WORKING PAPERS

EAST-WEST CENTER WORKING PAPERS

EAST-WEST CENTER

The East-West Center is an education and research organization established by the U.S. Congress in 1960 to strengthen relations and understanding among the peoples and nations of Asia, the Pacific, and the United States. The Center contributes to a peaceful, prosperous, and just Asia Pacific community by serving as a vigorous hub for cooperative research, education, and dialogue on critical issues of common concern to the Asia Pacific region and the United States. Funding for the Center comes from the U.S. government, with additional support provided by private agencies, individuals, foundations, corporations, and the governments of 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: ewcbooks@EastWestCenter.org
Website: www.EastWestCenter.org

EAST-WEST CENTER WORKING PAPERS

EAST-WEST CENTER

Population and Health Series

No. 119, October 2006 (revised March 2008)

Multivariate Analysis of Parity Progression-Based Measures of the Total Fertility Rate and Its Components Using Individual-Level Data

Robert D. Retherford, Naohiro Ogawa,
Rikiya Matsukura, and Hassan Eini-Zinab

Robert D. Retherford is Coordinator of Population and Health Studies at the East-West Center in Honolulu, Hawaii.

Naohiro Ogawa is Director of the Nihon University Population Research Institute in Tokyo, Japan.

Rikiya Matsukura is Staff Researcher at the Nihon University Population Research Institute in Tokyo, Japan.

Hassan Eini-Zinab is Graduate Assistant in Population and Health Studies at the East-West Center in Honolulu, Hawaii.

Revised version of a paper with the same title, presented at the annual meeting of the Population Association of America, Los Angeles, March 29–April 1, 2006.

TABLE OF CONTENTS

ABSTRACT

This paper develops multivariate methods for analyzing (1) effects of socioeconomic variables on the total fertility rate and its components and (2) effects of socioeconomic variables on the trend in the total fertility rate and its components. For the multivariate methods to be applicable, the total fertility rate must be calculated from parity progression ratios (PPRs), pertaining in this paper to transitions from birth to first marriage, first marriage to first birth, first birth to second birth, and so on. The components of the TFR include PPRs, the total marital fertility rate (TMFR), and the TFR itself as measures of the quantum of fertility, and mean and median ages at first marriage and mean and median closed birth intervals by birth order as measures of the tempo or timing of fertility. The multivariate methods are applicable to both period measures and cohort measures of these quantities. The methods are illustrated by application to data from the 1993, 1998, and 2003 Demographic and Health Surveys (DHS) in the Philippines.

This paper develops multivariate methods for estimating effects of socioeconomic predictor variables on the total fertility rate (TFR) and on the trend in the TFR. The methods utilize individual-level survey data and are applicable to both period measures and cohort measures of the TFR. The analysis of effects of socioeconomic variables on the trend in the TFR requires two or more surveys of the same population at different times.

The TFR is usually defined as the number of births that a woman would have by age 50 if, hypothetically, she lived through her reproductive years experiencing the age-specific fertility rates (ASFRs) that prevailed in the population in the particular calendar year. The TFR so defined is calculated by summing ASFRs (births per woman per year at each age) between the ages of 15 and 50. For the multivariate methods in this paper to be applicable, however, the TFR must be calculated from parity progression ratios (PPRs). A woman's parity is defined in the usual way as the number of children she has ever borne, but with parity zero subdivided into two states: never-married with no children and ever-married with no children. Parity progression ratios (PPRs) are the fractions of women who progress from their own birth to first marriage, from first marriage to first birth, from first birth to second birth, and so on. The PPRs so obtained are aggregated to a TFR and a total marital fertility rate (TMFR). TFR, TMFR, and PPRs are measures of the quantum of fertility. The multivariate methods are also applicable to measures of the tempo or timing of first marriage and births, as measured by mean and median ages at first marriage and mean and median closed birth intervals by birth order.

We focus on the TFR calculated from PPRs (TFR_{ppr}) instead of the TFR calculated from ASFRs (TFR_{asfr}) for several reasons: The first is that a multivariate method for analyzing factors affecting TFR_{asfr} calculated from individual data has already been developed and applied by Schoumaker (2004), who used Poisson regression for this purpose. The second reason is that, from an explanatory point of view, age-specific fertility rates are not ideal measures of the components of the total fertility rate. A woman's decision about whether to have a next birth does not depend primarily on her age. More important considerations are her marital status, time elapsed since marriage if she is married but does not yet have any children, time elapsed since her last birth if she already has children, and the number of children that she already has. The TFR calculated from PPRs takes all these considerations into account. Henceforth in this paper, "TFR" and "TMFR" refer to the total fertility rate and the total marital fertility rate calculated from PPRs, whether for periods or cohorts. We use a multivariate discrete-time survival model—the complementary log-log (CLL) model—to model parity progression. Because the CLL model was originally developed for application to cohort data, its application here to period data, yielding a multivariate analysis of the period TFR and its components, is the most innovative aspect of the paper.

By way of illustration, the methods are applied to both period and cohort data from three demographic and health surveys (DHS) undertaken in the Philippines in 1993, 1998, and 2003. Period measures are estimated for the 5-year period before each survey, and cohort measures are based on the earlier reproductive experience of women age 40–49 at the time of each survey. A ten-year age cohort is used instead of a five-year age cohort (such as women age 40–44 or 45–

49) in order to base the cohort analysis on a larger number of cases. Because the surveys are five years apart, the ten-year cohorts overlap from one survey to the next.

In the Philippines surveys, some regions were over-sampled, so weights must be used to restore representativeness. The over-sampled regions were more rural than average, so that, in effect, the surveys over-sampled rural. The design of the three surveys is described in more detail in the basic survey reports, which include questionnaires and more detailed information about sampling procedures (Philippines National Statistics Office and Macro International 1994; Philippines National Statistics Office, Philippines Department of Health, and Macro International 1999; Philippines National Statistics Office and ORC Macro 2004).

PARITY PROGRESSION-BASED MEASURES OF THE TFR AND ITS COMPONENTS

We define the following notation for PPRs and the parity transitions to which they refer:

p_B PPR for transition from a woman's own birth to her first marriage (B–M)
p_M PPR for transition from first marriage to first birth (M–1)
p_1 PPR for transition from first to second birth (1–2)
p_2 PPR for transition from second to third birth (2–3)
.
.
.
P_8 PPR for transition from eighth to ninth birth (8–9)
P_{9+} PPR for transition from ninth or higher-order birth to next higher-order birth (9+ to 10+)

The choice of a cutoff for the open-ended parity category depends on the overall level of fertility and the size of the sample, which together determine the parity at which one starts to run out of higher-order births in the sample survey. In this section, for purposes of explaining methodology, we assume a cutoff at 9+.

PPRs are calculated from life tables. In general, the life table method is appropriate when the input data indicate time elapsed between a starting event and a terminating event. The generic term for a terminating event is "failure," and we use this term throughout this paper. In the case of p_B, the starting event is a woman's own birth and the terminating event, or "failure", is her first marriage if a first marriage occurs. In the case of $p_M, p_1, \ldots, p_{9+},$ the starting event is either a first marriage or a birth of a particular order, "failure" is a next birth, and time elapsed since the starting event is referred to as duration in parity. Consistent with demographic usage, we refer to a birth-to-first-marriage life table also as a nuptiality table.

Because the number of first marriages that occur before age 10 or after age 40 is very small in the Philippines, we start our nuptiality tables at age 10 (instead of birth) and end them at age 40. We continue to refer to this transition, however, as birth to first marriage (B–M). Time in the nuptiality table ranges from 0 years (corresponding to age 10) to 30 years (corresponding to age 40). In the case of subsequent parity transitions $p_M, p_1, \ldots, p_{9+},$ the number of births that occur after 10 years of duration in parity is negligible, so we terminate life tables for these transitions at 10 years. Time in these life tables therefore ranges from 0 to 10 years.

A PPR is calculated from a life table by subtracting the proportion "surviving" at the end of the life table from one, yielding the proportion who "fail" by the end of the life table.

From the life table for each parity transition, we can also compute a mean failure time and a median failure time. In the case of the nuptiality table, the mean and median failure times (when added to 10, the age at the start of the nuptiality table) are measures of mean and median ages at first marriage. In the case of the life tables for higher-order parity transitions, the mean and median failure times (in years) are measures of mean and median closed birth intervals. The medians so calculated are true medians, based on all failures that occur over the course of the life table. Because of the problem of age truncation at time of survey in the case of cohort estimates, DHS survey reports define medians differently, as the duration in parity by which half of the starting cohort experience failure.

Once PPRs have been calculated using the life table method, TFR is calculated from the PPRs as

$$\begin{aligned}\mathrm{TFR} = {} & p_B\, p_M + p_B\, p_M\, p_1 + p_B\, p_M\, p_1\, p_2 + p_B\, p_M\, p_1\, p_2\, p_3 + p_B\, p_M\, p_1\, p_2\, p_3\, p_4 \\ & + p_B\, p_M\, p_1\, p_2\, p_3\, p_4\, p_5 + p_B\, p_M\, p_1\, p_2\, p_3\, p_4\, p_5\, p_6 + p_B\, p_M\, p_1\, p_2\, p_3\, p_4\, p_5\, p_6\, p_7 \\ & + p_B\, p_M\, p_1\, p_2\, p_3\, p_4\, p_5\, p_6\, p_7\, p_8 + p_B\, p_M\, p_1\, p_2\, p_3\, p_4\, p_5\, p_6\, p_7\, p_8\, p_{9+} /(1 - p_{9+}) \qquad (1)\end{aligned}$$

The term $p_B\, p_M$ is the expected number of first births, the term $p_B\, p_M\, p_1$ is the expected number of second births, and so on. As explained by Feeney (1986), the term $p_{9+} /(1 - p_{9+})$ is obtained by assuming that p_9 and all higher-order PPRs equal p_{9+} and pulling out a geometric series. (Recall that if r is a positive number less than one, the geometric series $r + r^2 + r^3 + ... = r/(1 - r)$.)

The formula for TMFR is the same as the formula for TFR in equation (1), except that p_B is set equal to one.

