitmn}$  for a particular set of values of the socioeconomic characteristics allow calculation of a global life table (discussed in the next section), from which one can calculate, for the specified values of the socioeconomic characteristics, not only model-predicted values of TFR and its components, but also child mortality measures and a replacement rate indicating the extent to which parents replace children who have died.

Equations (1) – (3) can also be used to construct unadjusted and adjusted probabilities  $P_{aitmn}$  by socioeconomic characteristics (such as residence and education). Unadjusted probabilities  $P_{aitmn}$  by values or categories of a particular socioeconomic variable are obtained by omitting all other socioeconomic variables from the equations. Adjusted probabilities by values or categories of a particular socioeconomic variable are obtained by retaining not only this particular socioeconomic variable as the main socioeconomic variable of interest but also all the other socioeconomic variables, which are controlled by holding them constant at their duration-specific mean values in the particular parity transition under consideration when varying the values or categories of the main socioeconomic variable. The procedure for doing this is described in more detail by Retherford et al. (2010b).

The unadjusted and adjusted probabilities  $P_{aitmn}$  are then used to calculate unadjusted and adjusted global life tables that incorporate child mortality. The adjusted global life table is multivariate in all the socioeconomic variables as well as the demographic variables corresponding to  $a$ ,  $t$ ,  $m$ , and  $n$  in the underlying CLL models for the various parity transitions. As already mentioned, this implies that all fertility and child mortality measures calculated from the adjusted global life table, including the child replacement measures, are also multivariate, as will be shown in the following section.

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

The model-predicted estimates of  $P_{aitmn}$  are the basic building blocks for constructing the global life table that incorporates child mortality. This global life table is sometimes referred to here in shorthand notation as the “GLT<sub>WCM</sub>,” where subscript WCM denotes “with child mortality.”  $S_{aitmn}$  and  $f_{aitmn}$  are additional building blocks.  $S_{aitmn}$  denotes number of “survivors” (women), and  $f_{aitmn}$  denotes the number of failures (births) 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 again that, because the child mortality variables are lagged by one year, these two ages are actually  $a-1$  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  $M_{m+1}$ .

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 they experienced a child death. This requires estimation of a partitioning factor, denoted by  $Q_{aitm}$ , which indicates the proportion who experience a child death over the interval. The proportion who do not experience a child death is then

$1-Q_{aitm}$ . The procedure for calculating the partitioning factor is explained later. For the moment, we shall assume that the values of  $Q_{aitm}$  have been calculated and are known.

With the partitioning factors defined and calculated, it is straightforward to construct a global life table that includes child mortality. 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 global life table are basically the same as in the global life table derived by the earlier  $P_{ait}$  method, since there is no previous child mortality in the case of 0–1 transition.

Equations (1) – (3) generate values of  $P_{Aitmn}$  for specified values of the socioeconomic variables. After these probabilities are generated, then, as explained by Retherford et al. (2010b), in order to simplify the global life table equations that come later, the variables  $A$  and  $t$  are re-labeled by replacing  $A$  with  $a = (A-10)+(t-1) = A+t-11$ , and by replacing  $t$  with  $t-1$ . Duration in parity  $t$  then starts at 0 instead of 1 (consistent with conventional demographic notation), and age in the global life table starts at 0 instead of 10. As an example of the re-labeling of  $A$  and  $t$ , the probability  $P_{Aitmn} = P_{27,4,3,1,1}$  is re-labeled as  $P_{aitmn} = P_{19,4,2,1,1}$ .

In the global life table, unlike the underlying CLL equations, duration in parity can be higher than 30 years for the 0–1 transition, and higher than 10 years for higher-order transitions, because everyone is followed from age 0 to age 40 (previously age 10 to age 50). This is handled in the global life table by setting  $P_{a,0,t,m,n} = 0$  when  $t > 29$ , and by setting  $P_{aitmn} = 0$  when  $i > 0$  and  $t > 9$ .

In sum, the ranges of  $a$ ,  $i$ , and  $t$  in the global life table equations shown below are the following: Age  $a$  ranges from 0 to 39 (prior to re-labeling, ages 10 to 49). Duration in parity  $t$  ranges from 0 to 39 in the 0–1 transition (but from 1 to 30 in the underlying CLL model for the 0–1 transition), and from 0 to 39 in each higher-order transition (but from 1 to 10 in the underlying CLL model for each higher-order transition).

Formulae for the 0–1 transition are then:

$$S_{0,0,0,0,0} = 1,000 \quad (4)$$

$$S_{a,0,t,0,0} = S_{a,0,a,0,0} = S_{a-1,0,a-1,0,0} (1 - P_{a-1,0,a-1,0,0}) \quad \text{for } a > 0 \quad (5)$$

$$f_{a,i,t,0,0} = S_{a,i,t,0,0} P_{a,i,t,0,0} \quad (6)$$

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, except that they are more elaborate because they incorporate non-zero child mortality. As already discussed, Panel 1 of Figure 1 for the 1–2 parity transition shows how women progress by age, duration in parity, and child mortality state. Panel 1

of Figure 2 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.

Formulae corresponding to Panel 2 of Figure 2 for the 1–2 transition are:

**If  $t=0$ :**

$$S_{a,1,0,0,0} = \sum (S_{a,0,t,0,0} P_{a,0,t,0,0}) = \sum f_{a,0,t,0,0} \quad \text{where the summation is over } t \quad (7)$$

$$S_{a,1,0,0,1} = 0 \quad (8)$$

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

**If  $t = 1$ :**

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

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

$$S_{a,1,1,1,1} = 0 \quad (12)$$

**If  $t = 2$ :**

$$S_{a,1,2,0,0} = S_{a-1,1,1,0,0} (1 - P_{a-1,1,1,0,0}) (1 - Q_{a,1,2,0}) \quad \text{for } a > 0 \quad (13)$$

$$S_{a,1,2,0,1} = S_{a-1,1,1,0,0} (1 - P_{a-1,1,1,0,0}) Q_{a,1,2,0} \quad \text{for } a > 0 \quad (14)$$

$$S_{a,1,2,1,1} = S_{a-1,1,1,0,1} (1 - P_{a-1,1,1,0,1}) \quad \text{for } a > 0 \quad (15)$$

**If  $t > 2$ :**

$$S_{a,1,t,0,0} = S_{a-1,1,t-1,0,0} (1 - P_{a-1,1,t-1,0,0}) (1 - Q_{a,1,t,0}) \quad \text{for } a > 0 \quad (16)$$

$$S_{a,1,t,0,1} = S_{a-1,1,t-1,0,0} (1 - P_{a-1,1,t-1,0,0}) Q_{a,1,t,0} \quad \text{for } a > 0 \quad (17)$$

$$S_{a,1,t,1,1} = (S_{a-1,1,t-1,0,1}) (1 - P_{a-1,1,t-1,0,1}) + (S_{a-1,1,t-1,1,1}) (1 - P_{a-1,1,t-1,1,1}) \quad \text{for } a > 0 \quad (18)$$

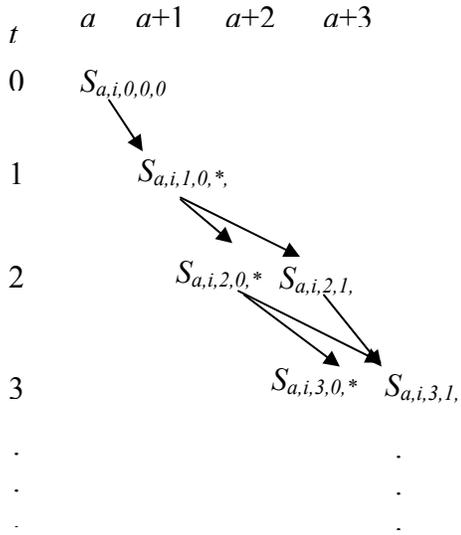
$$f_{aitmn} = S_{aitmn} P_{aitmn} \quad (19)$$

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

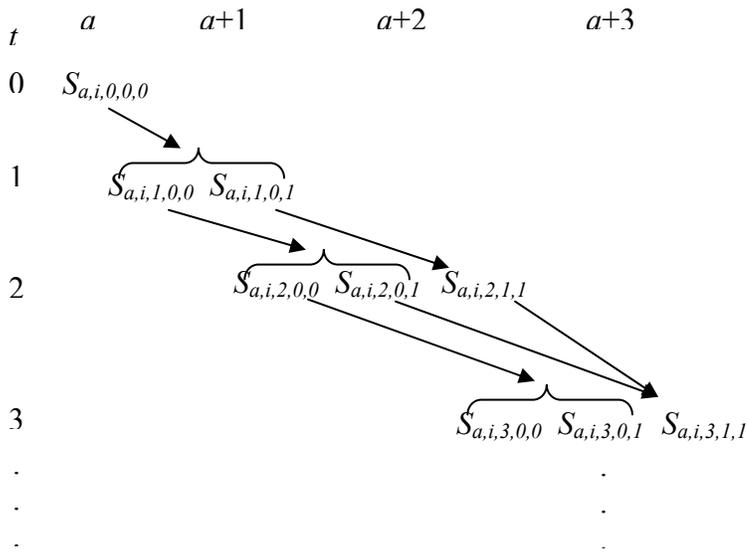
The formulae for parity transition 2–3 ( $i = 2$ ) provide an example. In order to clarify these formulae, Panel 2 of Figure 1 is elaborated in Figure 3 to show how women

Figure 2: Elaboration of panels 1 of Figure 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  $t > 0$ ,  $S_{a,i,t,m^*}$  in panel 1 is 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) can be either  $m$  or  $m+1$ .

are partitioned into two groups at each age, parity, duration in parity, and child mortality state.

**If  $t = 0$ :**

$$S_{a,2,0,0,0} = \left( \sum (S_{a,1,t,0,0} P_{a,1,t,0,0}) \right) (1 - Q_{a,2,0,0}) \quad \text{where the summation is over } t \quad (20)$$

$$S_{a,2,0,0,1} = \left( \sum (S_{a,1,t,0,0} P_{a,1,t,0,0}) \right) Q_{a,2,0,0} \quad \text{where the summation is over } t \quad (21)$$

$$S_{a,2,0,1,1} = \sum (S_{a,1,t,0,1} P_{a,1,t,0,1}) + \sum (S_{a,1,t,1,1} P_{a,1,t,1,1}) \quad \text{where the summation is over } t \quad (22)$$

$$S_{a,2,0,1,2} = 0 \quad (23)$$

$$S_{a,2,0,2,2} = 0 \quad (24)$$

**If  $t = 1$ :**

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

$$S_{a,2,1,0,1} = S_{a-1,2,0,0,0} (1 - P_{a-1,2,0,0,0}) Q_{a,2,1,0} \quad \text{for } a > 0 \quad (26)$$

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

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

$$S_{a,2,1,2,2} = 0 \quad (29)$$

**If  $t = 2$ :**

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

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

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

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

$$S_{a,2,2,2,2} = S_{a-1,2,1,1,2} (1 - P_{a-1,2,1,1,2}) \quad \text{for } a > 0 \quad (34)$$

**If  $t > 2$ :**

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

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

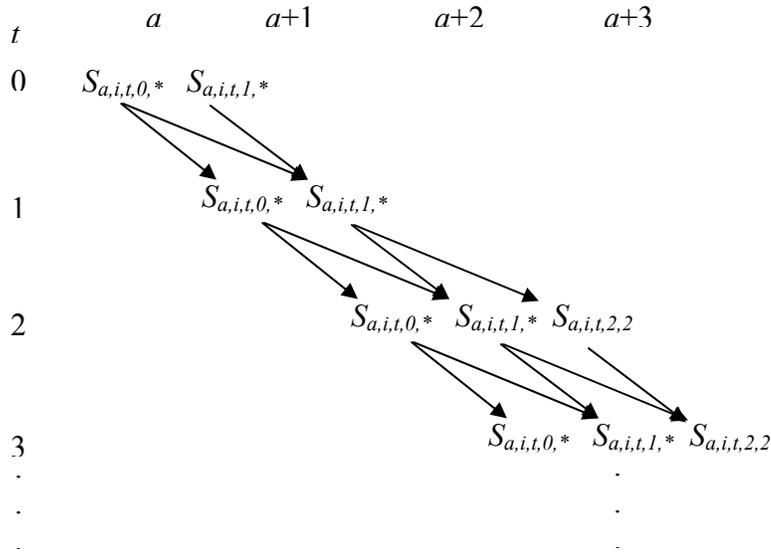
$$S_{a,2,t,1,1} = (S_{a-1,2,t-1,0,1} (1 - P_{a-1,2,t-1,0,1}) + S_{a-1,2,t-1,1,1} (1 - P_{a-1,2,t-1,1,1})) (1 - Q_{a,2,t,1}) \quad \text{for } a > 0 \quad (37)$$

$$S_{a,2,t,1,2} = (S_{a-1,2,t-1,0,1} (1 - P_{a-1,2,t-1,0,1}) + S_{a-1,2,t-1,1,1} (1 - P_{a-1,2,t-1,1,1})) Q_{a,2,t,1} \quad \text{for } a > 0 \quad (38)$$

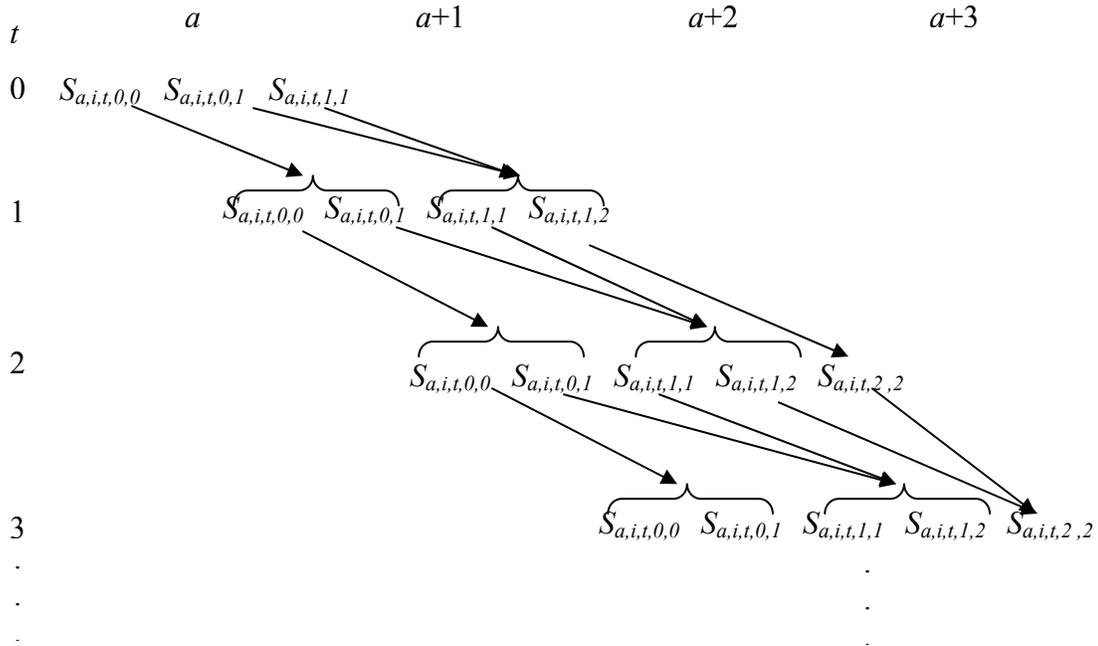
$$S_{a,2,t,2,2} = (S_{a-1,2,t-1,1,2} (1 - P_{a-1,2,t-1,1,2}) + (S_{a-1,2,t-1,2,2} (1 - P_{a-1,2,t-1,2,2})) \quad \text{for } a > 0 \quad (39)$$

Figure 3: Elaboration of panel 2 of Figure 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:  $S_{a,i,t,m^*}$  in panel 1 is 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 ending child mortality state can be  $m$  or  $m+1$ .

Formulae are based on the simplifying assumption that all births occur at the start of a one-year age or duration interval. This assumption allows more than one birth to occur at the same age  $a$  in the global life table. This is consistent with the way that the  $P_{aitmn}$  are estimated from separate CLL models for the various parity transitions, which makes it possible for a woman to experience multiple events in a one-year age interval (but not in a one-year duration interval, since  $t$  immediately changes to zero after the first event in the duration interval). The implications of this simplifying assumption are discussed in more detail in Retherford et al. (2010b).