In populations where a substantial proportion of births occur outside of marriage, an alternative approach would be to combine the first two parity transitions, B–M and M–1, into a single parity transition, 0–1,with p_0 defined as the fraction of women who progress from their own birth to a first birth. In our illustrative application to the Philippines, a substantial fraction of births occur in non-formalized unions. The three DHS surveys for the Philippines treat the first non-formalized union as a first marriage, however, and we also take this approach. We therefore retain the B–M and M–1 transitions in our analysis of these surveys.

Despite treating non-formalized unions in the same way as formalized marriages, there are still some births reported by ever-married women as having occurred before first marriage (i.e., before first formalized marriage or first non-formalized union), and there are also some births reported by never-married women, from whom birth histories were also collected. We refer to these births simply as premarital births. In the analysis, we do not exclude women who had a premarital birth. Instead, we treat all such women as newly married at the time of their first birth, by coding or re-coding date of first marriage back to the date of first premarital birth. This coding and re-coding introduce small biases in the estimates of mean age at first marriage and mean closed birth interval for the M–1 transition (only 6–8 percent of births were coded or re-coded in this way), but very little or no bias in the estimates of PPRs, median age at first marriage, median closed birth intervals, TFR, and TMFR.

MULTIVARIATE METHODOLOGY

Choosing a multivariate survival model

Because PPRs are derived from life tables, they can be modeled in a multivariate way using a multivariate survival model. It is useful to think of such a model as a multivariate life table from which PPRs and mean and median failure times can be calculated. Because TFR and TMFR can be calculated from the multivariate PPRs, TFR and TMFR can also be modeled in a multivariate way.

A number of multivariate survival models are available. We need a model that handles time-varying predictor variables and time-varying effects of predictor variables. Residence, for example, is properly specified as a time-varying predictor variable if some women move from a rural area to an urban area as they move through the life table. The effect of residence is also properly specified as time-varying if, as is usually the case, the effect of urban, relative to rural, is to lower the risk of first marriage at the younger reproductive ages and raise it at the older reproductive ages as a result of greater postponement of marriage in urban areas. If the effect of residence varies with time in this way, a proportional hazards model is not appropriate, because in a proportional hazards model the effect of urban, relative to rural, on the risk of first marriage is constrained to be constant over time in the life table.

Effects are time-varying not only for progression to first marriage but also for progression to higher-order parities. This is so not only because births may be postponed, but also because birth intervals (except for the interval between first marriage and first birth) tend not to change much as fertility falls (Pathak et al. 1998). This implies that the effect of residence on birth intervals can be small while its effect on PPRs is large. This is impossible to model with a proportional hazards model of parity progression. For example, in a proportional hazards model, if the probability of failure by the end of the life table is lower for urban than for rural, mean and median failure times must be higher for urban than for rural.

We also need a survival model that can handle left-censoring as well as right-censoring so that we can fit the model to period data. That is, we need to be able to censor not only the part of an individual's exposure that occurs after the period (right-censoring) but also the part that occurs before the period (left-censoring).

For our purposes, the survival model, when fitted to data, must also yield a baseline hazard function, so that we can estimate not only the effects of predictor variables on the risk of failure (as measured by the coefficients of the predictor variables) but also the model-predicted risk of failure itself (i.e., the hazard function on the left side of the model equation) for specified values of the predictor variables. Only then can we calculate predicted values of life table parameters such as PPRs and mean and median failure times for specified values of the predictors. This point will become clearer in the following paragraphs, which show the model equations.

One possible candidate for our multivariate survival model is the Cox model (Cox 1972). This model is usually stated in the form of a continuous-time proportional hazards model, although the model can also handle, up to a point, both time-varying predictors and time-varying effects of predictors, in which case the model is no longer proportional. Cox's continuous-time proportional hazards model is specified as

$$h_i(t) = h_0(t)\ \exp[b_1\, x_{i1} + \ldots + b_k\, x_{ik}] \tag{2}$$

where i denotes the i^{th} individual, t denotes continuous time in the life table, x_j ($j = 1, 2, ..., k$) is a set of k predictor variables (also called covariates), b_j ($j = 1, 2, ..., k$) is the set of coefficients of those predictors, $h_i(t)$ denotes the hazard rate for the ith individual at time t, and $h_0(t)$ is the baseline hazard function defined when all predictors have a value of zero. The continuous-time hazard rate $h_i(t)$ is defined as the individual's probability per unit time of experiencing failure in an infinitesimally small time interval centered on time t. A continuous-time hazard rate therefore has the dimensions of failures per person per unit time.

The Cox proportional hazards model is often stated alternatively in log-linear form as

$$\log h_i(t) = a_t + b_1\, x_{i1} + \ldots + b_k\, x_{ik} \tag{3}$$

where $a_t = \log h_0(t)$. As always in statistical models, logarithms are to the base e.

In equation (2), the exponential term is constant over time t in the life table, and that is what makes the model proportional. (Recall the definition of proportionality: two variables X and Y are proportional if $Y = kX$ for all values of X and Y, where k is the constant of proportionality. In equation (2), variation in $h_i(t)$ and $h_0(t)$ refers to variation over time t in the life table, and the exponential term, which does not vary over time, is the constant of proportionality.) The constant term in equation (2) is specified as an exponential function because the multiplicative effect of the predictors must be a positive number, and the function $\exp(x) \equiv e^x$ is defined and positive for all values of x and ranges over all positive real numbers. (Other functions, such as 3^x, could also be used, but e^x has mathematical properties that make it easier to work with.) In the exponential term in equation (2), not only the coefficients of the predictors but also the predictors themselves are time-invariant. Only then is the model proportional.

The continuous-time Cox model is fitted by the method of partial likelihood. The baseline hazard function $h_0(t)$ cancels out and does not appear in the likelihood function—hence the word "partial." Because of this, the partial likelihood method yields estimates of the coefficients of the predictors but not an estimate of the baseline hazard function $h_0(t)$ (equivalently, the term a_t in equation (3)). The output from the partial likelihood procedure is inputted into a second maximum likelihood procedure to obtain the baseline hazard function $h_0(t)$ (Allison 1995, p. 165). This second procedure does not work, however, when one or more predictor variables or their effects are time-varying (as in our application), in which case the Cox model does not yield a baseline hazard function. Because we need the baseline hazard function in order to calculate predicted values of the hazard function for specified values of the predictors (the necessity of this baseline hazard function is evident from equations (2) and (3)), the Cox model is not suitable for our purposes. A multivariate survival model that is suitable is the complementary log-log (CLL) model, which we consider next.

The complementary log-log (CLL) model

Basic form of the model

The CLL model is a discrete-time survival model. The general form of the model is

$$\log[-\log(1-P_{it})] = a_t + b_1 x_{i1} + \ldots + b_k x_{ik} \tag{4}$$

where i denotes the i^{th} observation, t is a counter variable denoting the t^{th} life table time interval ($t = 1, 2, \ldots$), P_{it} is the discrete probability of failure during the t^{th} life table time interval, a_t is a function of t (usually an unspecified function, in the sense of not having a particular functional form), and predictors and coefficients are as defined in the Cox model in equations (2) and (3). Equation (4) can be written more compactly as

$$\log[-\log(1-P_{it})] = a_t + \mathbf{bx} \tag{5}$$

where $\mathbf{b}$ is a vector of coefficients, $\mathbf{x}$ is a vector of predictor variables, and $\mathbf{bx}$ is the dot product of $\mathbf{b}$ and $\mathbf{x}$. The model is fitted by the method of maximum likelihood (Prentice and Gloeckler 1978), not partial likelihood, and therefore yields estimates of a_t as well as the coefficient vector $\mathbf{b}$.

In equations (4) and (5), the life table time intervals may be of variable length. If, however, the intervals are uniformly one time unit in length (as is assumed henceforth in this paper), then $t-1$ can be interpreted as exact time at the start of the interval to which P_t pertains. We can then re-label P_1 as P_0 and, more generally, P_t as P_{t-1}. The re-labeled P_t function, for $t = 0, 1, \ldots,$ will be used later when life table calculations are discussed in more detail. For now, we will stay with the original definition of P_t, defined for $t = 1, 2, \ldots$.

P_t is often called the discrete hazard, but it should be noted that P_t is defined quite differently from the continuous-time hazard $h(t)$ in the Cox model. In the Cox model, $h(t)$ is defined as the probability of failure per unit time, evaluated at time t, whereas P_t is defined as the probability that failure will occur in the t^{th} discrete time interval, whatever its length. If the interval is one time unit in length, the value of P_t and the average value of $h(t)$ over the interval will usually be close to each other but not quite identical. If the interval is more than one time unit in length, P_t and the average value of $h(t)$ over the interval can be very different.

If one solves equation (5) for P_t, one obtains an alternative form of the CLL model,

$$P_t = 1 - \exp[-\exp(a_t + \mathbf{bx})] \tag{6}$$

The right side of equation (6) specifies the functional form for the discrete hazard P_t, and this functional form of P_t is called the *link function*—in this case the CLL link function. Other link functions, such as the logit link function, are also possible. In this regard, it may be noted that the first part of the derivation of the log-likelihood function for a discrete-time survival model uses P_t without specifying the functional form of P_t. The second part of the derivation specifies the link function. Because of this, computer programs for estimating discrete-time survival models also require specification of the link function (Allison 1982; 1995, ch. 7).

Although it is not obvious, the CLL model is derived from the Cox proportional hazards model and therefore is itself a proportional hazards model. We consider a simplified derivation of the CLL model in equation (5) from the Cox model, pertaining to the case of one-year life table time intervals. The derivation begins with $\log[-\log(1-P_t)]$ and makes the substitutions $P_t = [S(t-1)-S(t)]/S(t-1)$ and $S(t) = [S_0(t)]^{\exp(\mathbf{bx})}$, where P_t denotes the probability of failure between exact times $t-1$ and t conditional on survival to time $t-1$, $S(t)$ denotes the unconditional probability of surviving to exact time t, and $S_0(t)$ denotes the value of $S(t)$ when all of the predictor variables equal zero. After these two substitutions and some algebraic manipulation, one obtains

$$\log[-\log(1-P_t)] = \log[-\log(1-P_{0,t})] + \mathbf{bx} \qquad (7)$$

where $P_{0,t}$ denotes the baseline P_t function defined when all predictors equal zero. Equation (7) is then the same as equation (5), in which $a_t = \log[-\log(1-P_{0,t})]$.

In the above derivation, the substitution of $[S_0(t)]^{\exp(\mathbf{bx})}$ for $S(t)$ is what makes equation (5) a proportional hazards model, because the relationship $S(t) = [S_0(t)]^{\exp(\mathbf{bx})}$ is valid only for a proportional hazards model (Retherford and Choe 1993, pp. 194−195). As will be explained shortly, however, it is possible to "trick" the CLL model in equation (5) to handle non-proportionality in the form of time-varying predictor variables or time-varying effects of predictor variables.

As already mentioned, a major advantage of a discrete-time survival model, such as the CLL model, over the continuous-time Cox model is that the CLL model, when fitted to data, yields a baseline hazard function (the $P_{0,t}$ function in equation (7)). This is so even when the CLL model is tricked to include time-varying predictors and time-varying effects of predictors. The CLL model yields this additional information because the terms a_t (actually the terms from which the values of a_t are calculated, as explained below) in equation (5) remain in the log-likelihood equations and can therefore be estimated.

The CLL model is superior to the discrete-time logit model, inasmuch as coefficients of predictors in the CLL model, but not in the discrete-time logit model, have the same relative-risk interpretation as coefficients of predictors in the continuous-time Cox model, namely that a one-unit increase in a predictor variable multiplies the underlying continuous-time hazard $h_i(t)$ by $\exp(b)$, where b is the coefficient of the predictor and $\exp(b)$ is the relative risk (Allison 1995, ch. 7). This is so because the CLL model, but not the discrete-time logit model, is derived from the continuous-time Cox model. (Due to differences in how the continuous-time Cox model and the discrete-time CLL model are formulated and estimated, however, these two models, when specified with the same predictor variables and applied to the same data, generally yield estimates of the coefficient vector **b** that are close to each other but not quite identical.)

Expanded data set of person-year observations for each parity transition

A discrete-time survival model, such as the CLL model, is fitted not to the original “person sample” but instead to an “expanded sample” of person-year observations created from the original person observations. In our analysis, these persons are women. More specifically, each woman’s survival history (beginning with the starting parity for the parity transition under consideration) is broken down into a set of discrete time segments, which in our analysis are person-years, up to the year of failure or censoring. Person-years after the year of failure are excluded, as are person-years before the year of the starting event. Thus, in the expanded sample, “year” in a person-year observation refers to life table time t. This means that, in the case of the B−M transition, as many as 30 person-year observations are created from a single person observation, and that, in the case of higher-order transitions, as many as 10 person-year observations are created from a single person observation.

Variables attached to the original woman record are carried over to the person-year records created from the woman record. Additional variables assigned to the person-year records are YEAR (life table time $t = 1, 2, \ldots$), a variable that we call CALTIME indicating the calendar year in which the person-year observation is located (e.g., 1999), and the dummy variable FAILURE indicating whether failure occurred during that person-year of exposure (1 if yes, 0 if no).[1] The value of YEAR for a particular person-year observation is calculated as the difference between CALTIME and the calendar year in which the woman reached the starting parity (which in the case of the B−M transition is the calendar year in which the person reached age 10). The values of the variables for each person-year observation are the input data for fitting the CLL model. The input datum for the dependent variable is the value of FAILURE (1 if yes, 0 if no) rather than a value of P_t, which is unobservable. The other variables attached to each person-year observation, such as residence and education, are potential predictor variables. (See Allison 1995, ch. 7, for details on how to create the person-year data set.)

For each of the three Philippines surveys, a separate expanded data set of person-year observations is created for each parity transition in the period analysis and for each parity transition in the cohort analysis.[2]

Because the CLL model is applied to a person-year data set, it easily handles censoring—both right-censoring and left-censoring. Censoring normally means “lost to observation,” but one can also treat an observation as censored even when it is not, if doing so furthers the aims of analysis. In our analysis of period data, right-censoring pertains to that part of an individual’s exposure to risk of failure that occurs after the calendar time period of interest, and left-censoring pertains to that part of an individual’s exposure that occurs before the calendar time period. The

[1] In our application to Philippine DHS data, calendar years refer to years before the survey. Our labeling convention for years before the survey is illustrated by the 1993 survey: The year before this survey falls partly in 1993 and partly in 1992; but it falls mostly in 1992 and is therefore labeled 1992.

[2] Multiple births are included in the analysis. The birth order of each birth within a set of multiple births is arbitrarily specified. Birth intervals between multiple births are coded as zero.

CLL model's way of handling censoring is quite simple: In a period analysis, the expanded data set includes only those person-year observations located within the period of interest. In a cohort analysis, the expanded data set includes only those person-year observations created from women in the cohort of interest.

The censoring criteria are illustrated diagrammatically in Figures 1 and 2 for the period and cohort cases of progression from 10th birthday to first marriage. Whether the analysis is a period analysis or a cohort analysis depends entirely on whether the expanded data set corresponds to person-year observations located within a rectangle or a diagonal cohort corridor in the Lexis diagram, as shown in the two figures. If the shaded area is a rectangle, life table time should be thought of as extending vertically in the diagram, and if the shaded area is a cohort corridor, life table time should be thought of as extending diagonally. The mathematics of the procedure for fitting a CLL model to the data is the same in either case.

As already mentioned, the methodology is applicable to parity transitions B–M, M–1, 1–2, 2–3, 3–4, 4–5, 5–6, …, up to some open-ended parity interval. The creation of the expanded data set for the open parity interval requires further explanation, and for purposes of explanation we consider again the example of the transition from 9+ to 10+. The approach is to create separate expanded data sets for transitions 9–10, 10–11, ..., up to the transition from $k-1$ to k, where k is the highest parity attained by any woman in the survey. These separate expanded data sets are pooled to form the expanded data set for the transition 9+ to 10+. A woman can contribute person-year observations to more than one of the individual data sets that are pooled. For example, a woman who was parity 11 at the time of the survey contributes person-year observations to the expanded data sets for the transitions 9–10, 10–11, and 11–12. Pooling requires that the set of predictor variables attached to person-year observations in each of the pooled data sets for transitions 9–10, 10–11, ..., and $k-1$ to k be the same and have the same variable names. In our application to Philippines data, the maximum value of k considered is 15. Higher-order transitions have a negligible impact on the TFR and are ignored.

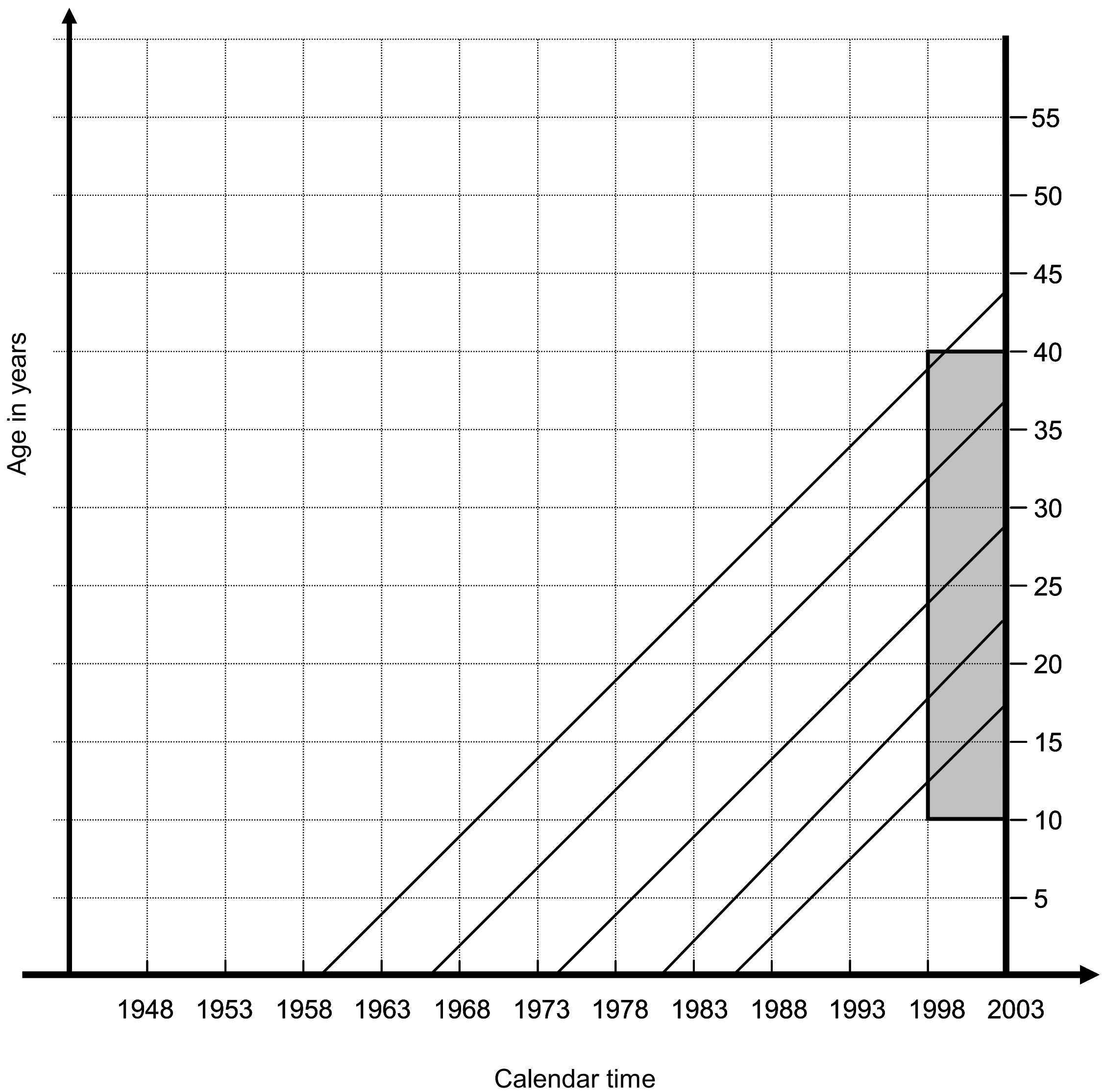
In the expanded data sets for the B–M transition, marriages occurring after age 40 are ignored. In the expanded data set for the M–1 transition, however, first marriages after age 40 (up to a maximum of 49) are included in the set of starting events. In the expanded data sets for the M–1 and higher-order transitions, all next births at durations 0–9, regardless of woman's age, are included in the set of terminal events. In the expanded data sets for the 1–2 and higher-order transitions, all births of the specified order, regardless of woman's age or duration in parity, are included in the set of starting events.