Once the global life table is constructed, the total fertility rate is calculated as

$$\text{TFR} = (\sum f_{a,i,t,m,n})/1,000 \quad (40)$$

In this equation the summation is over  $a$ ,  $i$ ,  $t$ ,  $m$ , and  $n$ .

TFR can also be calculated by collapsing the right side of equation (40) over  $a$ ,  $t$ ,  $m$ , and  $n$ , resulting in a sum of parity-specific terms that can be written as

$$\text{TFR} = (B_0 + B_1 + B_2 + \dots + B_{14})/1,000 \quad (41)$$

where  $B_0$  pertains to first births (i.e., births to zero-parity women),  $B_1$  pertains to second births, and so on.

Parity progression ratios can then be calculated as

$$p_0 = B_0/1,000 \quad (42)$$

$$p_i = B_i/B_{i-1} \quad i = 1, 2, 3, \dots, 14 \quad (43)$$

where  $p_0$  is the PPR from woman's own birth to her first birth,  $p_1$  is the PPR from first birth to second birth, and so on.

ASFRs for single-year age groups,  $F_a$ , are calculated for each value of  $a$  by summing the births  $f_{aitmn}$  over  $i$ ,  $t$ ,  $m$ , and  $n$  and dividing the result by 1,000. If desired, ASFRs for 5-year age groups can be obtained by summing the single-year  $F_a$  within a five-year age group and dividing the sum by five.

One can calculate numbers of births specific for parity and duration in parity,  $B_{it}$ , by aggregating the  $f_{aitmn}$  over  $a$ ,  $m$ , and  $n$ . One then computes mean and median closed birth intervals by birth order as

$$\text{Mean closed birth interval between } i^{\text{th}} \text{ and } (i+1)^{\text{th}} \text{ birth} = \sum [(B_{it}/B_i)(t)] \quad (44)$$

where the summation is over  $t$ .

$$\text{Median closed birth interval between } i^{\text{th}} \text{ and } (i+1)^{\text{th}} \text{ birth} =$$

$$[t, \text{ such that } (\sum B_{it})/B_i = 0.5] \quad (45)$$

where the summation ranges over duration in parity from 0 to  $t$ .

Mean and median ages at childbearing are calculated as

$$\text{Mean age at childbearing} = \sum [(F_a/\text{TFR})(a)] + 10 \quad (46)$$

where the summation is over  $a$ .

$$\text{Median age at childbearing} = [a, \text{ such that } \sum (F_a)/\text{TFR} = 0.5] + 10 \quad (47)$$

where the summation ranges over age from 0 to  $a$ .

Median age at childbearing for women of parity  $i$  bearing an  $(i+1)^{\text{th}}$  birth =

$$[a, \text{ such that } (\sum B_{ai})/B_i = 0.5] + 10 \quad (48)$$

where the summation ranges over age from 0 to  $a$ .

It might seem that the formulae for mean closed birth interval and mean age at childbearing should add 0.5 to duration  $t$  or age  $a$ , since the formulae are for discrete one-year time intervals. It turns out, however, that  $t$  should be used instead of  $t+0.5$  and  $a$  should be used instead of  $a+0.5$ , because of the way that the duration and age variables are constructed (Retherford 2010a, 2010b).

### Partitioning Factor

A major difference between the original global life table (GLT) and the global life table that includes child mortality (GLT<sub>WCM</sub>), is that women in some mortality transitions of the new global life table are partitioned into two groups, the first of which does not experience a child death over the following one-year interval and the second of which does experience a child death over this interval. In order to partition women at the start of each group, one needs a partitioning factor, denoted earlier as  $Q_{aim}$ , defined as the proportion who experience a child death over the interval. The proportion who do not experience a child death is then  $1-Q_{aim}$ . This section explains how  $Q_{aim}$  is calculated.

Two approaches for calculating  $Q_{aim}$  were tested. For ease of exposition, these two approaches are referred to here as the “overall probability” approach and the “logit model” approach. The overall probability approach uses the overall probability of failure (i.e., next birth) at age  $a$ , parity  $i$ , duration  $t$ , and child mortality state  $m$ , as part of the estimation procedure. The logit model approach regresses change in mortality state over the interval (yes or no) on age, parity, duration in parity, child mortality state at the start of the interval, and the socioeconomic variables. The fitted logit model yields a predicted proportion of women who experience a child death over the interval for specified values

of the predictor variables. For reasons that will become clear shortly, the overall probability approach is preferred over the logit model approach.

**“Overall Probability Approach” to Calculating the Partitioning Factor**

$S_{aitm^*}$  indicates number of women at age  $a$ , parity  $i$ , duration in parity  $t$ , and child mortality state  $m$  in the new global life table for a particular set of values of the socioeconomic characteristics. The asterisk indicates that child mortality state at the end of a one-year transition can be either  $m$  or  $m+1$ .  $S_{aitm^*}$  is partitioned into two groups, the first of which remains 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. (Recall again that child mortality is a lagged variable, so that the age interval is actually  $a-1$  to  $a$ .)

By way of illustration, suppose that there are 100 women age  $a$ , parity  $i$ , and duration  $t$  in child mortality state 2 ( $M_m = M_2$ ) at the start of a one-year interval  $a$  to  $a+1$  and  $t$  to  $t+1$ . I.e.,  $S_{aitm^*} = S_{a,i,t,2,*} = 100$ . During the interval  $a$  to  $a+1$  and  $t$  to  $t+1$ , 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  $S_{a,i,t,2,2}$ . The probability of having a birth for a woman in 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  $M_3$ , is denoted  $S_{a,i,t,2,3}$ . The probability of having a next birth for a woman in this group is  $P_{a,i,t,2,3}$ . It follows that  $S_{a,i,t,2,*} = S_{a,i,t,2,2} + S_{a,i,t,2,3}$ . The partitioning factor  $Q_{a,i,t,2}$  (to be calculated) indicates the proportion of  $S_{a,i,t,2,*}$  who end up in  $S_{a,i,t,2,3}$ .

The issue, then, is how to divide the  $S_{a,i,t,2,*}$  women into the two groups  $S_{a,i,t,2,2}$  and  $S_{a,i,t,2,3}$ . Define the overall probability of transition  $P_{aitm\bar{n}}$ , where  $\bar{n}$  is the average child mortality state (mean number of previous child deaths as of one year ago) at age  $a$  and duration  $t$  (actually age  $a-1$  and duration  $t-1$ ).  $\bar{n}$  is always equal to or greater than  $m$  and equal to or less than  $n$ . The calculation is best explained by an example, depicted in Figure 4, where  $\bar{n}$  is a number between 2 and 3.

The following equations are generated from the figure:

$$S_{a,i,t,2,*} = S_{a,i,t,2,2} + S_{a,i,t,2,3} \quad (49)$$

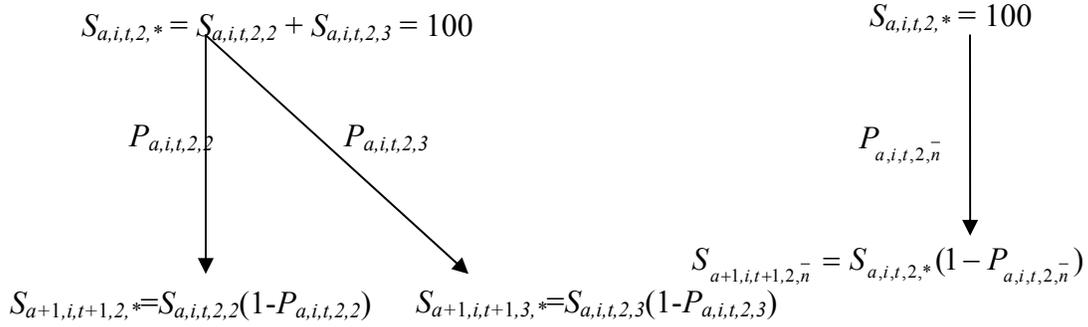
$$S_{a+1,i,t+1,2,*} = S_{a,i,t,2,2}(1-P_{a,i,t,2,2}) \quad (50)$$

$$S_{a+1,i,t+1,3,*} = S_{a,i,t,2,3}(1-P_{a,i,t,2,3}) \quad (51)$$

It is assumed that

$$S_{a+1,i,t+1,2,\bar{n}} = S_{a+1,i,t+1,2,*} + S_{a+1,i,t+1,3,*} \quad (52)$$

Figure 4: Logic of the “overall probability approach” for calculating the partitioning factor



Equation (52) states 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 (49) and (52) are solved simultaneously, using elements from equations (50) and (51), the numbers of women  $S_{a,i,t,2,2}$  and  $S_{a,i,t,2,3}$  are estimated as

$$S_{a,i,t,2,2} = \frac{S_{a,i,t,2,*}(P_{a,i,t,2,3} - P_{a,i,t,2,\bar{n}})}{(P_{a,i,t,2,3} - P_{a,i,t,2,2})} \quad (53)$$

$$S_{a,i,t,2,3} = \frac{S_{a,i,t,2,*}(P_{a,i,t,2,2} - P_{a,i,t,2,\bar{n}})}{(P_{a,i,t,2,2} - P_{a,i,t,2,3})} \quad (54)$$

From equation (54) it is evident that the partitioning factor can be calculated as

$$Q_{a,i,t,2} = \frac{P_{a,i,t,2,2} - P_{a,i,t,2,\bar{n}}}{P_{a,i,t,2,2} - P_{a,i,t,2,3}} \quad (55)$$

$P_{aitmn}$  is estimated by substituting values of age  $A$  (where  $A = a-t+11$ ) 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 for age  $a$ , parity  $i$ , duration  $t$ , starting child mortality state  $m$ , and values of the socioeconomic variables. For example, to estimate  $P_{25,3,4,2,\bar{n}}$ , one substitutes  $A = 31$ ,  $i = 3$ ,  $t = 5$ ,  $D_1 = 2$ , and  $D_2$  equal to the computed value of  $\bar{n}$  into the appropriate CLL equation ((1), (2), or (3)). Since the starting child mortality state in this example is 2, the mean of  $D_2$ , specified by age, parity, starting child mortality state, and values of the socioeconomic variables, is a number between 2 and 3.

Formula (55) works well as long as the two probabilities in the denominator of the right side of equation (55) are not equal. When equal, the formula is undefined, so that

one needs to find another solution to partition women into two groups. The logit model approach is used to do this.

### **“Logit Model Approach” to Calculating the Partitioning Factor**

Another way to partition women into two groups is the logit model approach. In this approach, for any given parity transition, one uses the same data set that is used to estimate the CLL model for that transition. The individual-level response variable in the logit regression is a dummy variable indicating whether a child death occurs in the following one-year interval (actually the previous one-year interval, since the child mortality variables are lagged). This dummy variable, denoted  $\Delta D$  (equal to  $D_2 - D_1$ ), is defined for each person-year observation in the data set. Predictor variables in the logit model include age, duration in parity, starting child mortality  $D_1$ , and socioeconomic variables, if any. The logit model is estimated for each parity transition higher than 0–1, where child mortality can occur.

The probability form of the logit model is:

$$P_{\Delta D} = \frac{1}{1 + e^{-(a+bt+ct^2+dA+eA^2+fD_1+gSES)}} \quad (56)$$

where  $P_{\Delta D}$  is the probability of a child death during the one-year interval,  $SES$  is a column vector of socioeconomic predictors, and  $g$  is a row vector of coefficients.  $Q_{aitm}$  for particular values of  $a, i, t, m$ , and the socioeconomic variables equals the model-predicted value of  $P_{\Delta D}$  for specified values of these same variables. The value of  $i$  is determined by which parity-specific CLL model is under consideration.

The overall probability approach is preferred over the logit model approach, because, in the former approach, all probabilities needed to calculate the partitioning factor are derived from the underlying CLL models, so that the  $Q_{aitm}$  values and the various probabilities used in calculating the global life table are internally consistent. Because of this internal consistency, tests on real data indicate that the overall probability approach works better than the logit model approach for estimating the partitioning factor (Eini-Zinab 2010a). For these reasons, the logit model approach is used here only when the denominator of the right side of equation (55) is zero. In almost all cases it is non-zero, so that the logit model approach is seldom used in any given application to real data.

### **Estimating Child Mortality Rates from the Global Life Table**

The global life table that incorporates child mortality allows calculation of child mortality rates. 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 one-year-lagged child death) can occur. This allows calculation of child mortality rates by age of woman in place of the more usual mortality rates by age of child. Child deaths can be at any age of child, no matter how old.

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 child mortality rates by age of woman and the overall mean number of child deaths per woman:

$$D_a = (\sum S_{aitmn})/1,000 \quad (57)$$

where  $D_a$  (not to be confused with the variables  $D_1$  and  $D_2$  in the underlying CLL models) is the child mortality rate by woman's age  $a$ , with child deaths lagged by one year. The summation is over  $i$ ,  $t$ ,  $m$ , and  $n$ , but only for cases where  $n = m + 1$ . The denominator of this single-year rate is always 1,000, regardless of the value of  $a$ , because although the global life table allows for child mortality, it assumes no mortality of women.

If the summation in equation (57) is also over  $a$ , the result is total number of child deaths per woman:

$$D = (\sum S_{aitmn})/1,000 \quad (58)$$

The latter of these two child mortality measures is affected by the level of fertility as well as by conventionally measured child mortality by age of child. The reason is that the more children a woman has, the more likely she is to experience a child death; thus the higher the level of fertility, the higher the value of  $D$ .

### **Estimating the Extent of Replacement of Dead Children from the Global Life Table**

One can construct a global life table using the  $P_{aitmn}$  approach by 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, so that it is always true that  $S_{a,i,t,m,n} = S_{a,i,t,0,0}$ . This global life table is referred to here as the child-mortality-effect-free global life table, denoted as  $GLT_{NCM}$ . (The word "effect" is retained because we are ultimately interested in the effect of child mortality on fertility.) The subscript NCM stands for "no child mortality effect". The child-mortality-effect-free GLT is constructed by using  $P_{a,i,t,0,0}$  estimated from CLL models that include the variables  $D_1$  and  $D_2$ , with non-zero child mortality taken into account.

All measures derived from the  $GLT_{NCM}$  are also child-mortality-effect-free. All of these measures can then be compared with corresponding measures calculated from the global life table that includes child mortality, in which the partitioning factors are not set to zero. The global life table that includes child mortality is referred to here as the child-mortality-effect-present GLT (denoted as  $GLT_{WCM}$ , where the subscript WCM stands for "with child mortality effect"). Any difference between the two GLTs is attributable to the effect of child mortality on fertility.

Several replacement measures can be constructed. One of them is

$$\Delta TFR = TFR_{WCM} - TFR_{NCM} \quad (59)$$

$\Delta TFR$  depends on the magnitude of child mortality, as measured by mean number of child deaths per woman. The higher the mean number of child deaths per woman, the higher the  $\Delta TFR$ .  $\Delta TFR$  is the additive effect of all child deaths per woman on the TFR. A related measure that could be used is the proportional increase in TFR due to all child deaths, calculated as  $[(TFR_{WCM}-TFR_{NCM})/TFR_{NCM}]$ .

Our preferred measure, however, is what is referred to here as the “replacement rate,” denoted by  $R$ .  $R$  is defined as the difference between the child-mortality-effect-present TFR and the child-mortality-effect-free TFR, divided by the total number of child deaths per woman,  $D$ , in the global life table ( $GLT_{WCM}$ ):

$$R = \frac{\Delta TFR}{D} \quad (60)$$

$R$  indicates the effect of a one-unit increase in number of child deaths per woman (all women, not just mothers) on the TFR. As such,  $R$  is a measure of the extent to which a dead child is replaced by another birth.

Because all of the above-mentioned replacement measures are derived from the global life table, and because global life table is multivariate, all the replacement measures are also multivariate, in the sense that they can be tabulated by values or categories of one socioeconomic variable while holding constant the other socioeconomic variables included in the underlying CLL models.