As discussed in more detail later, 96 expanded data sets are created for the analysis, some of which are pooled to form the data set for the open-ended parity transition.

Dummy variable specification of life table time

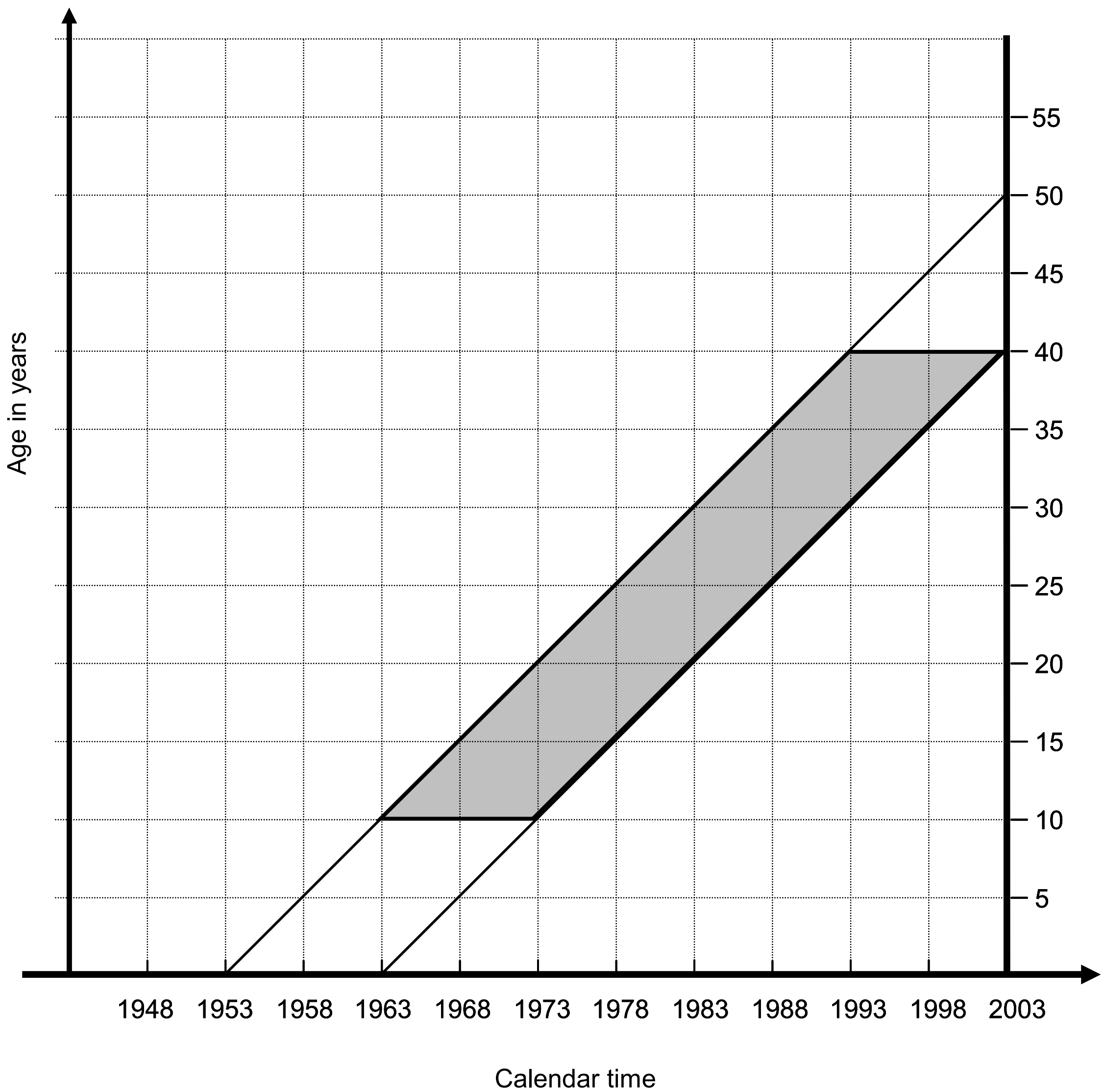
In our illustrative application to the Philippines data, life table time is modeled in two different ways: (1) a dummy variable specification and (2) a quadratic specification involving terms in t and t^2, where t is once again the counter variable $t = 1, 2, \ldots$. We consider the dummy variable specification first.

Figure 1: Lexis diagram illustrating censoring when setting up the expanded data set for calculating a multivariate period life table for progression from 10th birthday to first marriage, pertaining to the 5-year period preceding the 2003 survey



Notes: The shaded area, which is 5 years wide and 30 years high, represents the relevant period of exposure to the risk of first marriage. 45-degree lines are life-lines for particular individuals. Imagine that each life-line is divided into one-year segments, corresponding to person-years in the expanded data set. Person-years falling outside the shaded area are censored (or treated as censored) and not included in the expanded data set. Within the shaded area, the expanded data set includes only those person-years that occur up to the time of first marriage or censoring by reaching the survey date while still in the never-married state. If a first marriage occurs in a particular person-year within the shaded area, that person-year is also included in the expanded data set.

Figure 2: Lexis diagram illustrating censoring when setting up the expanded data set for calculating a multivariate cohort life table for progression from 10th birthday to first marriage, based on the 2003 survey



Notes: Within the shaded area, the expanded data set includes only those person-years up to the time of marriage or censoring by reaching age 40 while still in the never-married state. If a first marriage occurs in a particular person-year within the shaded area, that person-year is also included in the expanded data set.

The dependent variable P_t on the left side of equation (5) (or either of the equivalent equations (4) and (7)) is a function of t rather than a single value of P. Because of this, we can think of the model in equation (5) as representing a set of equations, one for each value of t. Equation (5) can thus be viewed as a multi-equation model. The units of analysis to which this multi-equation model pertains are persons (i.e., women who are at risk of either a first marriage or a next birth).

The model that is actually estimated, however, is a single-equation model that has the same form as equation (5), except that P_t is replaced by P, and the term a_t is replaced by an intercept and a set of dummy variables representing life table time intervals. This single-equation model is fitted to the expanded sample of person-year observations created from the original person observations. After the single-equation model is fitted, values of a_t are calculated from the estimates of the intercept and the coefficients of the dummy variables representing life table time intervals, thereby allowing the single-equation model to be rewritten in its original multi-equation form in equation (5).

By way of illustration, let us consider the single-equation model for progression to first marriage, in which there are 30 life table time intervals ($t = 1, 2, ..., 30$). Using the dummy variable specification of life table time, the single-equation model is

$$\log[-\log(1-P)] = a_{30} + c_1T_1 + c_2T_2 + ... + c_{29}T_{29} + \mathbf{bx} \tag{8}$$

where T_1, T_2, ..., T_{29} are dummy variables representing the first 29 life table time intervals (the 30th interval, for which $t = 30$, being the reference category), a_{30} is the intercept, and $c_1, c_2, ..., c_{29}$ are coefficients to be fitted to the data (i.e., to the expanded data set of person-year observations). Equation (8) specifies P rather than P_t, because inclusion of the subscript t would indicate one equation for each value of t instead of a single equation. In equation (8), time interval 1 is specified by $T_1 = 1$ and $T_2 = T_3 = ... = T_{29} = 0$. Time interval 2 is specified by $T_1 = 0$, $T_2 = 1$, and $T_3 = ... = T_{29} = 0$. And so on up to time interval 30 (the last interval), which is specified by $T_1 = T_2 = ... = T_{30} = 0$. Model fitting yields estimates of a_{30}, c_1, …, c_{29}, and the coefficient vector $\mathbf{b}$.

In equation (8), the intercept of the fitted model (which is the predicted value of $\log[-\log(1-P_{30})]$ when all predictors—including the dummy variables T_1, T_2, ..., T_{29}—are set to zero) is the same as a_{30} in equation (5) when t is set to 30, corresponding to the last time interval. The predicted value of $a_{29} = \log[-\log(1-P_{29})]$, with all the x_j and T_1, ..., T_{28} set to zero and T_{29} set to one, is $a_{30}+c_{29}$. More generally, $a_t = a_{30}+c_t$ for $t = 1, 2, ..., 29$. In this way, the single-equation model in equation (8), after being fitted to the expanded data set of person-year observations, can be rewritten in the same form as equation (5), which pertains to persons rather than person-year observations, with one equation for each value of t.

The dummy variable specification of life table time interval allows maximum flexibility in the way that a_t can vary over time. For this reason, the dummy variable specification of life table time interval in equation (8) is referred to as the *unrestricted* specification (unrestricted in the sense that a_t is not constrained to any particular functional form) (Allison 1995, ch. 7).

It might seem that, when the model in equation (8) is fitted to the data, standard errors of coefficients should be adjusted to take into account that the person-year observations created from a person record are not independent observations but instead are clustered. It has been shown, however, that adjustments for clustering are unnecessary for discrete-time survival models (including not only the CLL model but also the discrete-time logit model). The reason is that the log-likelihood function for the multi-equation model in equation (5), based on persons, and the log-likelihood function for the single-equation model in equation (8), based on person-year observations, are identical. Because of this, the models in equations (5) and (8) are equivalent, yielding identical estimates of coefficients, standard errors, and baseline hazard function when fitted to persons (in the case of equation (5)) or person-years created from those persons (in the case of equation (8)) (Allison 1982).

Quadratic specification of life table time

A potential problem with the unrestricted specification of life table time is that the computing algorithm for fitting the CLL model will not converge if there are any empty intervals (i.e., intervals in which there are no person-year observations) (Allison 1995, ch. 7). Non-convergence often arises at the higher parity transitions, where the expanded samples of person-year observations are smaller. In such cases the problem of non-convergence can be circumvented by using a quadratic specification of life table time; i.e., by replacing the dummy variables T_1, T_2, ..., T_{29} with t and t^2, as follows:

$$\log[-\log(1-P)] = a + c\,t + d\,t^2 + \mathbf{bx} \tag{9}$$

In our example of progression to first marriage, allowable values of t are $t = 1, 2, \ldots, 30$. This model is also fitted by procedures described by Allison (1995, ch. 7).

The non-convergence problem is the main reason why we specify life table time in years rather than months. Aggregation to years reduces the likelihood of no person-year observations in a life table time interval.

Non-convergence can also occur because of no person-year observations in at least one category of a socioeconomic predictor variable such as education. For example, few women with high education are found at the higher parities, in which case the high-education category for a particular parity transition may be empty. In this case a possible solution would be to combine some of the education categories at higher levels of education.

Non-convergence also occurs when one or more of the four cells in the 2x2 cross-classification of the dichotomous dependent variable FAILURE against a dichotomous predictor variable is empty. This cause of non-convergence is the most important reason for specifying life table time in years rather than months, as is evident from the following example: In the person-month data set for the 2−3 transition, there are no failures (next births) in the second month following the second birth; i.e., the cell in the cross-tabulation corresponding to FAILURE = 1 and $T_2 = 1$ is empty.

In the Philippines application, when we ran the CLL model with a dummy-variable specification of life table time without any socioeconomic predictors, we found 8+ to be the highest parity cutoff we could use for the open parity interval without running into convergence problems at least some of the time. We also had to use an 8+ cutoff when residence was the only socioeconomic predictor in the model. We had to use a 7+ cutoff when education was the only socioeconomic predictor in the model and when both residence and education were in the model. By contrast, when we used a quadratic specification of life table time, higher cutoffs were usually possible and were used.

Time-varying predictors

The CLL model easily handles time-varying predictor variables. In this context, "time-varying" refers to variation over life table time t (not calendar time). One simply assigns, where appropriate, different values of the predictor to different person-year observations created from a particular person observation. Although the value of a predictor can vary from one person-year to the next for a person, the CLL models in equations (8) and (9) assume only that the value of the predictor does not vary within a person-year. In other words, in the expanded sample of person-year observations, predictors are not time-varying, because the value of the predictor that is assigned to a person-year observation pertains to a particular value of t and therefore does not vary over time. In effect, the expansion of the person sample into a person-year sample converts time-varying predictors into time-invariant predictors.

Our illustrative application to Philippines DHS data includes only urban/rural residence and education as predictors. Both predictors are defined at the time of survey but not at earlier times, so in neither case can these predictors be treated as time-varying predictors. We are forced to treat them as time-invariant predictors. For example, if a woman was age 45 and urban at the time of survey, we are forced to assume (incorrectly in many cases) that she was also urban at all earlier ages, so that each person-year observation created from her person record is coded as urban.

In the Philippines analysis, residence is defined as a categorical variable with two categories: urban and rural. Education is defined as a categorical variable with three categories: less than secondary, some or completed secondary, and more than secondary. Henceforth we refer to these three education categories as low, medium, and high. Residence is specified by a dummy variable U, which is 1 if urban and 0 if rural. Education is represented by two dummy variables, M and H, where $M = 1$ if medium education and 0 otherwise, and $H = 1$ if high education and 0 otherwise. It follows that $(M, H) = (0, 0)$ for low education, $(1, 0)$ for medium education, and $(0, 1)$ for high education.

Even though U, M, and H are time-invariant predictors, interval-specific mean values of U, M, and H vary over time in the life table. (By "interval" here is meant life table time interval, indexed by the variable t.) In the period analysis, this is so because the group of women who reach a particular age during the period is not the same as the group of women who reach some other age during the period (although the two groups may overlap to some extent). For example, during the 5-year period immediately preceding any one of our three surveys, the women who had a 20th birthday during the period and the women who had a 35th birthday during the period are two completely different groups of women. In the case of the Philippines, these two groups differ substantially in population composition by residence and education, because the younger women tend to be more urban and more educated than the older women. In this regard, it should be noted that interval-specific mean values of the predictors for a particular time period, representing population composition, are calculated from the person-year observations in the expanded sample, not from the original woman records.

Even in the cohort analysis, interval-specific mean values of U, M, and H vary over time in the life table, because the expanded sample of person-year observations leaves room for "frailty" to operate. "Frailty" pertains to the effects of unobserved heterogeneity in the risk of failure in each life table time interval. Unobserved heterogeneity means that person-year observations at higher risk of failure are weeded out faster over the course of the life table. For example, in a life table for progression from third to fourth birth, rural persons have higher interval-specific risks of failure (fourth birth) than urban persons and are therefore weeded out faster than urban persons. This means that, in the expanded cohort data set for analyzing progression to fourth birth, interval-specific mean values of U tend to increase as life table time t increases. Similarly, interval-specific mean values of H tend to increase as life table time t increases.

Dummy variable specification of time-varying effects

The CLL model can also incorporate time-varying effects of predictors. Time-varying effects can be modeled in several ways. We consider first a dummy variable specification.

Suppose that, in our example of progression to first marriage, the predictor is urban/rural residence, specified by the dummy variable U. If the effect of urban/rural residence is not time-varying, the effect of residence on $\log[-\log(1-P_t)]$ is simply the coefficient of U, which we denote by b, which is constant over time in the life table. This is so regardless of whether U itself is time-varying. The effect of residence can be re-specified as time-varying by interacting U with the dummy variables representing life table time interval, resulting in an additional set of predictor variables that we can denote as $W_1 = U\,T_1$, $W_2 = U\,T_2$, ..., $W_{29} = U\,T_{29}$.

The effect of education is specified as time-varying in a similar fashion. Both M and H must be interacted with the dummy variables T_1, T_2, ..., T_{29}. The specification of this interaction requires the creation of the new variables $X_1 = MT_1$, $X_2 = MT_2$, ..., $X_{29} = MT_{29}$ with coefficients u_1, u_2, ..., u_{29}, and $Y_1 = HT_1$, $Y_2 = HT_2$, ..., $Y_{29} = HT_{29}$ with coefficients v_1, v_2, ..., v_{29}.

With the effects of residence and education specified in this way, the model in equation (5) becomes

$$\log[-\log(1-P_t)] = a_t + bU + d_1W_1 + \ldots + d_{29}W_{29} + fM + u_1X_1 + \ldots + u_{29}X_{29}$$

$$+ gH + v_1Y_1 + \ldots + v_{29}Y_{29} \qquad (10)$$

In equation (10), the terms containing U can be written as $bU+d_1W_1+d_2W_2+\ldots+d_{29}W_{29} = bU+d_1UT_1+d_2UT_2+\ldots+d_{29}UT_{29} = U(b+d_1T_1+d_2T_2+\ldots+d_{29}T_{29})$. It follows that the effect of a one-unit change in U is $b+d_1$ for the 1st time interval, $b+d_2$ for the 2nd time interval, ..., $b+d_{29}$ for the 29th time interval, and b for the 30th time interval. Thus, as long as $d_1, d_2, ..., d_{29}$ are not all zero, the effect of urban/rural residence on log[–log(1–P_t)] (and hence on P_t itself) is time-varying. Similarly, the effect of a change from low to medium education is $f+u_1$ for the 1st time interval, $f+u_2$ for the 2nd time interval, ..., $f+u_{29}$ for the 29th time interval, and f for the 30th time interval; the effect of a change from low to high education is $g+v_1$ for the 1st time interval, $g+v_2$ for the 2nd time interval, ..., $g+v_{29}$ for the 29th time interval, and g for the 30th time interval; and the effect of a change from medium to high education is the difference between the low-to-high effects and the low-to-medium effects.

When time-varying effects are incorporated into the CLL model in this way, it is the *effect* of residence—now represented by not only the coefficient of U but also the coefficients of $W_1, W_2, ..., W_{29}$ in the expression $b+d_1T_1+d_2T_2+...+d_{29}T_{29}$—that is time-varying, not the coefficients themselves, which are time-invariant. The same point applies to the effect of education and the coefficients of the education-related variables. Again, because of the dummy-variable specification of the life table time interval variable, this approach to time-varying effects achieves maximum flexibility in the way that time-varying effects are modeled. For this reason, this approach to modeling time-varying effects is also referred to as *unrestricted*.

At first blush, equation (10) is still a proportional hazards model. This is so because the computer does not see the time variation when it fits the model. It sees only person-year observations that are time-invariant, predictors that are time-invariant, and coefficients to be fitted that are time-invariant. Time variation is hidden in the definitions of sample observations (person-years) and in the definitions of the variables $W_1, W_2, ..., W_{29}$; $X_1, X_2, ..., X_{29}$; and $Y_1, Y_2, ..., Y_{29}$. In this way, we "trick" a model designed for time-invariant predictors and time-invariant effects into including time-varying predictors and time-varying effects. This means, among other things, that the model with time-varying predictors and time-varying effects is fitted in exactly the same way as the model with time-invariant predictors and time-invariant effects. (A somewhat similar situation arises in ordinary multiple regression when we trick a model that is linear in X into becoming nonlinear by introducing a new predictor variable $Z = X^2$.)

Quadratic specification of time-varying effects

A problem with the dummy variable specification of time-varying effects of predictor variables is that, as mentioned earlier, the algorithm for fitting the CLL model will not converge unless each of the four cells in the cross-classification of the dichotomous dependent variable FAILURE against each dichotomous predictor variable contains at least one person-year observation. This problem arises in the case of the dummy variables $W_1, W_2, ..., W_{29}, X_1, X_2, ..., X_{29}, Y_1, Y_2, ..., Y_{29}$. For example, consider the variable $Y_{29} = MT_{29}$ in the multivariate nuptiality analysis. This variable is 1 if the person-year observation has medium education and is age 39 (10 plus 29), and 0 otherwise. In the cross-classification of FAILURE against Y_{29}, it could easily be the case that there are no person-year observations in the cell for which FAILURE = Y_{29} = 1, because all persons with medium education who are going to get married may already have gotten married before age 39. A possible solution to this problem is to combine time intervals until the cross-classification has at least one person-year observation in each of the four cells, but this gets messy. A second, preferable solution is to use an alternative specification of time-varying effects that not only avoids this problem and but also requires many fewer predictor variables.