A seemingly illogical property of the global life table, in both the  $P_{ait}$  and  $P_{aitmn}$  methods, is that the total number of life table 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 a woman can experience more than one birth in any given one-year age interval. For reasons explained by Retherford et al. (2010b), however, this feature of the global life table does not bias the estimates of the fertility and child mortality measures in this paper.

## **VALIDATION AND APPLICATION OF THE $P_{aitmn}$ METHOD**

This section tests the validity of the  $P_{aitmn}$  method by comparing, where possible, selected fertility and child mortality measures derived by the  $P_{aitmn}$  method with corresponding estimates derived by the birth history method, the  $P_{it}$  method, and the  $P_{ait}$  method. The  $P_{aitmn}$  methodology is considered valid if, for real cohorts, the estimates derived by the various methods are close to the same.

This validation procedure is applied to not only cohort data but also period data from India’s first National Family Health Survey (NFHS-1), conducted in 1992–93. The cohort analysis pertains to the birth and child-mortality histories of all women age 45–49 at time of survey. The period analysis pertains to the fertility and child mortality experienced by all women age 10–49 during the five-year period preceding the survey

(1988 to 1992)<sup>6</sup>. In the case where socioeconomic variables are omitted from the underlying CLL models, it is expected that the estimates derived by the various methods will agree closely when based on cohort data. As will be explained later, it is not expected that the estimates derived by the various methods will necessarily be the same in the case of period data, however. Following validation of the methodology using NFHS-1 data, the methodology is applied more extensively to not only NFHS-1 data but also NFHS-2 and NFHS-3 data.

The socioeconomic predictors included in both the validation and the application of the  $P_{aitmn}$  method are the following: Residence is represented by a dummy variable *URB* (urban), with rural as the reference category. Education is represented by two dummy variables *LOW* (low education) and *HIG* (high education), with illiterate as the reference category. Low education comprises some primary or completed primary education, and high education comprises all levels of education beyond primary. Wealth, based on a wealth index constructed by Macro International (Rutstein and Johnson, 2004) is represented by two dummy variables, *MWI* (denoting medium wealth) and *HWI* (denoting high wealth), with low wealth as the reference category. Break points in the wealth index for low, medium, and high are chosen with the goal (imperfectly achieved) of having one-third of the women in each of the three categories. The break points vary from one survey to the next. Religion is represented by the dummy variable *HIN* (Hindu), with “other” as the reference category. When *HIN* is used as a control variable, “other” includes all non-Hindu women. When *HIN* is used as the main socioeconomic predictor variable, “other” includes Muslims only. Scheduled caste/scheduled tribe (SC/ST) is represented by the dummy variable *SCT* (yes, no), with “no” as the reference category. Region is a categorical variable which includes North, East, South, and West, represented by the dummy variables *NOR*, *EAS*, and *SOU*, with West as the reference category. North includes the states of Haryana, Himachal Pradesh, Jammu, Madhya Pradesh, Punjab, Rajasthan, Uttar Pradesh, New Delhi, Uttaranchal (only in NFHS-3), and Chhattisgarh (only in NFHS-3). East includes the states of Assam, Bihar, Manipur, Meghalaya, Mizoram, Nagaland, Orissa, West Bengal, Arunachal Pradesh, Tripura, Sikkim (only in NFHS-2 and NFHS-3), and Jharkhand (only in NFHS-3). South includes the states of Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu. West includes the states of Goa, Gujarat, and Maharashtra. Except for Sikkim, the states included only in NFHS-2 or NFHS-3 are new states that have been split off from existing states. Sikkim could not be included in NFHS-1 because of political disturbances.

Table 1 shows how sample women are distributed on the socioeconomic characteristics in each of the three surveys. The table shows increasing proportions with urban residence, high education, and SC/ST membership over the three surveys. The proportions by religion and region are fairly stable over the three surveys. The

---

<sup>6</sup> Calendar years in this analysis do not exactly match calendar years that run from January 1 to December 31. Instead, calendar years are demarcated by counting by increments of 12 months backward from the survey date for each respondent, starting with the first complete survey month for that respondent. In the case of the 1992-93 NFHS-1, for example, “1992” in the present analysis refers to the first 12 complete months preceding date of interview. The second 12 months preceding date of interview are labeled 1991, and so on. The month in which the interview took place is not included in the analysis because it is an incomplete month for most women.

proportions in the wealth categories are difficult to interpret because the wealth categories are defined somewhat differently in each survey, with the goal of keeping the proportion in each wealth category close to one-third.

Table 1: Percentaged distribution of sample women by socioeconomic variables included in the analysis

Predictor variable	Women age 45-49			Women age 10-49		
	NFHS-1	NFHS-2	NFHS-3	NFHS-1	NFHS-2	NFHS-3
<b>Residence</b>						
Urban	28	30	35	28	28	32
Rural	72	70	66	72	72	68
<b>Education</b>						
Illiterate	71	61	57	50	41	35
Low	16	17	16	26	23	22
High	13	22	26	24	36	43
<b>Wealth Index</b>						
Low	39	36	39	41	41	45
Medium	32	31	30	31	32	31
High	29	32	31	28	27	24
<b>Religion</b>						
Hindu	83	83	83	81	80	80
Other	17	17	17	19	20	20
<b>SC/ST</b>						
Yes	20	25	26	20	27	28
No	80	75	74	80	73	72
<b>Region</b>						
North	36	33	35	36	36	38
East	25	25	25	26	26	27
South	25	29	26	23	24	21
West	15	14	14	14	15	14
No. of women	8,592	8,661	9,807	144,626	149,016	153,348

Notes: Women 45-49 are the basis of the cohort analysis. Women 10-49 are the basis of the period analysis. The "other" category for religion includes all non-Hindu women. Numbers of missing cases are negligible except for the Scheduled Caste/Scheduled Tribe (SC/ST) variable in the NFHS-2 and NFHS-3 surveys; about 3 percent of women in these two surveys having missing information on the SC/St variable. The percentaged distributions exclude missing cases. The calculation of percentages incorporates individual-level weights as specified in each of the three surveys.

Table 2 shows the expanded person-year data sets. The expanded data sets for the cohort analysis were created from women age 45–49 at time of survey, and the expanded data sets for the period analysis were created from women age 10–49 at time of survey. The numbers of women from whom expanded data sets were created are shown in the

last row of Table 1. In the case of open parity intervals, some of the person-year data sets shown in Table 2 were aggregated, as explained earlier.

The CLL model for the  $P_{aitmn}$  method with all socioeconomic predictors for parity transitions higher than 0–1 is:

$$\begin{aligned}
P = 1 - \exp\{ & -\exp[a + b_1T_1 + b_2T_2 + \dots + b_9T_9 + A(c_1+c_2t+c_3t^2) + A^2(d_1+d_2t+d_3t^2) \\
& + D_1(e_1+e_2t+e_3t^2) + D_2(f_1+f_2t+f_3t^2) + URB(g_1+g_2t+g_3t^2) + LOW(h_1+h_2t+h_3t^2) \\
& + HIG(j_1+j_2t+j_3t^2) + HIN(k_1+k_2t+k_3t^2) + SCT(l_1+l_2t+l_3t^2) \\
& + MWI(m_1+m_2t+m_3t^2) + HWI(n_1+n_2t+n_3t^2) + NOR(o_1+o_2t+o_3t^2) \\
& + EAS(q_1+q_2t+q_3t^2) + SOU(r_1+r_2t+r_3t^2) + D_1(e_4A + e_5A^2) + D_2(f_4A + f_5A^2) \\
& + URB(g_4A + g_5A^2) + LOW(h_4A + h_5A^2) + HIG(j_4A + j_5A^2) + HIN(k_4A + k_5A^2) \\
& + SCT(l_4A + l_5A^2) + MWI(m_4A + m_5A^2) + HWI(n_4A + n_5A^2) + NOR(o_4A + o_5A^2) \\
& + EAS(q_4A + q_5A^2) + SOU(r_4A + r_5A^2) + D_1(s_1URB + s_2LOW + s_3HIG + s_4HIN \\
& + s_5SCT + s_6MWI + s_7HWI + s_8NOR + s_9EAS + s_{10}SOU) + (u_1URB.LOW \\
& + u_2URB.HIG + u_3URB.HIN + u_4URB.SCT + u_6URB.MWI + u_7URB.HWI \\
& + u_8URB.NOR + u_9URB.EAS + u_{10}URB.SOU + u_{11}LOW.HIN + u_{12}HIG.HIN \\
& + u_{13}LOW.SCT + u_{14}HIG.SCT + u_{15}LOW.MWI + u_{16}HIG.MWI + u_{17}LOW.HWI \\
& + u_{18}HIG.HWI + u_{19}LOW.NOR + u_{20}HIG.NOR + u_{21}LOW.EAS + u_{22}HIG.EAS \\
& + u_{23}LOW.SOU + u_{24}HIG.SOU + u_{25}HIN.SCT + u_{26}HIN.MWI + u_{27}HIN.HWI \\
& + u_{28}HIN.NOR + u_{29}HIN.EAS + u_{30}HIN.SOU + u_{31}MWI.NOR + u_{32}HWI.NOR \\
& + u_{33}MWI.EAS + u_{34}HWI.EAS + u_{35}MWI.SOU + u_{36}HWI.SOU)\} \quad (61)
\end{aligned}$$

where,  $URB, LOW, HIG, HIN, \dots, SOU$  are the socioeconomic predictors defined above and  $a, b_i, c_i, d_i, \dots, u_i$  are coefficients to be estimated. The CLL model for the 0–1 transition does not include  $A$  (age at starting parity) and  $D_1$  and  $D_2$  (the child mortality variables). The CLL model for the open parity interval is similar to equation (61), except that it includes the additional predictors  $I$  and  $I^2$ , as explained earlier. (The CLL models for the  $P_{ait}$  method are similar to the CLL models for the  $P_{aitmn}$  method, except that the CLL models for the  $P_{ait}$  method exclude the child mortality variables  $D_1$  and  $D_2$ . The CLL models for the  $P_{it}$  method exclude the variable  $A$  as well.)

One expects estimates derived by the  $P_{aitmn}$  method to agree closely with estimates derived by the other methods when the estimates are based on cohort data. When using period data, however, one is dealing with a synthetic cohort instead of a real cohort, which makes comparisons between estimates based on the birth history method and estimates based on the life table methods more complicated. When using period data, one expects the biggest differences to occur between estimates based on the birth history method and estimates based on either of the two global life table methods (the  $P_{ait}$  method and the  $P_{aitmn}$  method). The reason is that, in a global life table at any given age, the distribution of women by parity and duration in parity (in the  $P_{ait}$  method) or by parity, duration in parity, and the two child mortality states at the beginning and end of the age interval  $a$  to  $a+1$  (in the  $P_{aitmn}$  method) differs from the distribution of women by these variables in the actual population. Differences in age structure do not account for any of the differences, because the TFR does not depend on the age structure of the population, regardless of which method is used to calculate the TFR.

Table 2: Expanded sample sizes

Parity transition	Cohort Analysis			Period Analysis		
	NFHS			NFHS		
	1	2	3	1	2	3
0-1	102,787	104,924	124,947	316,308	327,129	276,090
1-2	33,596	33,421	38,979	69,399	69,493	75,315
2-3	34,747	35,872	44,941	73,472	77,659	86,021
3-4	33,946	34,850	40,226	63,108	61,953	59,190
4-5	30,992	29,823	29,615	45,562	42,701	36,238
5-6	24,580	21,682	18,913	30,871	26,995	22,628
6-7	18,608	15,917	11,983	20,041	17,136	13,451
7-8	12,814	9,868	7,067	12,433	10,172	8,100
8-9	7,863	6,086	4,101	7,523	5,838	4,725
9-10	4,324	3,369	2,234	4,074	3,209	2,659
10-11	2,306	1,645	1,119	2,021	1,864	1,243
11-12	1,007	711	547	918	858	555
12-13	330	289	181	370	273	243
13-14	98	78	22	106	79	40
14-15	47	29	6	38	13	14

Notes: Expanded sample sizes are numbers of person-year observations. Each cell in the table corresponds to a separate data set, for which the sample size (number of person-year observations) is shown. There are 90 data sets. For each data set, weighted and unweighted sample sizes are the same. 0-1 denotes the transition from a woman's own birth to first birth, and 1-2 denotes the transition from first birth to second birth. In the period analysis, periods are the five-year period before each survey. In the cohort analysis, cohorts are defined as women age 45-49 at the time of survey.

### ***Validation of the $P_{aitmn}$ Method***

By way of validating the  $P_{aitmn}$  methodology, selected measures of fertility and child mortality are calculated using each of the four methods—the birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods—insofar as this is possible. In order to increase the precision of comparisons between estimates derived by the birth history method and estimates derived by  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods, the data set for the birth history method is constructed in the same way as the data set for the other three 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 shows unadjusted and adjusted TFRs derived by the birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods applied to both cohort and period data from NFHS-1. With the birth history estimates of the TFR as the base of comparison, the table shows that all three life table methods ( $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$ ) slightly underestimate the unadjusted cohort  $TFR_{PPR}$

(the birth history-derived  $TFR_{PPR}$  being a more appropriate basis for comparison than the birth history-derived  $TFR_{ASFR}$ , because the  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods are also fundamentally PPR-based.) One expects some differences, because the  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods all impose functional forms on the data, which the birth history method does not impose. In the case of adjusted estimates, those derived by the  $P_{it}$  and  $P_{ait}$  methods agree closely, while those derived by the  $P_{aitmn}$  method are somewhat lower. Some differences among the adjusted estimates are expected, because of the artificial way in which the socioeconomic variables are controlled, combined with the non-linear nature of the  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  models.

Figures 5–10 supplement the above unadjusted cohort estimates of the TFR with additional detail pertaining other cohort fertility measures. Figure 5 shows that the  $P_{aitmn}$ -derived unadjusted estimates of cohort PPRs agree closely with corresponding estimates derived by the birth history,  $P_{it}$ , and  $P_{ait}$  methods, especially at the lower parities. The four methods produce slightly different results at higher parities where the expanded sample sizes are small. Figure 6 shows that the  $P_{aitmn}$ -derived unadjusted estimates of cohort ASFRs agree closely with the birth-history and  $P_{ait}$ -derived estimates. Figures 7 and 8 show that the  $P_{aitmn}$ -derived unadjusted cohort estimates of mean and median age at childbearing by child's birth order agree closely with corresponding estimates derived by the birth history and  $P_{ait}$  methods. In the case of all births (regardless of child's birth order), the various unadjusted cohort estimates of mean age at childbearing approximately coincide, as do the various unadjusted cohort estimates of median age at childbearing. Figures 9 and 10 show that the  $P_{aitmn}$ -derived unadjusted cohort estimates of mean and median closed birth intervals agree closely with corresponding estimates derived by the birth history,  $P_{it}$ , and  $P_{ait}$  methods for all but the very high-order parity transitions where numbers of cases are small.

In the case of period estimates, Table 3 shows that the estimate of the period TFR derived by the  $P_{it}$  method is closest to the estimate of the TFR derived by the birth history estimate. This is not surprising, because in this case the birth history method is Feeney's PPR-based method of estimating a period TFR (Feeney and Yu 1987), and because the  $P_{it}$  method is a multivariate generalization of Feeney's PPR-based method. Once age is introduced into the global life tables derived by the  $P_{ait}$  and  $P_{aitmn}$  methods, the estimate of the period TFR drops substantially. As already mentioned, this drop occurs apparently because, at any given age in the global life table, the distribution of women by parity, duration in parity, and the two child mortality states at the beginning and end of the age interval  $a$  to  $a+1$  differs between the actual population and the global life table population.