Again we use progression to first marriage as an example, this time with residence as the sole predictor and its effect modeled as time-varying. Instead of interacting U with the 29 dummy variables representing life table time intervals in order to model time-varying effects, we interact U with life table time t (the counter-variable analogue of the 29 dummy variables representing discrete life table time intervals) by creating the variables $Z_1 = Ut$ and $Z_2 = Ut^2$, and we add these variables to each person-year record. If we want to use a linear specification of the time-varying effect, we include the variable Z_1 in the set of predictors in the CLL model. If we want to use a quadratic specification, we include both Z_1 and Z_2.

Suppose that we use the quadratic specification, and suppose that the fitted coefficients of Z_1 and Z_2 are c and d. The right side of the model equation then includes the terms $bU+cZ_1+dZ_2 = bU + cUt+dUt^2 = U(b+ct+dt^2)$. The expression $(b+ct+dt^2)$ is now the "coefficient" of U, and this "coefficient," representing the effect of U, is a function of t and is therefore time-varying. At any given value of t, the effect of a one-unit change in U is to increase $\log[-\log(1-P_t)]$ by $b+ct+dt^2$ units. Equivalently, the effect is to multiply the underlying continuous-time hazard $h(t)$ by $\exp(b+ct+dt^2)$. Similar reasoning applies to the time-varying effects of M and H. Using the Philippines data, we experimented with the linear and quadratic specifications and found that the quadratic specification was always significantly better than the linear specification.

We therefore opted for the quadratic specification. The multi-equation model in equation (5) (based on persons) for progression to first marriage can then be written as

$$\log[-\log(1-P_t)] = a_t + U(b+ct+dt^2) + M(f+gt+ht^2) + H(j+kt+mt^2) \tag{11}$$

and the equivalent single-equation model in equation (8) (based on person-years) can be written as

$$\log[-\log(1-P)] = a_{30} + c_1T_1 + c_2T_2 + \ldots + c_{29}T_{29} + U(b+ct+dt^2) + M(f+gt+ht^2) + H(j+kt+mt^2) \tag{12}$$

(In the model that the computer sees and fits, however, the term $U(b+ct+dt^2)$ appears as $bU + cZ_1 + dZ_2$, where $Z_1 = Ut$ and $Z_2 = Ut^2$. The terms $M(f+gt+ht^2)$ and $H(j+kt+mt^2)$ also appear in analogously altered form.)

Note that life table time is specified in two different ways in the same model in equation (12) — both as 29 dummy variables $T_1, T_2, \ldots, T_{29}$ and as a counter variable t. This works because t and t^2 are hidden in the interaction variables, which in the case of residence are Z_1 and Z_2.

In the previous section that used dummy variable specifications of time-varying effects, we needed 90 coefficients to model the time-varying effects of residence and education (30 for residence and 60 for education). In the alternative approach using a quadratic specification, we need only 9 coefficients (3 for residence and 6 for education).

Equations (11) and (12) are our preferred models for analyzing progression to first marriage in the three Philippines DHS surveys. In cases where non-convergence was a problem, however, we were sometimes able to use the following equation in place of equation (12):

$$\log[-\log(1-P)] = a_{30} + c_1 t + c_2 t^2 + U(b+ct+dt^2) + M(f+gt+ht^2) + H(j+kt+mt^2) \quad (13)$$

Equation (13) illustrates two meanings of "quadratic specification". The first refers to the specification of life table time intervals, and the second refers to the specification of time-varying effects of predictor variables.

Weights

The three Philippines DHS survey samples for 1993, 1998, and 2003 are weighted samples. In each survey, sample weights are normalized so that the weighted number of cases is identical to the unweighted number of cases in the full DHS data set. In other words, the weights sum to the total survey sample size.

In our analysis, when the expanded data sets are created, the original weight for a woman carries over to the person-year observations created for that woman; i.e., the same weight is attached both to the original woman record and to each person-year record created from the original woman record. Each time a CLL model is fitted to an expanded data set (recall that we have 96 such data sets), however, it is important that the weights attached to the person-year records are re-normalized so that the re-normalized weights sum to the number of unweighted person-year observations in the particular expanded data set.

When calculating re-normalized weights for person-year observations in the expanded data set for a particular parity transition, we use the following notation pertaining to the particular expanded data set:

N	The number of unweighted person-year observations in the data set
w_i	The original weight attached to the i^{th} person-year observation in the data set
W	The sum of the w_i over the person-year observations in the data set
w_i^*	Re-normalized weight for the i^{th} person-year observation in the data set

We would like the weighted data set to sum to N (i.e., the re-normalized weights should sum to N). Re-normalized weights are accordingly calculated as

$$w_i^* = w_i(N/W) \quad (14)$$

When the w_i^* are summed over person-year observations in the data set, the result is $W(N/W) = N$, as desired.

In the case of the data set for the open parity interval, for example 9+ to 10+, the re-normalization of weights is done in the following way. One first creates the expanded data subsets for parity transitions 9–10, 10–11, ..., 14–15. One then pools these data subsets. Finally, one re-normalizes the weights in this pooled data set using equation (14). For the merge to be done properly, all variables carried over into the merged data set must have the same names in each and every data subset.

Calculating predicted values of the P_t function and derived life table from the fitted CLL model

Let us return to the example of progression to first marriage. Once we have fitted values of the terms a_t and the various coefficients in equation (11), we can predict values of log[–log(1–P_t)] on the left side of the equation for specified values of *U, M, H,* and *t*. For each value of *t*, we can then solve for the value of P_t, which is assumed to be the same for all persons with the specified values of the predictors. If there are 30 time intervals, as in the case of our nuptiality tables, there are 30 such equations, which are solved for 30 values of P_t for $t = 1, 2, ..., 30$. These 30 values of P_t constitute the discrete-time hazard function for women with the specified values of the socioeconomic predictors.

A baseline hazard function $P_{0,t}$ is obtained by setting all predictors (i.e., all socioeconomic predictors) equal to zero in the fitted model in equation (11). Equation (11) then reduces to

$$\log[-\log(1-P_{0,t})] = a_t \tag{15}$$

where the subscript 0 in $P_{0,t}$ denotes the baseline value of P_t with all predictors in equation (11) set to zero. Equation (15) represents 30 separate equations, one for each value of *t*. Each equation can be solved for P_t, yielding

$$P_{0,t} = 1 - \exp[-\exp(a_t)] \tag{16}$$

From equations (15) and (16) it is evident that the function a_t is a simple mathematical transformation of the baseline $P_{0,t}$ function, and *vice versa.*

In the more general case, for arbitrary values of the predictors,

$$P_t = 1 - \exp[-\exp(a_t + \mathbf{bx})] \tag{17}$$

The P_t function (evaluated with all predictors set either to zero or to other specified values of the predictors) determines an entire life table. In presenting calculation formulae for the life table measures of interest, we use the re-labeled P_t function, mentioned earlier, defined for $t = 0, 1, 2, \ldots$ instead of $t = 1, 2, 3, \ldots$). In the example of a life table for progression to first marriage, *t* ranges from 0 to 29, and *t* is interpreted as exact time at the start of the life table time interval ranging from exactly *t* to exactly *t*+1.

Given predicted values of P_t calculated from a fitted CLL model, values of the survivorship function $S(t)$ at exact time t are calculated sequentially as

$$S(0) = 1$$
$$S(t{+}1) = S(t)\,(1{-}P_t),\ \ t = 0, 1, ..., 29 \tag{18}$$

The unconditional probability of failure between t and t+1 is calculated as

$$f_t\ = S(t)\,P_t \tag{19}$$

The unconditional probability of failure by time t is calculated as

$$F(t) = 1{-}S(t) \tag{20}$$

The parity progression ratio is calculated (in the case of progression to first marriage) as

$$\text{PPR} = F(30) \tag{21}$$

The mean age at failure is calculated as

$$\text{Mean failure time} = \sum\ [f_t/F(30)](t + 0.5) \tag{22}$$

where the summation ranges from $t = 0$ to $t = 29$. The median failure time is calculated as

$$\text{Median failure time} = t\text{, such that } F(t)/F(30) = 0.5 \tag{23}$$

The mean and median failure times, when added to age 10 (woman's age at the start of the nuptiality table), are mean and median ages at first marriage.

P_t functions and life tables for progression from first marriage to first birth, first birth to second birth, second birth to third birth, and so on, defined at specified values of the predictor variables, are calculated in a similar manner, except that life tables have 10 one-year time intervals instead of 30 one-year time intervals.

Using equation (1), TFR for specified values of the predictor variables is then calculated from the PPRs for specified values of the predictor variables. TMFR is similarly calculated, with p_B set to one in equation (1).

Calculating unadjusted and adjusted values of the P_t function and derived life tables from the fitted CLL model

By "unadjusted" we mean "without controls", and by "adjusted" we mean "with controls".

To obtain unadjusted values of the P_t function and derived life table for each category of a predictor such as urban/rural residence, we run the CLL model with residence as the sole socioeconomic predictor variable with time-varying effects, with a quadratic specification of the time variation. (In the case of progression to first marriage, for example, the model equation is the same as equation (12) or (13) above, but with the education terms deleted.) Once the model is fitted, we use it to calculate two life tables, one for urban and one for rural, by alternatively setting U to 1 and 0 in the model equations. From these two life tables we calculate urban and rural values of the PPR and mean and median failure times. We refer to the life tables for urban and rural as *unadjusted life tables*, and to the values of the PPR and mean and median failure times calculated from these urban and rural life tables as *unadjusted values*.

To obtain adjusted estimates of the P_t function and derived life tables for urban and rural, we run the model again, this time with not only residence but also all the other predictors included in the model. (In our Philippines example, the only other predictor is education, but for the moment we shall speak in terms of the more general case of more than two socioeconomic predictors.) The other predictor variables serve as controls. The procedure is the same as in the unadjusted case, except that we hold the control variables constant at their interval-specific mean values when varying residence from urban to rural. By "interval" is meant life table time interval. For example, in the analysis of progression to first marriage, where the model can be thought of as comprising 30 equations (one for each value of t), 30 separate means of M and 30 separate means of H, representing education, are used as controls when using equation (11) to compute adjusted values of the P_t function and derived life tables for urban and rural. These interval-specific means (one mean value of M and one mean value of H for each value of t) are calculated from the expanded data set (using the re-normalized weights) to which the CLL model for a particular parity transition is fitted.

We refer to the two life tables for urban and rural calculated in this way as *adjusted life tables*—adjusted in the sense that the other socioeconomic predictors are controlled by holding them constant at their interval-specific mean values when residence is varied from urban to rural (i.e., when U is varied from 1 to 0)—and to the values of the PPR and mean and median ages at marriage calculated from these adjusted urban and rural life tables as *adjusted values*.

In both the period analysis and the cohort analysis, interval-specific means rather than the overall means of M and H must be used as control values when calculating adjusted P_t values and derived life tables for urban and rural. It is especially important to do this in the period analysis, because the use of overall means results in younger women being treated as less educated than they really are and older women being treated as more educated than they really are. In models with time-varying effects of education, this kind of distorting effect is amplified if the interval-specific mean of U, M, or H increases or decreases monotonically over time in the life table (as it typically does in the case of U and H), because the mean values of U, M, and H are weighted by a "coefficient" that contains terms in t and t^2, as seen earlier in equation (11). We use interval-specific means throughout, in both the period analysis and the cohort analysis, when calculating adjusted values.

The above procedure for calculating unadjusted and adjusted values is then repeated, with another of the predictors considered as the principal predictor in place of residence. In the unadjusted case, a new model must be run each time another predictor is selected as the sole predictor variable in the model. In the adjusted case, however, a new model need not be run, because all the predictors are already in the model the first time around. One needs only to change the way in which the predictor variables are set to particular values. In the adjusted case, when a predictor other than residence is chosen as the principal predictor variable, the set of control variables again includes all of the other predictors, so that this time residence is included in the set of control variables. One proceeds in this way until each and every predictor variable has been treated as the principal predictor variable.

Unadjusted and adjusted P_t functions and life tables for categories of each predictor variable are calculated in this way for each of the parity transitions from 10th birthday to first marriage, from first marriage to first birth, from first birth to second birth, and so on. For each parity transition, the unadjusted and adjusted life tables yield unadjusted and adjusted values of the PPR and mean and median failure times for each category of each predictor variable. The unadjusted and adjusted PPRs for the various parity transitions are then substituted into equation (1) to yield unadjusted and adjusted values of TFR and (with p_B set to one) TMFR for each category of each predictor variable.

When adjusted PPRs are chained together to calculate an adjusted TFR or TMFR, the set of socioeconomic predictor variables must be the same in the CLL model for each PPR. The set of dummy variables representing life table time intervals can vary, however, as it does between the B−M transition, where there are 29 dummy variables representing 30 intervals (equation (12) or (13) above), and higher-order parity transitions, where there are 9 dummy variables representing 10 intervals.

Further explanation is needed about how to handle time-varying effects of predictors when holding predictors constant at their interval-specific mean values when the predictor variables involve interactions, as in the case of $Z_1 = Ut$ and $Z_2 = Ut^2$. In this case, one does not take the means of Z_1 and Z_2 when holding predictors constant at their interval-specific mean values. Instead, in the expressions for $Z_1 = Ut$ and $Z_2 = Ut^2$, mean values are substituted for U but not for t or t^2, which are left as they are. The same point refers to terms that interact M and H with t and t^2.

This mode of presentation of regression results, in the form of unadjusted and adjusted values of the response variable, is sometimes referred to in the demographic literature as multiple classification analysis (MCA), which was originally developed for ordinary multiple regression (Andrews, Morgan, and Sonquist 1969). The same logic applies to other forms of regression as well, including the CLL regression models used here. In general, the MCA approach focuses on predicted values of the response variable (PPR, mean or median failure time, TFR, or TMFR, in the present instance) classified by categories of each predictor variable with other predictor variables held constant (usually at their mean values, and in the case of the CLL model, at their interval-specific mean values). The MCA mode of presenting results has the advantage of transforming rather complicated and voluminous regression results into simple bivariate tables that are readily understood not only by statisticians and demographers but also by policy makers and the intelligent layman.

Tests of statistical significance

Ideally, model results should be accompanied by tables of estimated coefficients and their standard errors. But because the number of coefficients is large, and also because each coefficient by itself is difficult to interpret, we do not present coefficients. Instead, we have used the jackknife method to compute standard errors of the estimates of PPRs , TFR, TMFR, and mean and median failure times. These standard errors are then used for tests of statistical significance (see Appendix B).

Consistency checks

A number of consistency checks on the methodology, using Philippines data, are presented in Appendix A.

Computer programs for fitting the CLL model

We originally used PROC GENMOD in SAS to fit CLL models (Allison 1995, ch. 7). Subsequently we produced a second version of the programs, written in STATA. The current version of the programs for jackknife estimates of standard errors are written in SAS and have not yet been rewritten in STATA. The core computer programs, in both SAS and STATA, including documentation that is currently under preparation, will be placed in the public domain at a later date.

THE PHILIPPINES DATA

Table 1 shows the distribution of the original Philippines survey samples for 1993, 1998, and 2003 by residence and education. Distributions are shown for women age 10–49 and women age 40–49. Expanded samples of person-year observations for the period analysis and the cohort analysis, shown in Table 2, are created from the two groups of women in Table 1. The sample sizes in Table 2 indicate number of person-year observations in the data sets to which CLL models are fitted. For each of the three surveys, two separate data sets, one for the period analysis and one for the cohort analysis, are created for each of 16 parity transitions (B−M, M−1, 1−2, …, 14−15), for a total of 96 data sets. When fitting CLL models for open parity intervals, some of these person-year data sets are combined, as explained earlier.

MULTIVARIATE ANALYSIS OF EACH SURVEY CONSIDERED SEPARATELY

As already noted, the CLL model can be run without any predictors except the variables representing life table time (either the dummy variable specification or the quadratic specification of life table time, as illustrated by equations (8) and (9) above). In this case one obtains, for each survey, a basic period life table and a basic cohort life table for each parity transition, pertaining to all persons regardless of their socioeconomic characteristics. PPRs, mean and median failure times, TFR, and TMFR are calculated from these basic life tables, as shown in Table 3. PPRs and mean and median failure times are shown only up to the 9–10 transition, but TFR and TMFR are calculated using a higher cutoff (in each case, as high as possible without running into non-convergence problems) and an open parity interval beyond the cutoff.

Table 1: Percent distribution of women by urban/rural residence and education: 1993, 1998, and 2003 DHS surveys, Philippines

Survey		Women age 10-49			Women age 40-49		
year	Education	Urban	Rural	Total	Urban	Rural	Total
1993	Low	17.9	26.8	44.7	21.5	34.0	55.5
	Medium	19.6	15.8	35.3	13.9	9.6	23.6
	High	13.6	6.4	20.0	14.9	6.0	20.9
	Total	51.1	48.9	100.0	50.4	49.7	100.0
				(N =19586)			(N =2741)
1998	Low	13.9	28.7	42.6	15.0	34.3	49.3
	Medium	17.8	17.8	35.6	13.6	12.6	26.2
	High	14.3	7.4	21.8	18.0	6.5	24.5
	Total	46.1	53.9	100.0	46.6	53.4	100.0
				(N = 17857)			(N =2693)
2003	Low	13.8	22.7	36.5	15.4	26.4	41.8
	Medium	22.1	17.9	39.9	18.2	13.4	31.6
	High	16.4	7.1	23.6	18.8	7.9	26.6
	Total	52.3	47.7	100.0	52.4	47.6	100.0
				(N =17515)			(N = 2884)

Note: "Low" education means less than secondary, "medium" means some or completed secondary, and "high" means more than secondary. The samples for which the distributions are shown include single women as well as ever-married women. Numbers in this table incorporate sample weights. The weighted N equals the unweighted N for each of the six samples in the table.

In the case of the cohort estimates, the estimates of PPRs, mean and median failure times, TFR, and TMFR are virtually identical to those calculated from Kaplan-Meier life tables (see Appendix A for details). A similar comparison cannot be made for period estimates, however, because Kaplan-Meier life tables cannot be calculated from period data.

In the period analysis in Table 3, an unexpected finding is that p_B rose and age at first marriage fell over the three surveys. By contrast, in the cohort analysis p_B hardly changed, and age at first marriage rose slightly. The difference occurs because the cohort estimates of mean and median ages at first marriage pertain to marriages that occurred roughly two decades before survey interview when age at first marriage was slowly rising rather than falling.

In the period analysis, a falling mean age at first marriage causes a compression of marriages in calendar time (Bongaarts and Feeney 1998), thereby contributing to a temporary rise in the period estimate of p_B. This “tempo effect” may be part of the reason for the rise in the period estimate of p_B in Table 3. Supporting evidence for this tentative conclusion is that residence and education explain neither the upward trend in p_B nor the downward trend in mean

Table 2: Expanded sample sizes: 1993, 1998, and 2003 DHS surveys, Philippines

Parity transition	Period analysis 1993	Period analysis 1998	Period analysis 2003	Cohort analysis 1993	Cohort analysis 1998	Cohort analysis 2003
B-M	44205	39466	37096	35846	35654	38226
M-1	3788	3846	4210	5627	5443	5945
1-2	7235	6838	7929	8107	8317	9581
2-3	7573	6964	7507	8888	9163	10426
3-4	7041	6526	6200	9304	9225	10333
4-5	5747	5134	4699	8411	7894	8555
5-6	4054	3697	3214	6491	6112	5817
6-7	2871	2678	2227	4636	4475	4155
7-8	2047	1815	1600	3493	3216	2794
8-9	1587	1154	1006	2504	1949	1788
9-10	987	845	601	1613	1356	1033
10-11	619	558	453	1043	809	676
11-12	336	317	242	510	418	341
12-13	211	149	91	304	193	147
13-14	79	71	37	118	97	60
14-15	25	45	26	41	47	40

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 96 data sets. For each data set, weighted and unweighted sample sizes are the same. B-M denotes the transition from a woman's own birth to first marriage, and M-1 denotes the transition from first marriage to first birth. In the period analysis, periods are the five-year period before each survey. In the cohort analysis, cohorts are defined as women age 40-49 at the time of survey.

and median age at first marriage, as will be seen in the multivariate analysis of trends in the next section.