Figures 11–16 supplement the estimates of the unadjusted period TFR in Table 3 with additional detail for the other fertility measures. As expected, Figure 11 shows that the  $P_{it}$ -derived estimates of unadjusted period PPRs agree closely with corresponding birth history estimates derived by Feeney's method, even at the higher parities. By contrast, the  $P_{ait}$ -derived and  $P_{aitmn}$ -derived unadjusted estimates of PPRs, though in close agreement with each other, are much lower than those derived by the  $P_{it}$  method or by Feeney's method, except at the very high parities. These differences parallel the

Table 3: Comparison of total fertility rates derived by the birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: NFHS-1

	Birth History	$P_{it}$ method	$P_{ait}$ method	$P_{aitmn}$ method
<b>COHORT ANALYSIS</b>				
<b>TFR<sub>PPR</sub></b>				
Unadjusted	5.03	4.93	4.96	4.89
Adjusted	N/A	4.82	4.85	4.69
<b>TFR<sub>ASFR</sub></b>				
Unadjusted	5.09	N/A	4.96	4.89
Adjusted	N/A	N/A	4.85	4.69
<b>CEB</b>	5.03			
<b>PERIOD ANALYSIS</b>				
<b>TFR<sub>PPR</sub></b>				
Unadjusted	3.53	3.50	3.19	3.13
Adjusted	N/A	3.49	3.19	3.16
<b>TFR<sub>ASFR</sub></b>				
Unadjusted	3.44	N/A	3.19	3.13
Adjusted	N/A	N/A	3.19	3.16

Notes: In the cohort case, what is labeled as the birth history method is simply the average number of children ever born in the age group 45-49. In the period case, what is labeled as the birth history method is actually Feeney's PPR-based method (Feeney and Yu 1987). In the case of the  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods, "unadjusted" means that no socioeconomic predictor variables are included in the underlying CLL models (but at least some of the basic predictor variables—duration in parity, age at starting parity, and the two child mortality variables  $D_1$  and  $D_2$ —are included, depending on the method). "Adjusted" means that the socioeconomic variables shown in Table 1 are included in the set of predictor variables in the underlying CLL models, and that these socioeconomic variables are held constant at their duration-specific mean values in the expanded data set pertaining to the CLL model for any particular parity transition. The predictor variables in the CLL models underlying the adjusted estimates additionally include duration in parity in the case of the  $P_{it}$  method, both duration in parity and age at starting parity in the case of the  $P_{ait}$  method, and both duration in parity, age at starting parity, and the two child mortality variables  $D_1$  and  $D_2$  in the case of the  $P_{aitmn}$  method. These variables are allowed to vary while the socioeconomic variables are held constant.

Figure 5: Comparison of unadjusted PPRs derived by the birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 45-49, cohort analysis, NFHS-1

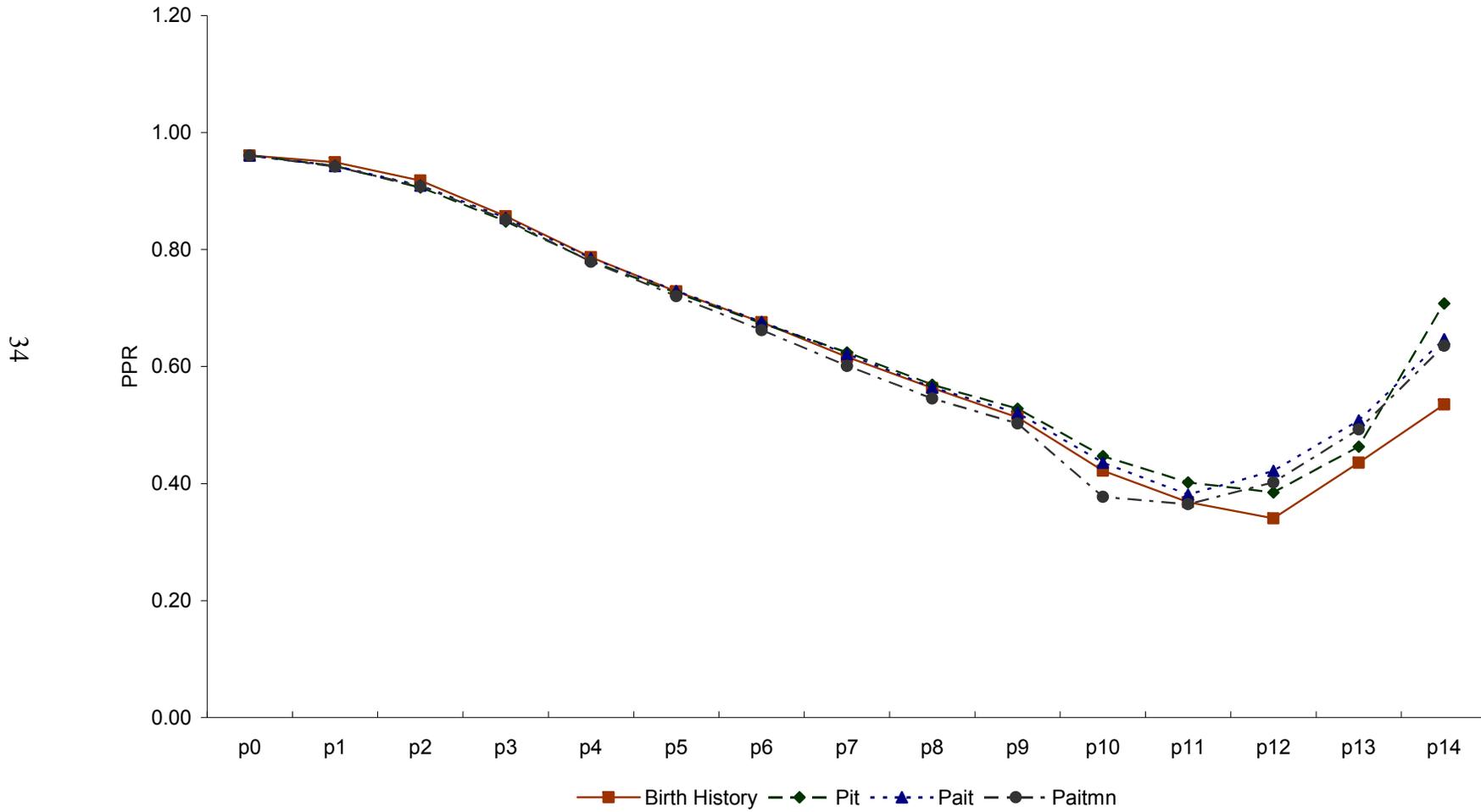


Figure 6: Comparison of unadjusted ASFRs derived by the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 45-49, cohort analysis, NFHS-1

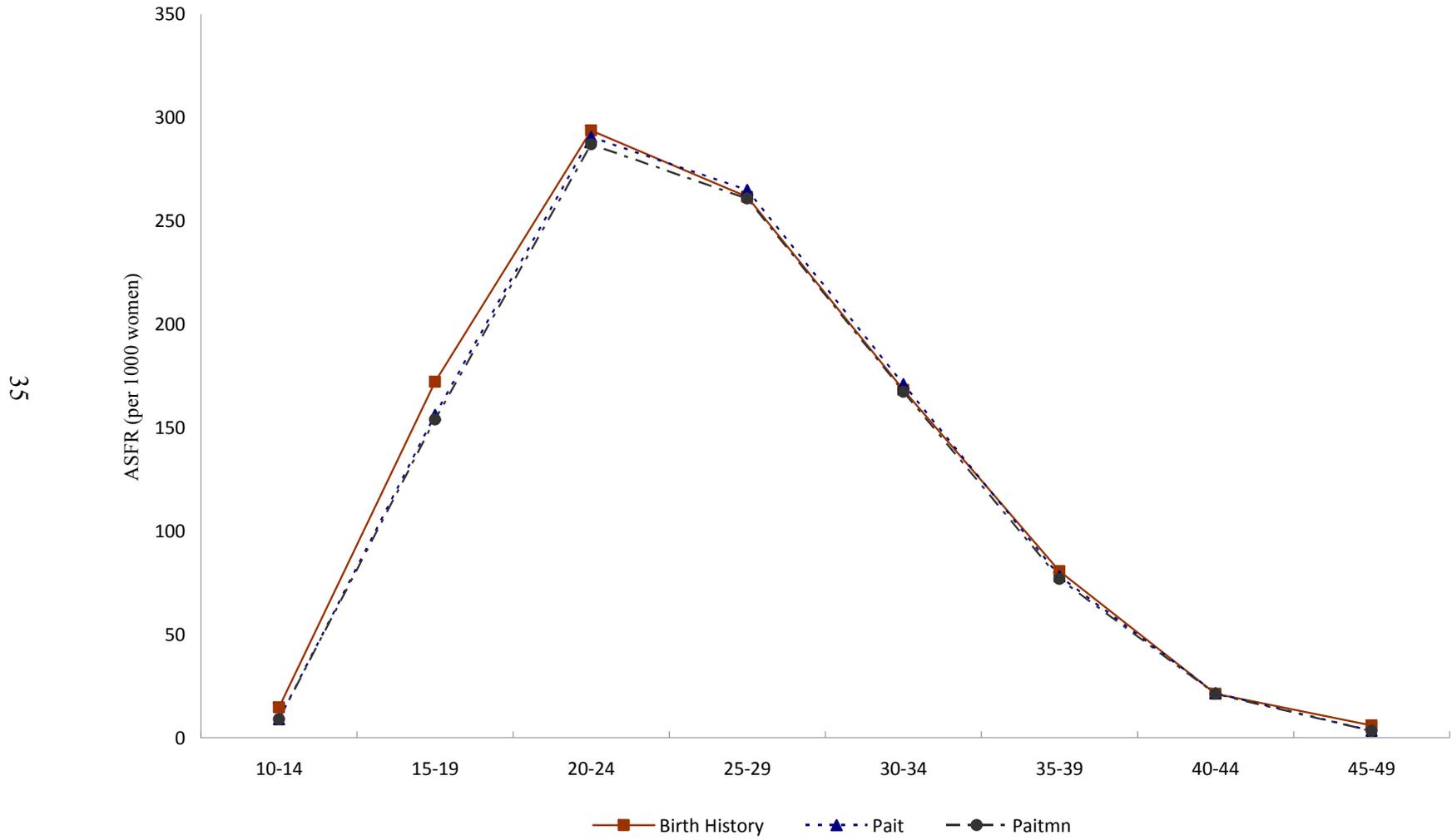


Figure 7: Comparison of unadjusted mean age at childbearing derived by the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 45-49, cohort analysis, NFHS-1

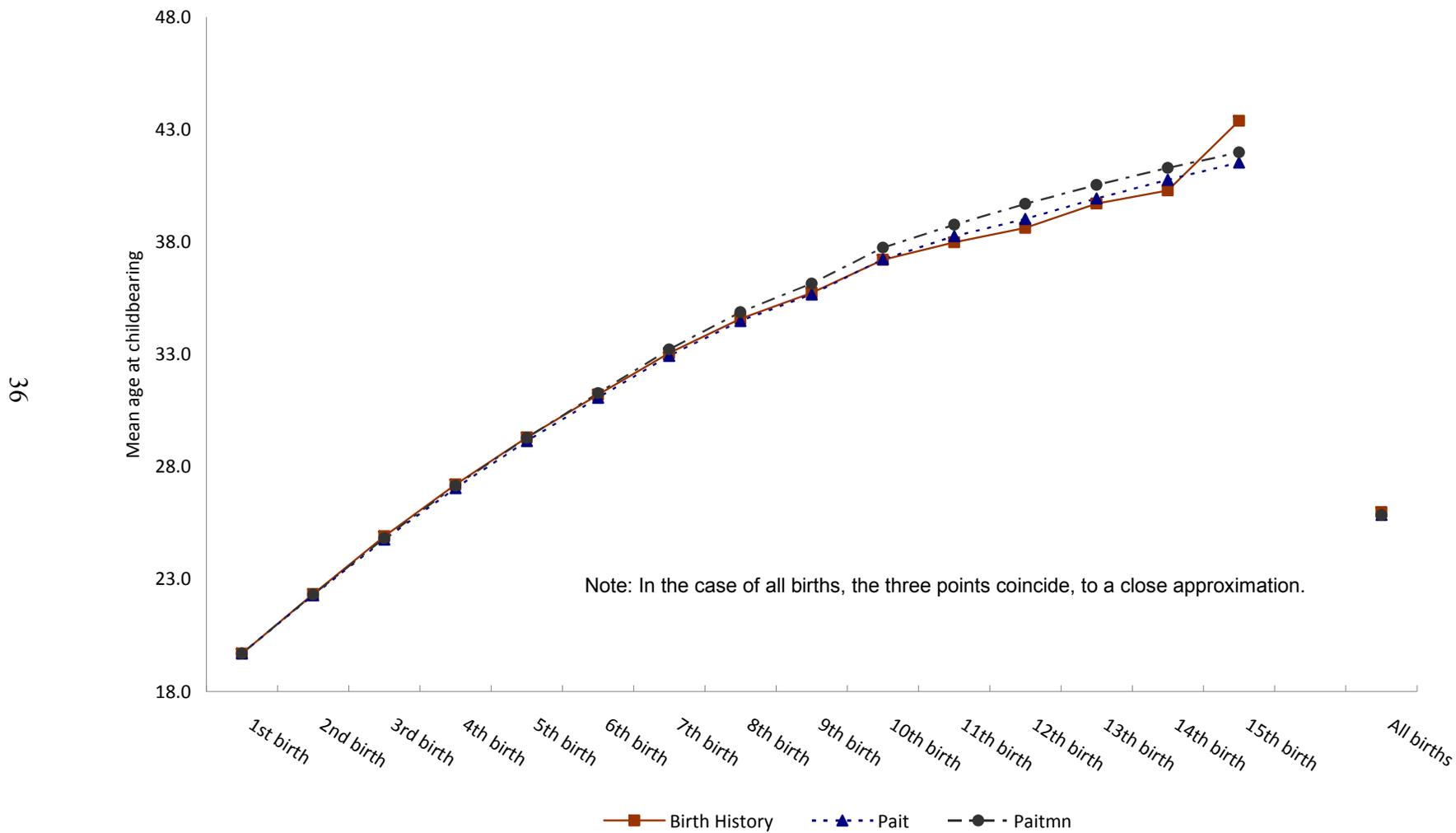


Figure 8: Comparison of unadjusted median age at childbearing derived by the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women 45-49, cohort analysis, NFHS-1

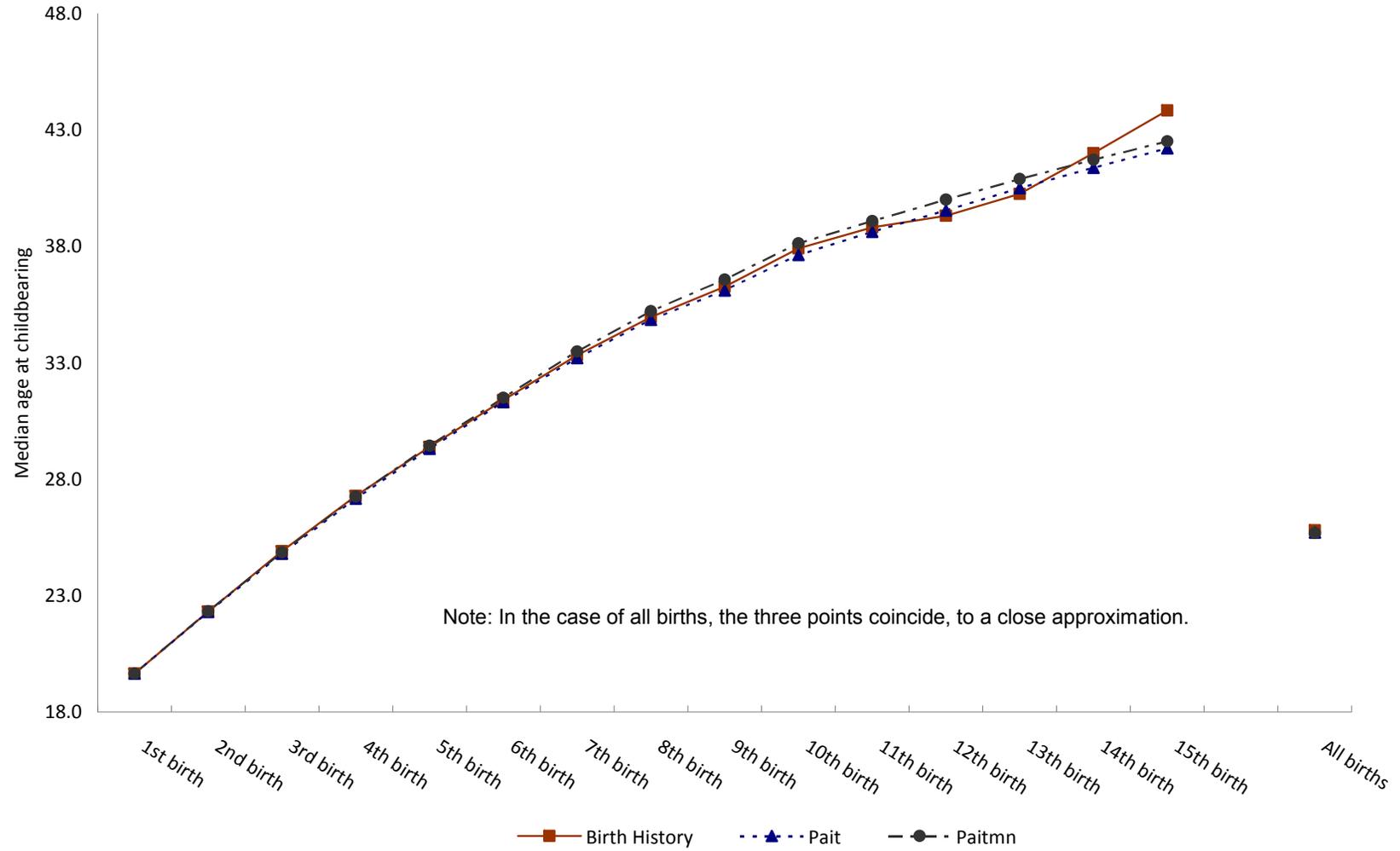


Figure 9: Comparison of unadjusted mean closed birth intervals derived by the birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 45-49, cohort analysis, NFHS-1

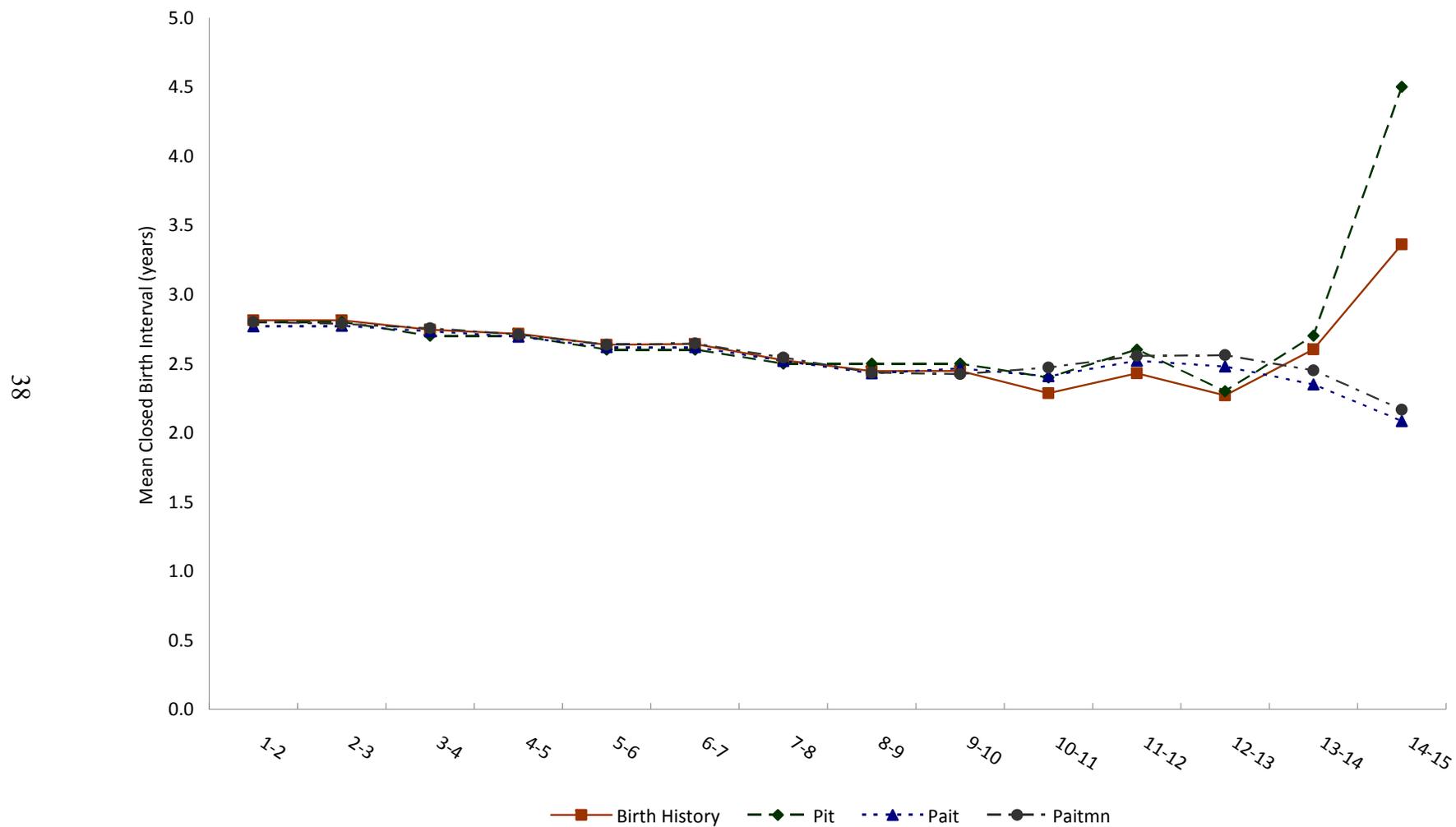
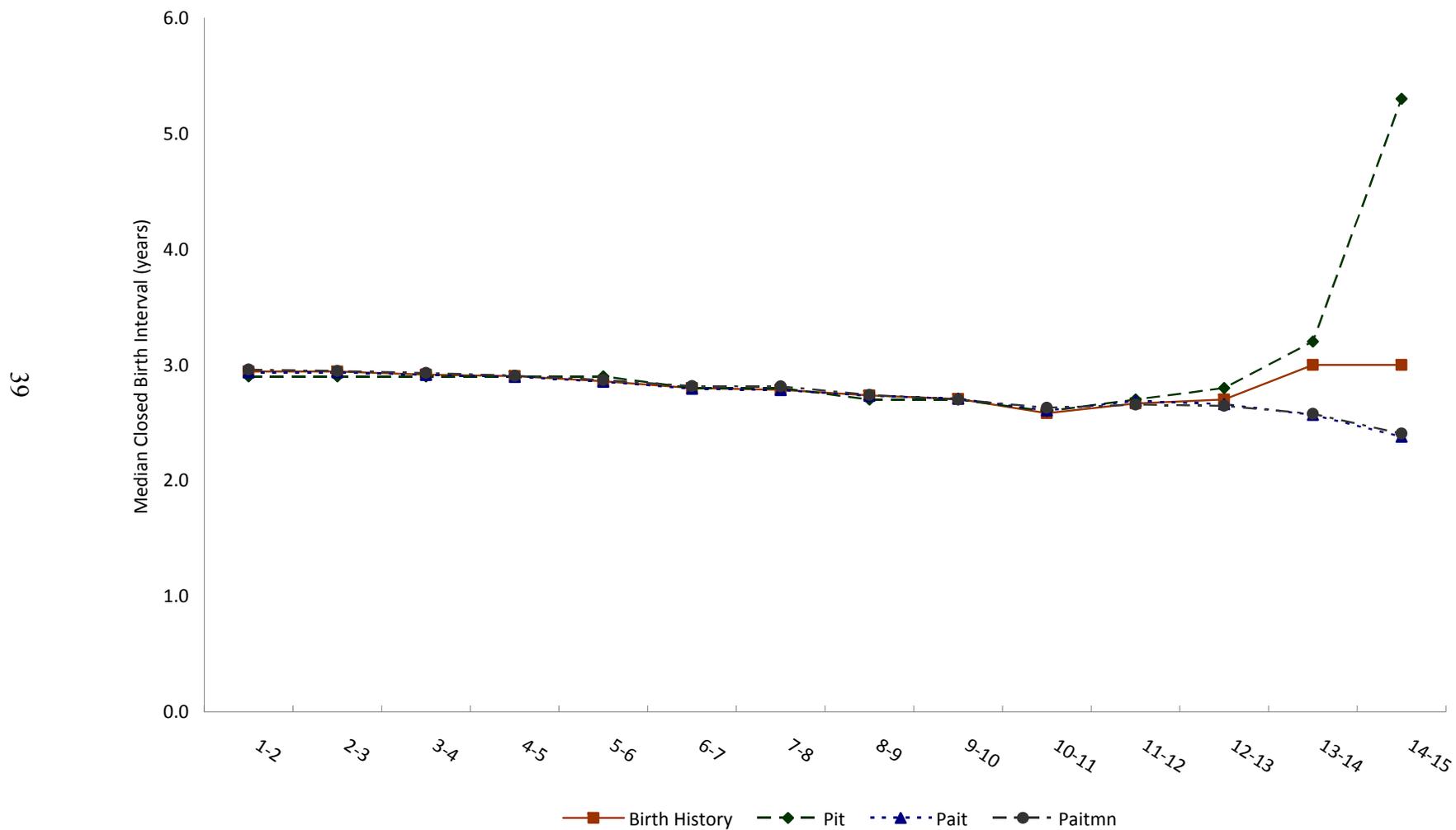


Figure 10: Comparison of unadjusted median closed birth intervals derived by the birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 45-49, cohort analysis, NFHS-1



unadjusted period TFR differences in Table 3. Figure 12 shows that the  $P_{ait}$ -derived and  $P_{aitmn}$ -derived unadjusted estimates of period ASFRs are lower than conventional birth history estimates of period ASFRs, except at age 25–29. (It is not clear why 25–29 is an exception.) Figures 13 and 14 show that the  $P_{ait}$ -derived and  $P_{aitmn}$ -derived unadjusted estimates of mean and median ages at childbearing are higher than corresponding estimates derived by the conventional birth history method. Figures 15 and 16 show that the unadjusted period estimates of mean and median closed birth intervals derived from birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods agree closely, except at the very high parities.

Table 4 compares unadjusted child mortality measures and fertility measures derived by the birth history method and the  $P_{aitmn}$  method for both cohort and period data from NFHS-1. As the table indicates, the two methods yield numbers of child deaths per woman ( ${}_5D_a$  and  $D$ ) that agree more closely for cohort data than for period data, the reason being that, in the period case but not the cohort case, the distribution of women by age, parity, duration in parity, and the two lagged child mortality variables differs between the synthetic global life table population of women and the actual population of women. This difference in the distribution of women tends to result in fewer births in the global life table population of women than in the actual population of women (Retherford et al. 2010b). Because of fewer births, there are also fewer child deaths.

In the cohort case but not in the period case, perfect model fits to the data (including not only the underlying CLL models but also the partitioning factors  $Q_{aitm}$ ) should result in perfect agreement between the birth history estimates and the  $P_{aitmn}$ -derived estimates of  ${}_5D_a$ ,  $D$ , and mean and median  $A_d$ . In the cohort case in Table 4, the  $P_{aitmn}$ -derived estimates of  ${}_5D_a$  and  $D$  are slightly lower than the corresponding birth history estimates (by 0.09 child death per woman in the case of  $D$ ), and the  $P_{aitmn}$ -derived estimates of mean  $A_d$  and median  $A_d$  are somewhat higher (by 0.04 year and 0.07 year, respectively) than the corresponding birth history estimates. These comparisons show that the fit of the underlying CLL models in the  $P_{aitmn}$  method and the partitioning factors  $Q_{aitm}$  to the cohort data is fairly close but not perfect. Figure 17 shows that most of the underestimation of child deaths per woman per year by the  $P_{aitmn}$  method occurs at woman's ages 20–24 and 25–29.

As already mentioned, in the period case in Table 4 one expects the  $P_{aitmn}$ -derived estimates of  ${}_5D_a$  and  $D$  to be lower than the corresponding birth history estimates. Table 4 and Figure 18 show how much lower in the case of NFHS-1. Table 4 shows that the  $P_{aitmn}$ -derived period estimate of  $D$  is about half as large as the corresponding birth history estimate of  $D$  that the birth history method generates for the period data. Table 4 and Figure 18 show that this difference, part of which is expected and part of which probably results from underestimation of child deaths (judging from the cohort results), occurs at all ages of women. By contrast, Table 4 shows that the  $P_{aitmn}$ -derived and birth history-derived estimates of mean and median  $A_d$  do not differ noticeably more in the period case than in the cohort case. In the case of median  $A_d$ , the agreement is even closer in the period case than in the cohort case.

Figure 11: Comparison of unadjusted PPRs derived by the birth history,  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 10-49, period analysis for 1988-92, NFHS-1

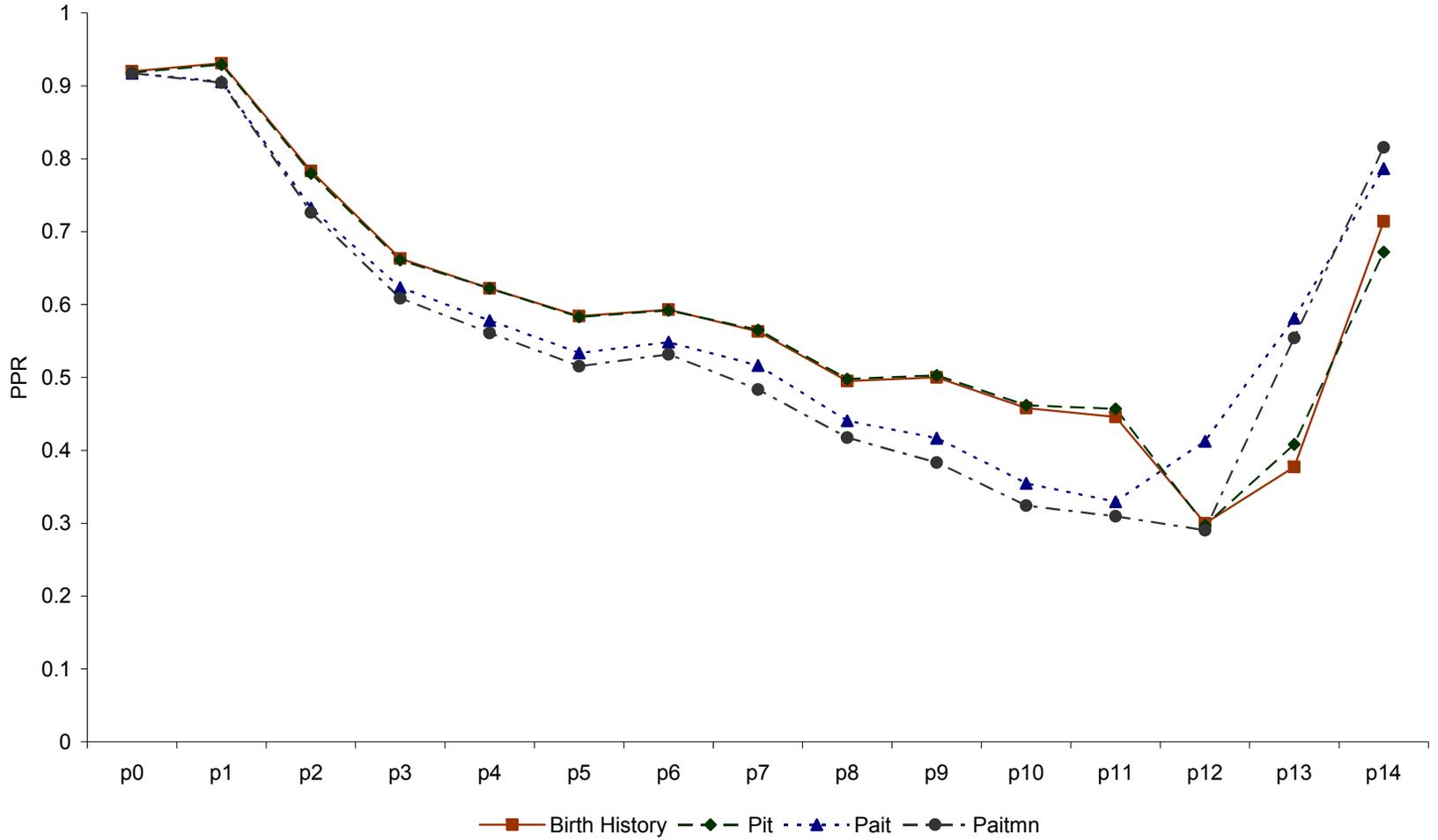


Figure 12: Comparison of unadjusted ASFRs derived by the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 10-49, period analysis for 1988-92, NFHS-1

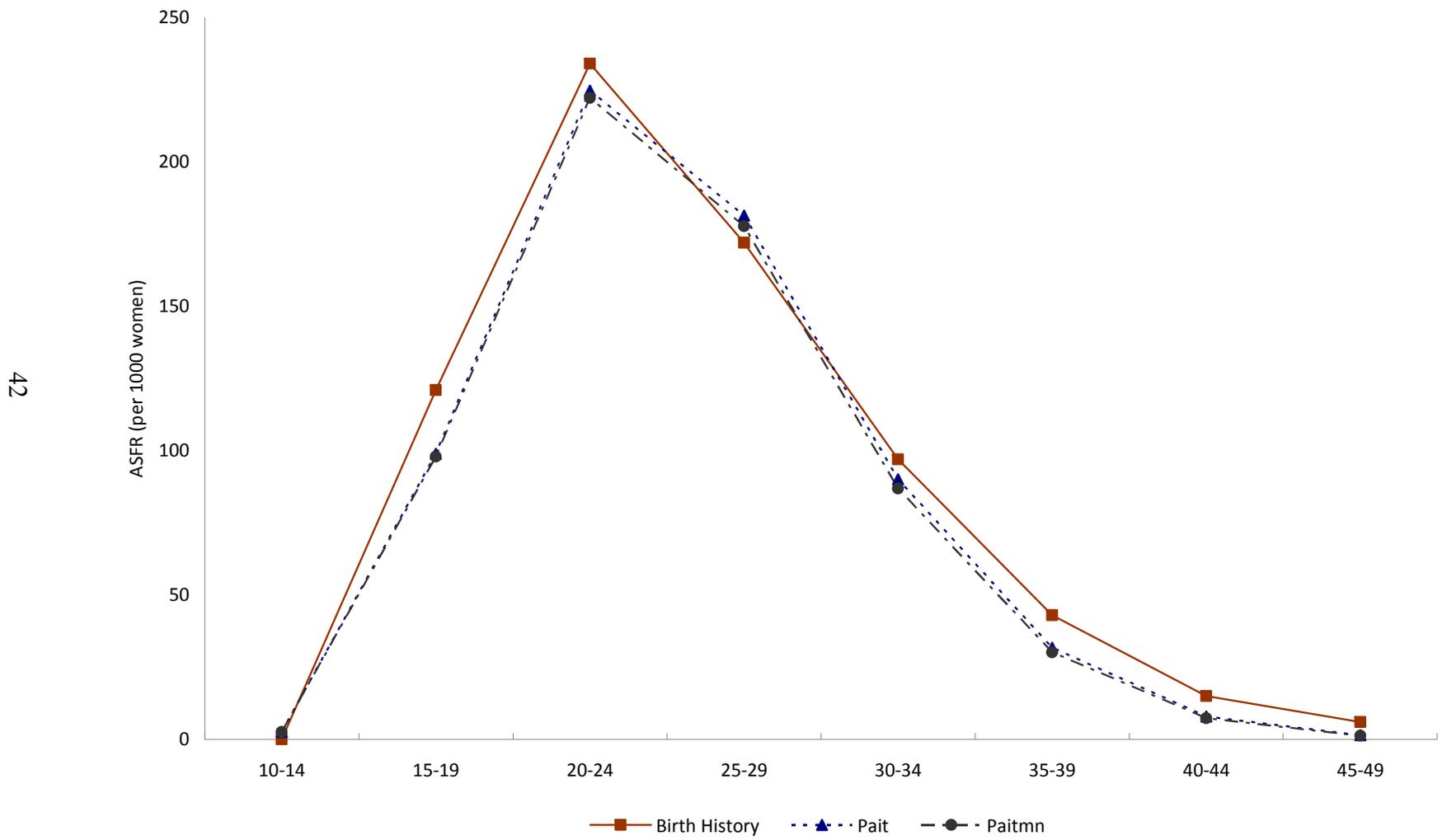


Figure 13: Comparison of unadjusted mean age at childbearing derived by the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 10-49, period analysis for 1988-92, NFHS-1

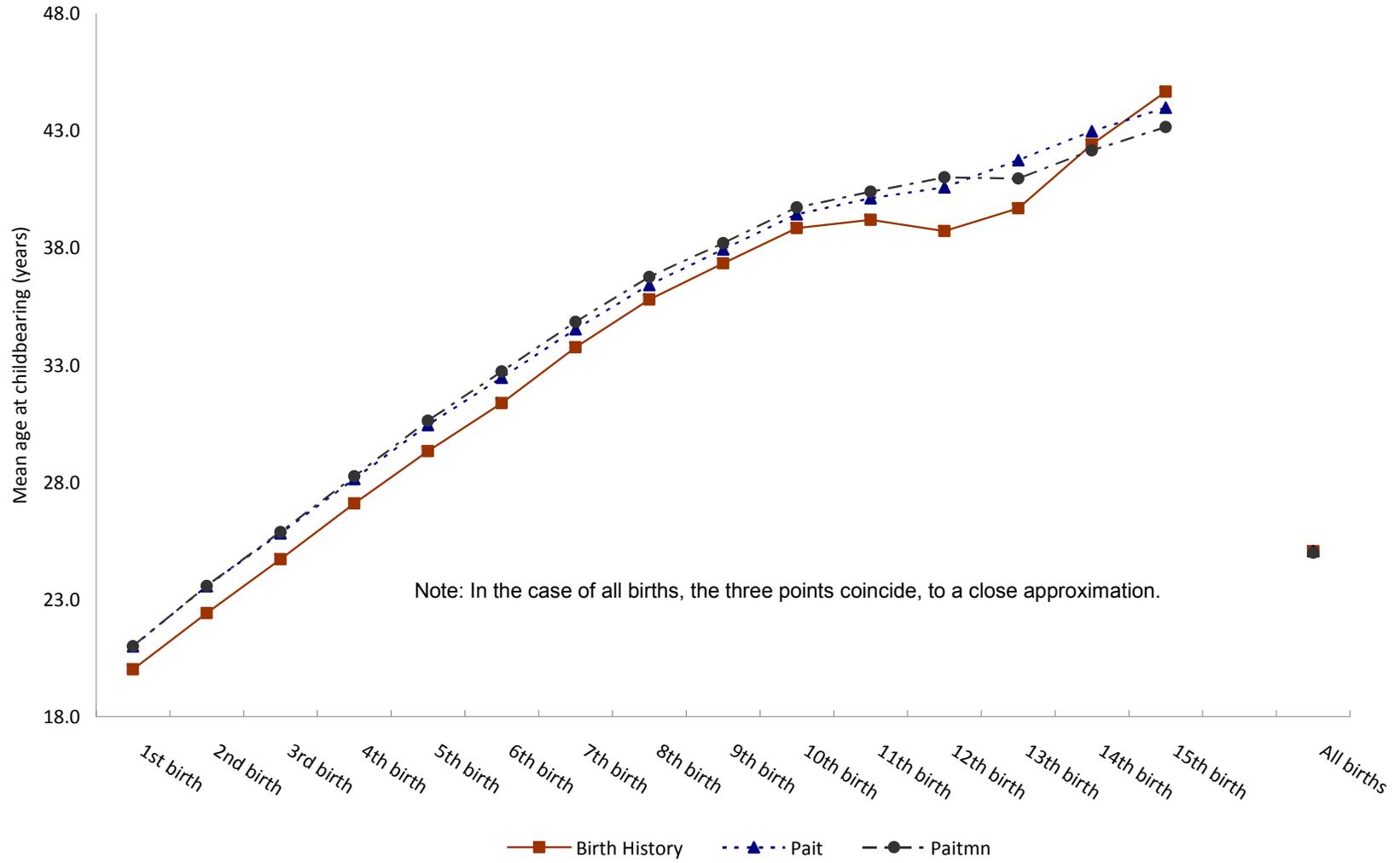


Figure 14: Comparison of unadjusted median age at childbearing derived by the birth history,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 10-49, period analysis for 1988-92, NFHS-1

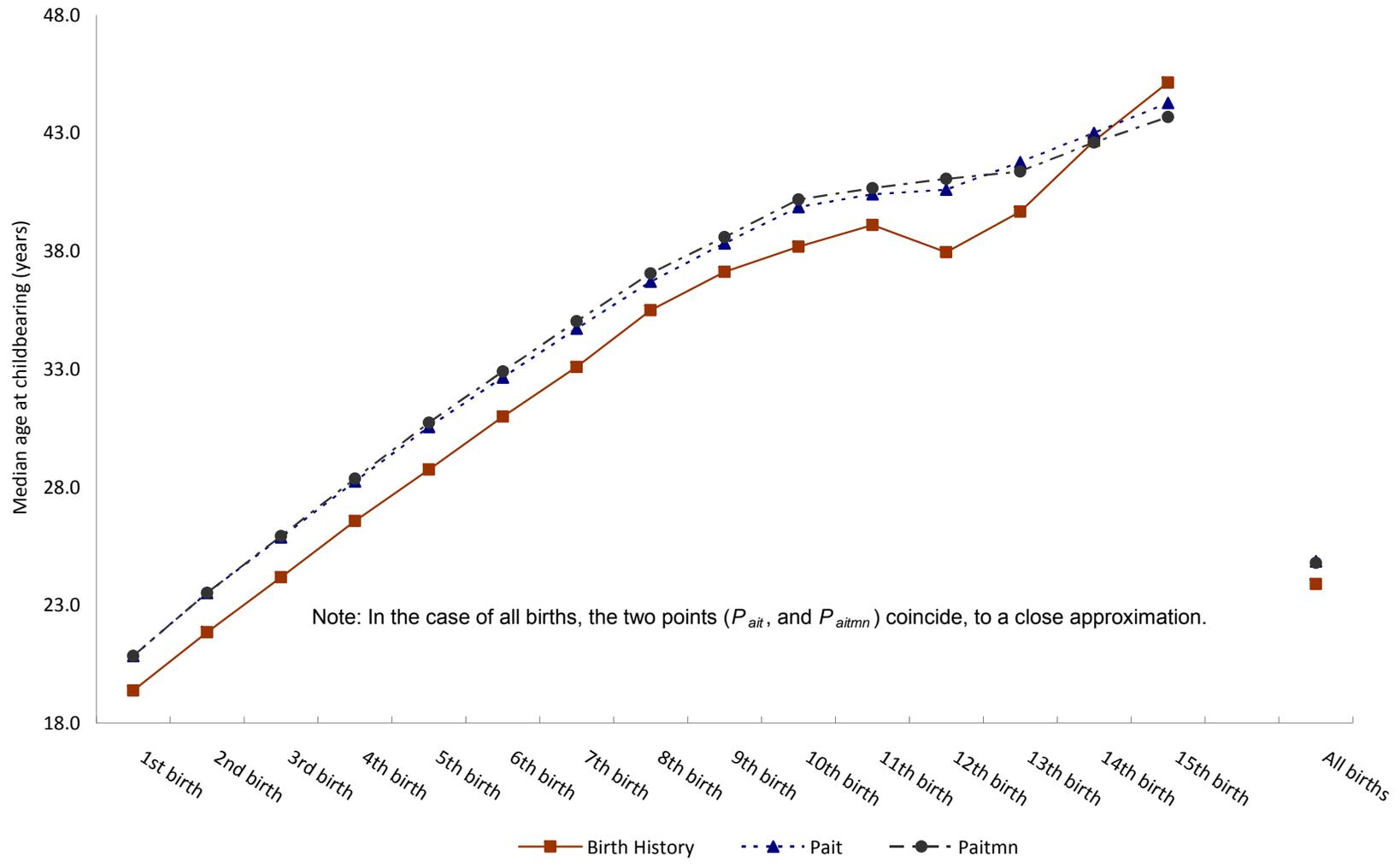


Figure 15: Comparison of unadjusted mean closed birth intervals derived by the  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 10-49, period analysis for 1988-92, NFHS-1

45

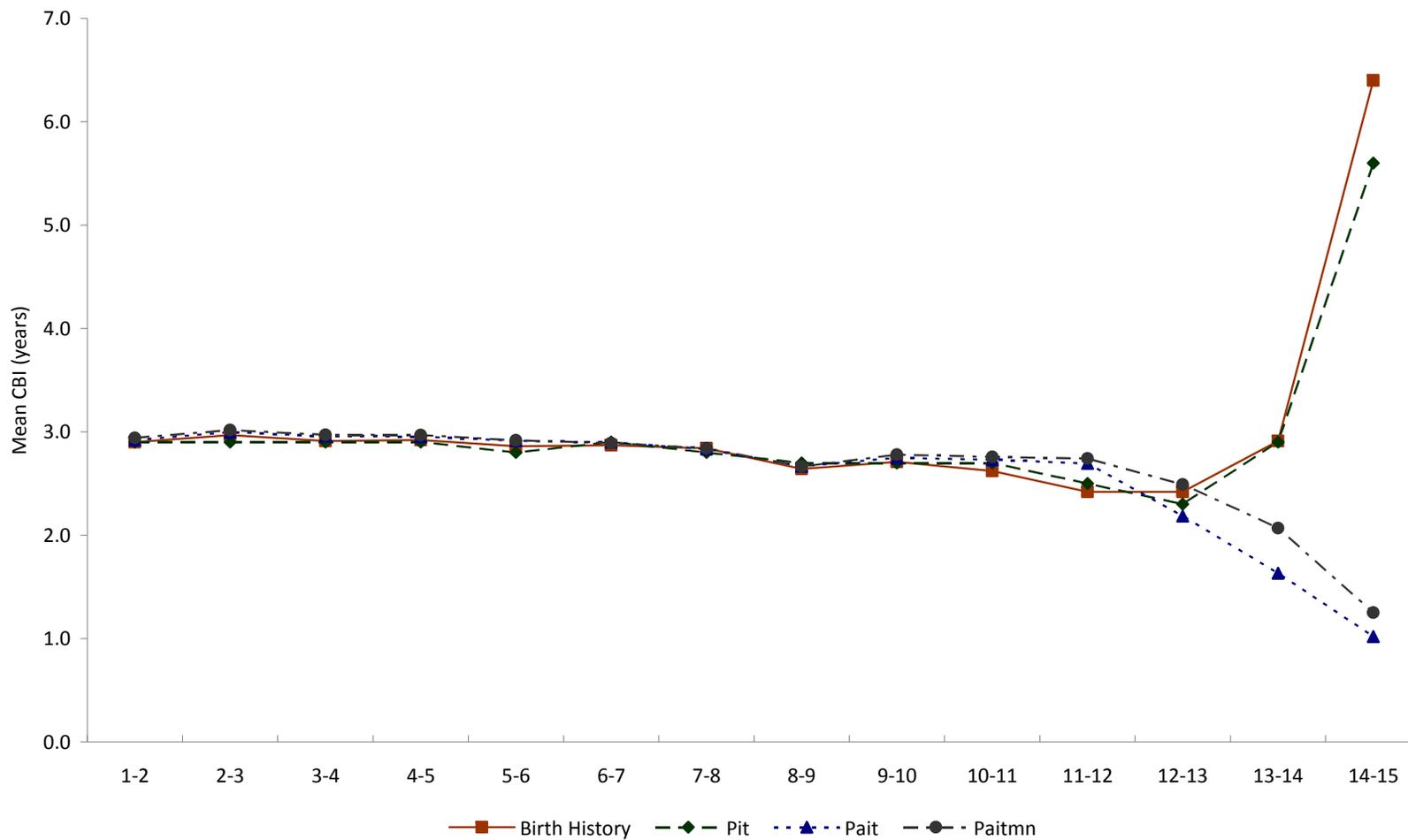


Figure 16: Comparison of unadjusted median closed birth intervals derived by the  $P_{it}$ ,  $P_{ait}$ , and  $P_{aitmn}$  methods: Women age 10-49, period analysis for 1988-92, NFHS-1

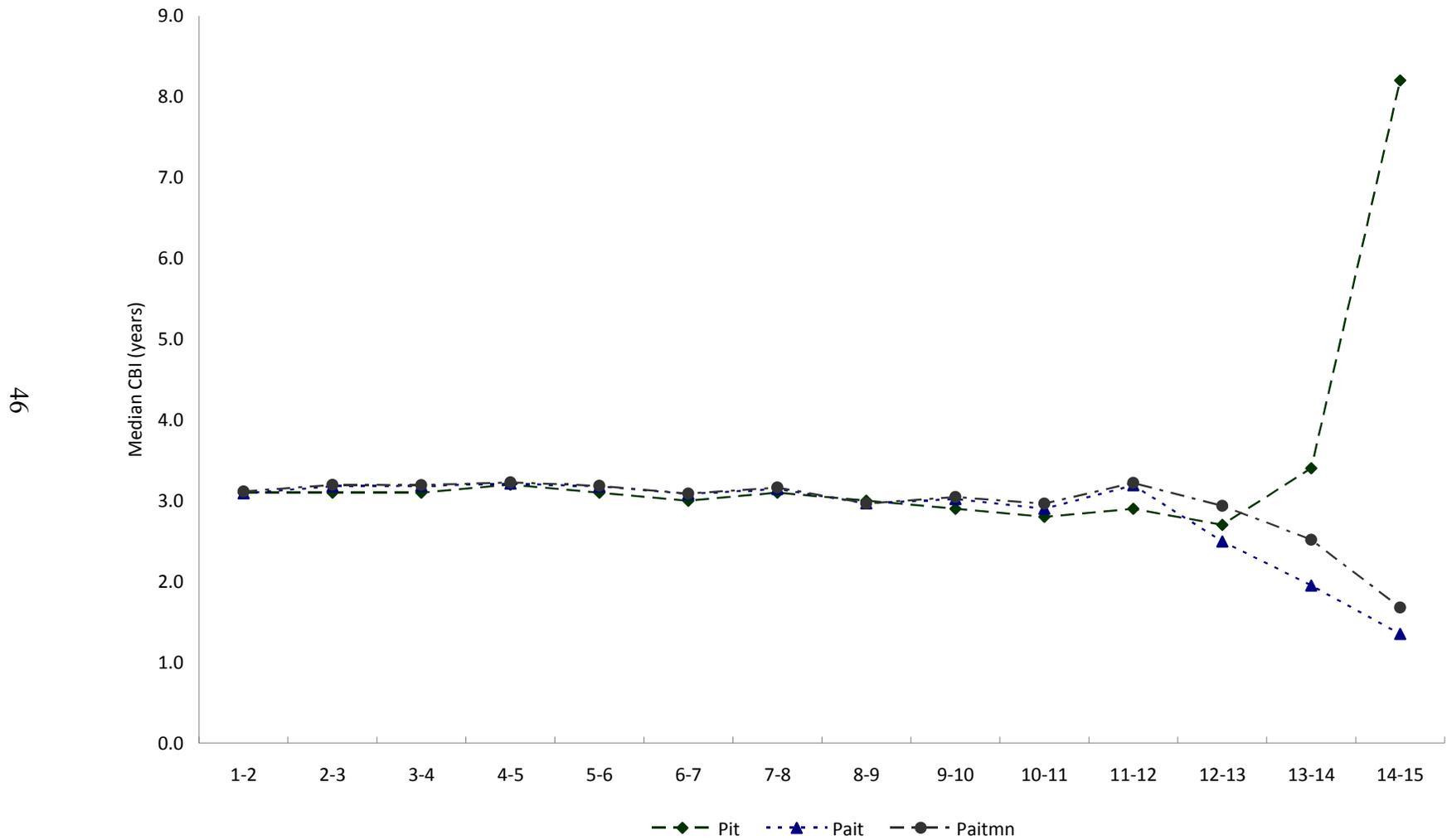


Table 4: Comparison of unadjusted child mortality measures derived by the birth history and  $P_{aitmn}$  methods: cohort and period analysis, India's NFHS-1

Measure	COHORT		PERIOD	
	Birth History	$P_{aitmn}$ Method	Birth History	$P_{aitmn}$ Method
${}_5D_a$				
10-14	0	0	0	0
15-19	16	13	7	4
20-24	48	40	24	15
25-29	53	48	27	17
30-34	42	41	21	12
35-39	24	28	15	7
40-44	14	15	10	3
45-49	8	6	8	1
$D$	1.03	0.94	0.56	0.29
Mean $A_d$	29.2	29.6	29.0	28.5
Median $A_d$	28.7	29.4	27.9	28.1
$TFR_{WCM}$		4.89		3.13
$TFR_{NCM}$		4.57		2.98
$R$		0.34		0.51

Notes:  ${}_5D_a$  is child deaths (regardless of child's age) per 1,000 women per year between woman's ages  $a$  and  $a+5$ , and  $D$  is total number of child deaths per woman, regardless of child's age or woman's age, with child deaths (but not woman's age) lagged by one year (see equations (55) and (56)).  $A_d$  is mean age of woman at child death in the global life table.  $TFR_{WCM}$  is the child-mortality-effect-present TFR.  $TFR_{NCM}$  is the child-mortality-effect-free TFR. The replacement rate is  $(TFR_{WCM} - TFR_{NCM})/D$ .

Figure 17: Comparison of unadjusted child mortality rates (child deaths but not woman's age lagged by one year) by woman's age derived by the birth history and  $P_{aitmn}$  methods: Women age 45-49, cohort analysis, NFHS-1

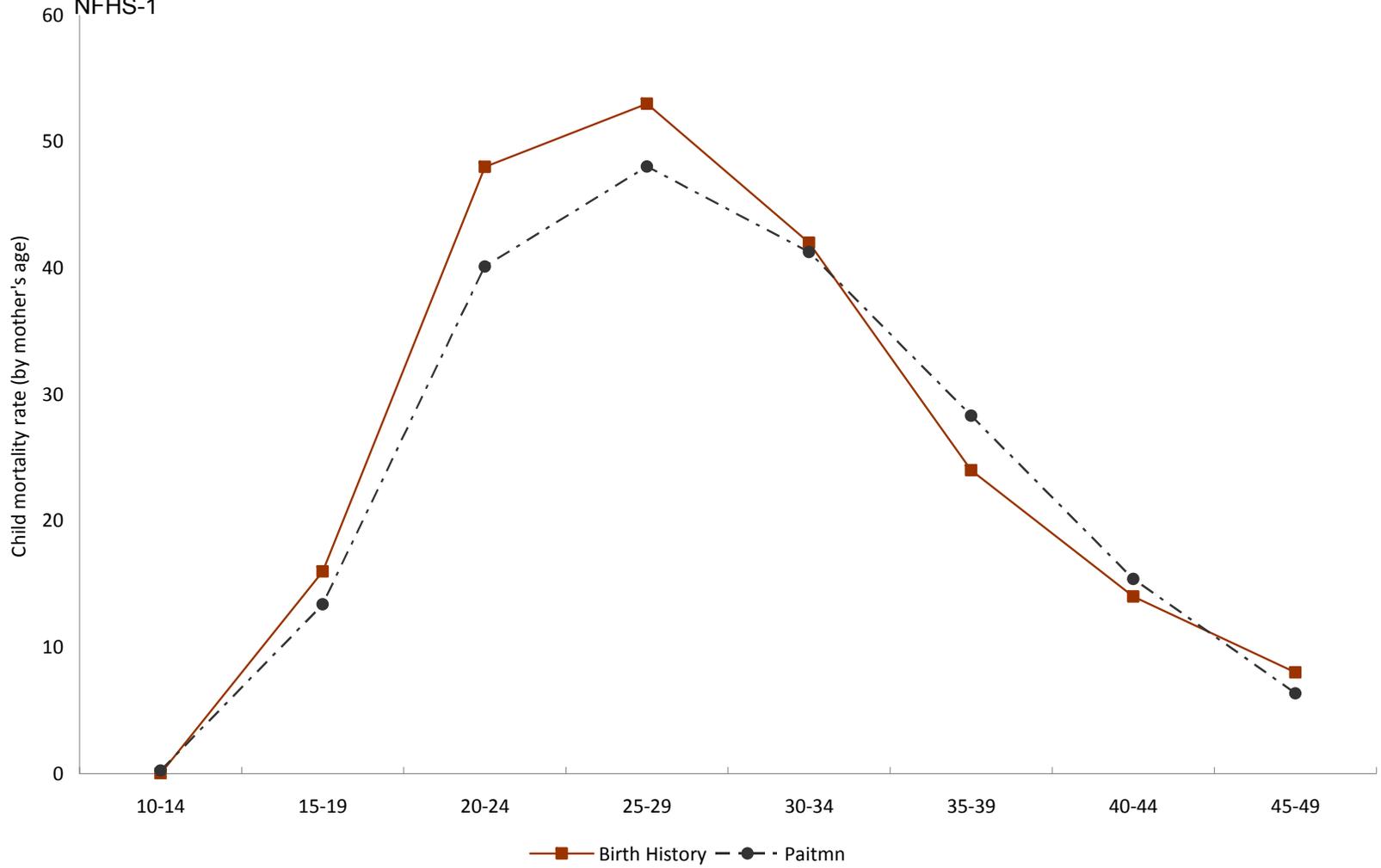
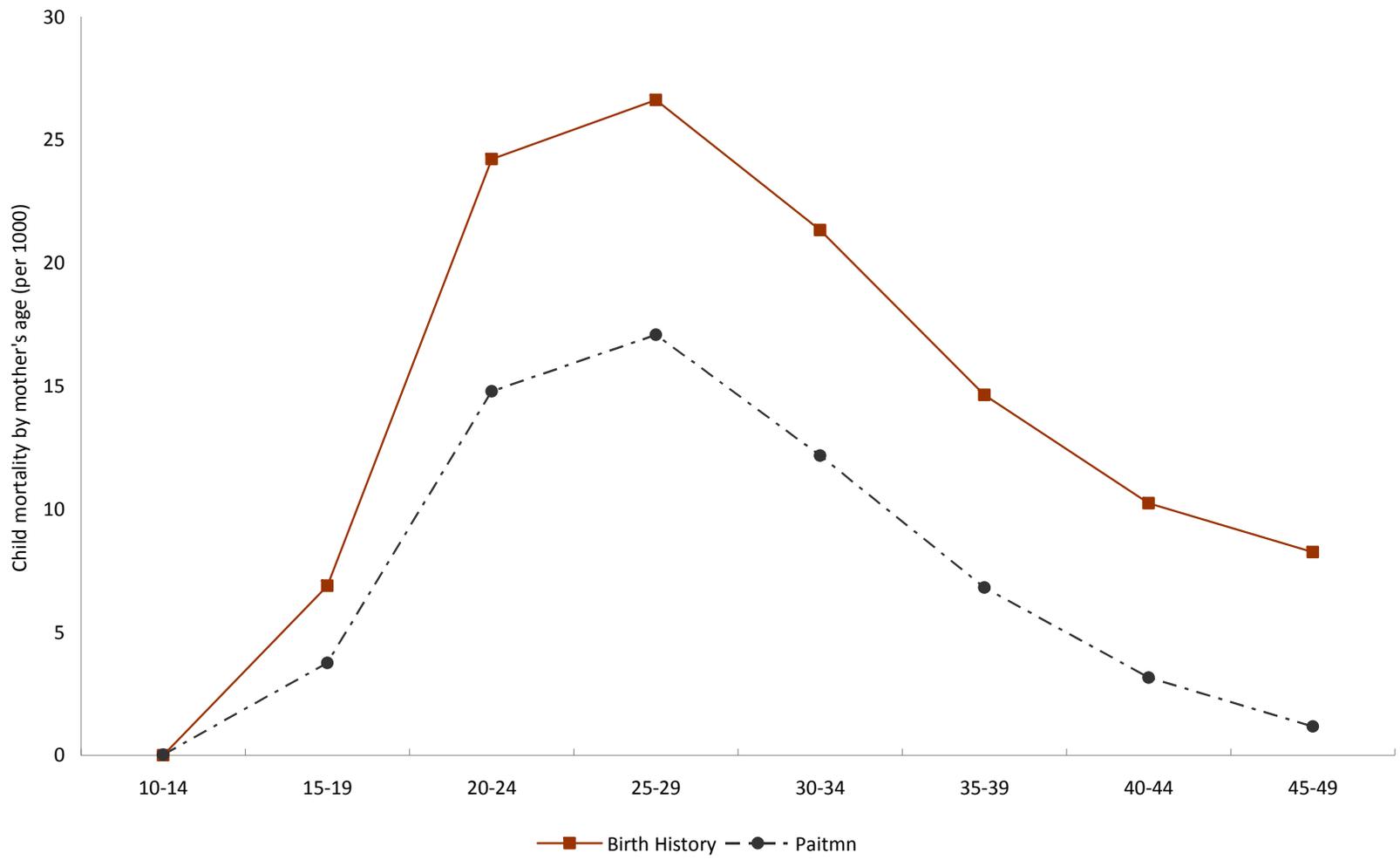


Figure 18: Comparison of unadjusted child mortality rates (child deaths but not woman's age lagged by one year) by woman's age derived by the birth history and  $P_{aitmn}$  methods: Women age 10-49, period analysis for 1988-92, NFHS-1



Estimates of  $TFR_{WCM}$ ,  $TFR_{NCM}$ , and  $R$  can be calculated by the  $P_{aitmn}$  method but not by the birth history method. As shown in Table 4, the replacement rate is 0.34 in the cohort case and 0.51 in the period case. In other words, a dead child is replaced by about one-third of a birth in the cohort case and about one-half of a birth in the period case. One expects  $R$  to be higher in the period case, because desired family size and fertility have been falling and contraception has been increasing over time, implying substantially more fertility control among women age 10–49 during the 5-year period before survey than among the cohort of women age 45–49 at time of survey who had most of their children during earlier periods when fertility control was not so widely practiced.

When fertility is high, replacement is relatively low (due not only to a low level of volitional replacement but also to non-volitional replacement as a consequence of truncated breastfeeding leading to early resumption of ovulation after an early childhood death). But, when fertility is low, replacement is relatively high due to increased volitional replacement due to earlier completion of childbearing (as a consequence of falling fertility) and higher rates of contraception and abortion.

### ***Validation of the Partitioning Factor***

For reasons explained earlier, the “overall probability approach” to calculating partitioning factors is preferred over the “logit model approach.” This section tests the sensitivity of the global life tables to the choice of partitioning factor.

The sensitivity analysis is based on four different global life tables. The first uses the overall probability approach, except when the denominator in the formula for the partitioning factor  $Q_{aitm}$  is zero, in which case  $Q_{aitm}$  is set to zero. The second global life table is done the same way, except that  $Q_{aitm}$  is set to one when the denominator is zero. The third global life table always uses the logit model approach to calculate the partitioning factor. The fourth global life table uses the overall probability approach throughout, except when the denominator of the formula for  $Q_{aitm}$  is zero, in which case the “logit model approach” is used. The four global life tables are constructed using cohort data (for women age 45–49) from NFHS-1. All four global life tables are based on similar CLL models that include no socioeconomic predictors; i.e., the only predictor variables that are included are those that represent the basic dimensions of the global life table. In the CLL model for parity transition  $i$  to  $i+1$ , these variables pertain to the dimensions  $a$ ,  $t$ ,  $m$ , and  $n$ .

Table 5 compares GLT-derived values of TFR, mean  $A_d$ , and  $R$  over the four global life tables. The three GLTs that use the overall probability approach as the main approach (GLTs 1, 2, and 4) yield TFRs that are identical out to two decimal places. The reason for the close agreement is that instances where the denominator of the  $Q_{aitm}$  formula is zero are rare. GLTs 1, 2, and 4 do yield some different values of  $D$  and  $R$ , however. Virtually identical estimates of  $D$  and  $R$  for GLTs 1 and 4 occur because (1) a zero denominator in the formula for  $Q_{aitm}$  is a rare event, and (2) a child death over a one-year age interval is also a rare event. In cases of a zero denominator it does appear

preferable, however, to set  $Q_{aim}$  equal to the estimate of  $Q_{aim}$  derived by the logit model approach, rather than setting it equal to zero. This is why approach used in GLT 4 is the preferred approach in this paper.

Table 5: Sensitivity of selected global life table results to the choice of the partitioning factor  $Q_{aim}$ ; Women age 45-49, cohort analysis, NFHS-1

Method	TFR	$D$	$R$
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:

**Method 1:** The "overall probability" approach is used to estimate  $Q_{aim}$  if the denominator on the right side of equation (53) is non-zero. probability of a birth during the interval differs between the two groups.  $Q_{aim}$  set to zero if the above-mentioned denominator is zero.

**Method 2:** The "overall probability" approach is used to estimate  $Q_{aim}$  if the denominator on the right side of equation (53) is non-zero.  $Q_{aim}$  set to one if the above-mentioned denominator is zero.

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

**Method 4:** The "overall probability" approach is used to estimate  $Q_{aim}$  if the denominator on the right side of equation (53) is non-zero. The "logit model" approach is used if the above-mentioned denominator is zero. (This is the preferred approach that is actually used in the application to the three NFHS surveys.)

## APPLICATION OF THE METHODOLOGY TO ALL THREE NFHS SURVEYS

Table 6 presents unadjusted and adjusted values of selected global life table measures, based alternatively on cohort and period data from NFHS-1, NFHS-2, and NFHS-3. The cohort estimates of  $D$  show that number of child deaths per woman declined considerably over the three surveys, with little or no difference between the unadjusted and adjusted estimates of  $D$  in each survey. The decline in the cohort estimate of  $D$  occurred not only because mortality declined but also because fertility declined (fewer births mean fewer child deaths). The cohort estimate of  $\Delta TFR$  (the incremental increase in TFR due to child mortality) remained roughly constant over the three surveys. The cohort estimate of the replacement rate  $R$  (calculated as  $\Delta TFR/D$ ) rose substantially, however, from about 0.3 in NFHS-1 to about 0.5 in NFHS-3, reflecting both fertility decline, mortality decline, and increased control over fertility through increased use of contraception and abortion. With fertility decline, women have more time to replace dead children as age at completion of childbearing declines. The cohort estimates of mean  $A_d$  indicate a decline of somewhat more than a year over three surveys. This occurred because the reduction of births (and therefore early childhood deaths) was concentrated at higher parities and older ages of

women (for evidence of these reductions, see International Institute for Population Sciences (IIPS) 1995; International Institute for Population Sciences (IIPS) and ORC Macro 2000; International Institute for Population Sciences (IIPS) and Macro International 2007).

Table 6: Unadjusted and adjusted values of  $TFR_{WCM}$ ,  $TFR_{NCM}$ , mean number of child deaths per woman, child replacement rate, and mean age of women at child death: Cohort and period analysis, NFHS surveys

	NFHS-1		NFHS-2		NFHS-3	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
<b>COHORT</b>						
$TFR_{WCM}$	4.9	4.8	4.5	4.4	4.0	3.9
$TFR_{NCM}$	4.6	4.6	4.1	4.1	3.7	3.7
$\Delta TFR$	0.31	0.25	0.30	0.27	0.30	0.24
$D$	0.94	0.92	0.72	0.74	0.55	0.56
$R$	0.33	0.27	0.42	0.36	0.55	0.44
Mean $A_d$	29.3	29.2	28.6	28.5	28.0	27.7
<b>PERIOD</b>						
$TFR_{WCM}$	3.1	3.2	2.8	2.8	2.7	2.7
$TFR_{NCM}$	3.0	3.0	2.7	2.7	2.6	2.6
$\Delta TFR$	0.15	0.14	0.11	0.11	0.08	0.09
$D$	0.29	0.31	0.22	0.23	0.16	0.16
$R$	0.51	0.45	0.53	0.49	0.53	0.56
Mean $A_d$	28.4	28.5	27.9	27.9	28.2	28.0

Notes: For definitions of measures in this table, see the notes to Tables 3 and 4. The unadjusted cohort  $R$  for NFHS-1 in Table 6 is not exactly the same as the unadjusted  $R$  in the validation analysis in Table 4, because the open parity interval used in the calculation of unadjusted  $R$  in Table 6 is chosen to be the same as the open parity interval used in the calculation of adjusted  $R$ , in order to assess more accurately the effect of the adjustment.

The period results in Table 6 are less dramatic, because the period of major decline in fertility and major increase in contraceptive use preceded NFHS-1. As the table shows, the period  $TFR$  declined by only about a third of a child over the three surveys (much less than the decline of 0.9 child in the cohort  $TFR$ ). The major decline in  $D$  explains why the period estimate of  $\Delta TFR$  declined over the three surveys. The modest decline in fertility and the modest increase in contraceptive use that accompanied the decline in  $\Delta TFR$  are probably the main reasons why the period estimate of  $R$  increased very little over the three surveys (for the trends in fertility and contraceptive use, see International Institute for Population Sciences (IIPS) 1995; International Institute for

Population Sciences (IIPS) and ORC Macro 2000; International Institute for Population Sciences (IIPS) and Macro International 2007).

Table 7 shows unadjusted and adjusted estimates of total fertility rates ( $TFR_{WCM}$ ) and replacement rates ( $R$ ) by socioeconomic characteristics for periods and cohorts, based on the three NFHS surveys. The discussion here focuses on the replacement rates  $R$ . In the results for cohorts, when the TFR is low,  $R$  tends to be high.  $R$  does not differ much between urban and rural.  $R$  tends to be higher for those with more education in NFHS-1 but higher for those with less education in NFHS-3.  $R$  is consistently higher for those with high wealth than those with medium or low wealth.  $R$  is also consistently much higher for Hindus than for Muslims. Differences in  $R$  between those who are not SC/ST and those who are SC/ST tend to be small. Differences in  $R$  between regions also tend to be small. In almost all cases, for reasons already discussed,  $R$  increases over the three surveys. In the case of education, however,  $R$  increases a great deal for illiterates and those with low education but changes hardly at all for those with high education. Those with high wealth have the highest replacement rate. Adjustment for the other socioeconomic variables usually has the effect of reducing differences in  $R$  by categories of the main socioeconomic variable.

For each category of each socioeconomic variable, the period estimate of  $R$  for each survey in Table 7 tends to be higher than the cohort estimate, and the trend tends to be flatter (especially the unadjusted trend), for reasons mentioned in the discussion of Table 6.

The more detailed results are not always easy to interpret, however. For example, it is not clear why both the unadjusted and adjusted cohort estimates of  $R$  for Muslims rose between NFHS-1 and NFHS-2 and then fell between NFHS-2 and NFHS-3. And it is not clear why the unadjusted period estimate of  $R$  for Muslims rose between NFHS-1 and NFHS-2 and then fell between NFHS-2 and NFHS-3, while the adjusted period estimate of  $R$  for Muslims fell between NFHS-1 and NFHS-2 and then rose between NFHS-2 and NFHS-3.

## CONCLUSION

Although the underlying methodology is complex, involving the construction of multivariate 5-dimensional global life tables of fertility (the dimensions being age, parity, duration in parity, and two dimensions pertaining to child mortality), final results are in the form of simple bivariate tables and graphs that are easy to understand by not only demographers and statisticians but also policy makers and program managers in the population field. The global life table is multivariate in the sense that any given measure derived from it can be tabulated by values or categories of one socioeconomic variable while holding constant the other socioeconomic variables included in the underlying discrete-time survival models (one model for each parity transition). The global life table can be calculated from either cohort data or period data. Validation of the methodology

Table 7: Comparison of unadjusted and adjusted total fertility rates ( $TFR_{WCM}$ ) and child replacement rate by categories of socioeconomic predictor variables: Cohort and period analysis, India's NFHS surveys

Predictor Variables	$TFR_{WCM}$						Replacement Rate ( $R$ )					
	Unadjusted			Adjusted			Unadjusted			Adjusted		
	NFHS			NFHS			NFHS			NFHS		
	1	2	3	1	2	3	1	2	3	1	2	3
<b>COHORT</b>												
<b>Residence</b>												
Urban	4.2	3.8	3.4	4.6	4.1	3.9	0.39	0.42	0.46	0.27	0.34	0.40
Rural	5.1	4.7	4.3	4.9	4.5	4.0	0.31	0.40	0.54	0.27	0.39	0.46
<b>Education</b>												
Illiterate	5.2	5.0	4.6	5.0	4.7	4.3	0.28	0.36	0.47	0.25	0.37	0.43
Low	4.5	4.2	3.7	4.6	4.3	3.8	0.31	0.42	0.47	0.29	0.37	0.42
High	3.3	3.2	2.8	4.4	3.8	3.3	0.39	0.39	0.38	0.34	0.37	0.38
<b>Wealth Index</b>												
Low	5.4	5.0	4.8	4.9	4.2	4.1	0.27	0.35	0.44	0.25	0.32	0.39
Medium	4.9	4.6	3.9	4.9	4.5	3.9	0.22	0.33	0.41	0.22	0.31	0.38
High	4.1	3.7	3.1	4.6	4.3	3.7	0.44	0.48	0.53	0.34	0.48	0.52
<b>Religion</b>												
Hindu	4.8	4.3	3.9	4.8	4.3	3.9	0.35	0.44	0.58	0.28	0.38	0.46
Muslim	5.9	5.8	5.1	5.8	5.5	5.0	0.18	0.46	0.33	0.11	0.28	0.16
<b>SC/ST</b>												
Yes	5.2	4.8	4.5	4.9	4.2	4.1	0.30	0.38	0.50	0.26	0.38	0.41
No	4.8	4.3	3.8	4.8	4.4	3.9	0.34	0.43	0.56	0.27	0.37	0.45
<b>Region</b>												
North	5.4	5.2	4.7	5.2	5.0	4.6	0.34	0.39	0.53	0.29	0.34	0.43
East	5.0	4.6	4.3	4.9	4.5	4.1	0.28	0.43	0.48	0.24	0.41	0.44
South	4.3	3.8	3.1	4.4	3.9	3.2	0.27	0.36	0.39	0.23	0.35	0.41
West	4.5	3.8	3.4	4.7	4.0	3.8	0.31	0.40	0.43	0.28	0.35	0.33
<b>PERIOD</b>												
<b>Residence</b>												
Urban	2.6	2.3	2.1	3.1	2.7	2.5	0.50	0.53	0.53	0.46	0.48	0.56
Rural	3.4	3.0	2.9	3.2	2.9	2.7	0.48	0.50	0.50	0.44	0.51	0.57
<b>Education</b>												
Illiterate	3.8	3.5	3.5	3.5	3.2	3.1	0.43	0.42	0.41	0.41	0.43	0.47
Low	2.9	2.8	2.7	2.9	2.8	2.5	0.48	0.55	0.63	0.44	0.52	0.61
High	2.3	2.2	2.1	2.9	2.6	2.5	0.52	0.56	0.57	0.52	0.53	0.63
<b>Wealth Index</b>												
Low	3.7	3.3	3.3	3.4	3.0	2.9	0.42	0.40	0.41	0.43	0.41	0.45
Medium	3.2	2.8	2.5	3.1	2.7	2.5	0.45	0.54	0.54	0.45	0.53	0.54
High	2.4	2.1	1.9	2.9	2.6	2.4	0.49	0.57	0.57	0.44	0.54	0.70
<b>Religion</b>												
Hindu	3.1	2.5	2.6	3.1	3.8	2.6	0.53	0.54	0.55	0.46	0.34	0.56
Muslim	4.0	3.0	3.2	4.1	4.9	3.4	0.42	0.49	0.45	0.34	0.27	0.43
<b>SC/ST</b>												
Member	3.5	3.0	3.0	3.1	2.8	2.8	0.45	0.44	0.52	0.41	0.44	0.58
Other	3.0	2.7	2.5	3.1	2.8	2.6	0.52	0.56	0.52	0.46	0.52	0.55
<b>Region</b>												
North	3.8	3.4	3.1	3.8	3.4	3.2	0.52	0.53	0.47	0.46	0.48	0.47
East	3.1	2.7	2.8	3.0	2.6	2.7	0.42	0.48	0.47	0.41	0.52	0.52
South	2.5	2.2	2.0	2.7	2.4	2.2	0.30	0.31	0.29	0.41	0.46	0.54
West	2.8	2.6	2.3	3.1	2.9	2.7	0.48	0.46	0.59	0.45	0.38	0.62

using cohort data from India's first National Family Health Survey (NFHS-1) shows that, where comparisons can be made, global life table estimates of fertility and child mortality measures of interest agree closely with corresponding estimates derived directly from birth histories and from application of earlier versions of the methodology that do not incorporate child mortality in the set of predictor variables.

The principal measure of interest derived from this global life table is the replacement rate  $R$ , which is defined as the incremental increase in the TFR due to a one-child increase in the number of child deaths per woman,  $D$ , where the value of  $D$  is also derived from the global life table.  $R$  is calculated from India's three National Family Health Surveys. In the application of the methodology to India's three National Family Health Surveys, Results show that replacement of dead children is incomplete. In the cohort analysis, unadjusted values of  $R$  increased from 0.33 to 0.42 to 0.55 over the three surveys. In the case of adjusted values of  $R$ , the increase is smaller, from 0.27 to 0.36 to 0.44 over the three surveys. In the period analysis, unadjusted values of  $R$  increased from 0.51 to 0.53 to 0.53, and the adjusted values of  $R$  increased from 0.45 to 0.49 to 0.56. The values of  $R$  tend to be higher in the period analysis, because women's exposure to risk of both birth and child death is more recent in the period analysis than in the cohort analysis, leading to more volitional replacement of dead children due to earlier completion of childbearing (as a consequence of falling fertility) and higher rates of contraception and abortion. The increase in  $R$  over the three surveys is smaller in the period analysis than in the cohort analysis because the period of rapid increase in contraceptive use preceded the earliest five-year period (1988–92) considered in the period analysis.

A more general conclusion is that when fertility is high, replacement is relatively low (due not only to a low level of volitional replacement but also to non-volitional replacement as a consequence of truncated breastfeeding, leading to early resumption of ovulation after an early childhood death); and that when fertility is low, replacement is relatively high, due to increased volitional replacement.

## **ACKNOWLEDGMENTS**

Support for this research was provided by Grant 1R01HD057038 from the U.S. National Institutes of Health and by a grant obtained by the Nihon University Population Research Institute from the “Academic Frontier” Project for Private Universities: matching fund subsidy from MEXT (Japan Ministry of Education, Culture, Sports, Science and Technology), 2006–2010. I also thank Robert Retherford for many helpful comments on earlier drafts.

## 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.
- Easterlin R.A. 1975. An economic framework for fertility analysis. *Studies in Family Planning*, Vol. 6, No. 3: 54–63.
- Eini-Zinab, H. 2010a. *The effect of child mortality on fertility in India*. Ann Arbor, MI: ProQuest LLC.
- Eini-Zinab, H. 2010b. *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.  
<http://www.measuredhs.com/pubs/pdf/WP70/WP70.pdf>
- Feeney, Griffith, and Jingyuan Yu. 1987. Period parity progression measures of fertility in China. *Population Studies* 41:77–102.
- 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.
- International Institute for Population Sciences (IIPS) and Macro International. 2007. *National Family Health Survey (NFHS-3), 2005–06: India: Volume I*. Mumbai: IIPS.
- International Institute for Population Sciences (IIPS) and ORC Macro. 2000. *National Family Health Survey (NFHS-2), 1998–99: India*. Mumbai: IIPS.
- International Institute for Population Sciences (IIPS). 1995. *National Family Health Survey (MCH and Family Planning), India 1992–93*. Bombay: IIPS.
- 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. 1985. A theory of marital fertility transition. *Population Studies*. 39: 249–268.

- 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*. East-West Center Working Paper 122. Honolulu: East-West Center.  
<http://www.eastwestcenter.org/fileadmin/stored/pdfs/popwp122.pdf>
- Rutstein, Shea O., and K. Johnson. 2004. *The DHS Wealth Index*. DHS Comparative Reports 6. Calverton, MD: ORC Macro.