Tempo effects are, however, only part of the story. Another, perhaps more important cause of not only the upward trend in p_B but also the downward trend in age at first marriage is the rising prevalence of non-formalized unions. Mean age at first union (calculated directly from reported first unions occurring in the 5-year period before each survey) was about two years younger for non-formalized unions than for formalized marriages in all three surveys, while the proportion that non-formalized unions are of all unions increased over the three surveys. At ages 15–19 this proportion was 35 percent in the 1993 survey, 36 percent in the 1998 survey, and 55 percent in the 2003 survey; and at ages 20–24 it was 13 percent in the 1993 survey, 16 percent in the 1998 survey, and 25 percent in the 2003 survey. The biggest increases in the proportions at 15–19 and 20–24 and the biggest decline in age at first marriage all occurred between the second and third surveys, a pattern that also suggests a causal effect of prevalence of non-formalized unions on mean and median ages at first marriage and p_B.

Table 3 also shows that both period and cohort estimates of p_M hardly changed over the three surveys, and that p_1 declined only modestly. Both period and cohort estimates of p_2, p_3, …,

Table 3: Period and cohort estimates of parity progression ratios, mean and median ages at first marriage (A_m), mean and median closed birth intervals (CBI), TFR, and TMFR, derived from CLL models in which the only predictor variables are the variables representing life table time intervals: Based on Philippines DHS surveys for 1993, 1998, and 2003

Parity transition	Period analysis			Cohort analysis		
Life table measure	1988-92	1993-97	1998-02	1993	1998	2003
B-M						
PPR (p_B)	0.87	0.89	0.92	0.94	0.93	0.95
Mean A_m	24.0	23.8	23.2	21.7	22.0	22.1
Median A_m	23.0	23.1	22.4	20.9	21.2	21.2
M-1						
PPR (p_M)	0.96	0.96	0.94	0.97	0.97	0.96
Mean CBI	1.3	1.3	1.4	1.4	1.4	1.4
Median CBI	1.0	1.0	1.0	1.0	1.1	1.0
1-2						
PPR (p_1)	0.89	0.85	0.83	0.95	0.92	0.92
Mean CBI	2.9	2.9	3.2	2.4	2.5	2.6
Median CBI	2.3	2.4	2.6	2.0	2.0	2.1
2-3						
PPR (p_2)	0.81	0.78	0.73	0.90	0.87	0.86
Mean CBI	3.1	3.2	3.5	2.6	2.8	2.9
Median CBI	2.6	2.6	2.9	2.3	2.3	2.4
3-4						
PPR (p_3)	0.74	0.67	0.64	0.84	0.79	0.77
Mean CBI	3.2	3.2	3.5	2.8	2.9	3.0
Median CBI	2.6	2.6	2.8	2.4	2.5	2.5
4-5						
PPR (p_4)	0.68	0.66	0.59	0.77	0.75	0.71
Mean CBI	3.0	3.2	3.4	2.9	3.0	3.0
Median CBI	2.6	2.8	3.0	2.4	2.6	2.6
5-6						
PPR (p_5)	0.71	0.65	0.61	0.77	0.76	0.73
Mean CBI	3.1	3.2	3.3	2.9	3.0	3.0
Median CBI	2.7	2.7	2.8	2.5	2.6	2.6
6-7						
PPR (p_6)	0.71	0.66	0.69	0.79	0.75	0.73
Mean CBI	3.0	3.2	3.4	2.8	3.0	3.0
Median CBI	2.6	2.7	2.8	2.4	2.6	2.5

Table 3, continued: Period and cohort estimates of parity progression ratios, mean and median ages at first marriage (A_m), mean and median closed birth intervals (CBI), TFR, and TMFR, derived from CLL models in which the only predictor variables are the variables representing life table time intervals: Based on Philippines DHS surveys for 1993, 1998, and 2003

Parity transition	Period analysis			Cohort analysis		
Life table measure	1988-92	1993-97	1998-02	1993	1998	2003
7-8						
PPR (p_7)	0.71	0.64	0.59	0.79	0.69	0.72
Mean CBI	2.9	3.1	2.9	2.9	2.9	2.9
Median CBI	2.5	2.6	2.6	2.5	2.5	2.6
8-9						
PPR (p_8)	0.67	0.60	0.66	0.74	0.73	0.71
Mean CBI	3.1	2.9	3.4	2.8	2.8	3.0
Median CBI	2.7	2.8	2.8	2.4	2.5	2.6
9-10						
PPR (p_9)	0.59	0.64	0.61	0.69	0.69	0.72
Mean CBI	2.6	3.5	3.9	2.7	3.9	3.8
Median CBI	2.6	2.6	2.5	2.4	2.5	2.5
TFR	3.59	3.23	2.99	5.15	4.47	4.39
TMFR	4.15	3.64	3.23	5.47	4.83	4.61

Notes: In the period analysis, the time periods are the 5-year period before each of the 1993, 1998, and 2003 surveys. Separate CLL models are calculated for the 5-year period before each survey, using data from only that survey. In the cohort analysis, three cohorts are defined as women age 40-49 at the time of each of the three surveys. A separate CLL model is calculated for the cohort from each survey, using data from only that survey. The CLL models use a dummy variable specification of life table time up to the parity transition where non-convergence occurs, after which a quadratic specification is used, up to the cutoff, which is chosen as high as possible. Reading across the row labeled "TFR", the cutoffs are 12+, 12+, 13+, 13+, 13+, 13+. Results for parity transitions higher than 9-10 are not shown, but TFR and TMFR are calculated using PPRs for transitions higher than 9-10, including the PPR for the open-parity interval. In this table and in all subsequent tables, births of order 16 and over are ignored.

p_8 declined more substantially. PPRs at higher-order transitions also declined in most cases, but less regularly. Birth intervals between first marriage and first birth are very short, reflecting the fact that many first births were conceived shortly before marriage. Our recoding of date of first marriage back to age at first birth in cases where the first birth was a premarital birth also contributes to the short intervals between first marriage and first birth, but not by very much because only 6–8 percent of births were recoded in this way. Mean and median closed birth intervals tended to increase over the three surveys, more so in the period case than in the cohort case. In the period case the increases again occurred mainly between the second and third surveys. Also in this table, mean age at first marriage always exceeds median age at first marriage, and mean closed birth intervals always exceed median closed birth intervals; this pattern occurs because distributions of failures tend to be skewed toward higher ages (in the case of first marriage) and higher durations in parity (in the case of next births).

The period TFRs in Table 3 are about half a child lower than published values of TFR_{asfr} in the Philippines DHS reports. Differences between TFR_{ppr} and TFR_{asfr} occur because of the different ways that these two measures are calculated. In theory, though rarely in practice, differences even larger than half a child are possible. This is illustrated by the following hypothetical example: Suppose that all never-married women suddenly decide to postpone marriage until the end of the following year, so that no first marriages occur in that year. Then $p_B = 0$ in that year, and, as is evident from equation (1), $TFR_{ppr} = 0$. TFR_{asfr}, on the other hand, is unaffected by the absence of first marriages, except for a relatively small number of births to women who would otherwise have both married and had a child in that year.

Tables 4–12 show unadjusted and adjusted estimates of PPRs, mean and median ages at marriage, mean and median closed birth intervals, TFR, and TMFR by residence and education for the three surveys. Each survey is analyzed independently. In the case of adjusted estimates, this means that control variables are held constant at interval-specific mean values specific to each survey. Thus In other words, control variables are held constant within surveys but not between surveys.

In each survey in Table 4, pertaining to progression to first marriage, unadjusted and adjusted estimates of p_B tend to be higher for rural than for urban, and higher for those with less education. Unadjusted and adjusted estimates of mean and median ages at marriage tend to be lower for rural than for urban and lower for those with less education. Urban/rural differences in p_B and mean and median ages at first marriage are affected little by the introduction of controls for education. By contrast, the differences by education in p_B and mean and median ages at first marriage are affected more substantially by the introduction of controls for residence—but only in the case of the period estimates. Thus education does not explain the effects of residence, but residence does explain some of the effects of education.

Table 5, pertaining to the M–1 transition, shows little variability in p_M or mean or median age at first marriage by residence and education. This is expected, because virtually all women who marry in the Philippines want to have at least one child and to have the first child soon after marriage.

The adjusted period estimates in Tables 6–12, pertaining to the transitions from first to second birth, second to third birth, …, sixth to seventh birth, and seventh or higher-order birth to next birth, show regular patterns by residence and education and over time. For each period and each cohort, PPRs tend to be higher and birth intervals shorter for rural than for urban, and PPRs tend to decrease and birth intervals to increase as education increases. At the higher parity transitions, birth intervals tend to be quite short for the relatively few women with high education. It seems likely that such women are highly selected for relatively short breastfeeding, which shortens the period of post-partum amenorrhea, and for lack of use of contraception. PPRs tend to be lower and birth intervals longer in the period analysis than in the cohort analysis, indicating a long-term trend toward lower PPRs and longer birth intervals. Adjustment of the estimates tends to reduce residence differentials and education differentials in PPRs and birth intervals.

Table 4: Unadjusted and adjusted estimates of parity progression ratios and mean and median ages at first marriage (A_m) for progression from birth to first marriage (B-M): 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
		UNADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_B)	0.84 **	0.83 **	0.90 **	0.92 **	0.91 **	0.94 *
	Mean A_m	24.8 **	24.7 **	23.8 **	22.3 **	22.5 **	22.6 **
	Median A_m	23.8 **	24.1 **	23.1 **	21.6 **	21.7 **	21.9 **
Rural†	PPR (p_B)	0.91	0.96	0.96	0.96	0.95	0.96
	Mean A_m	22.8	22.6	22.1	21.0	21.3	21.4
	Median A_m	21.8	21.9	21.1	20.2	20.5	20.5
Education							
Low†	PPR (p_B)	0.93	0.91	0.97	0.96	0.95	0.96
	Mean A_m	22.0	21.6	20.4	20.4	20.4	20.4
	Median A_m	20.5	20.4	19.4	19.6	19.6	19.5
Medium	PPR (p_B)	0.88 *	0.89	0.94	0.94	0.93	0.97
	Mean A_m	23.1 **	23.2 **	22.4 **	21.8 **	21.8 **	21.7 **
	Median A_m	22.1 **	22.2 **	21.5 **	21.2 **	21.0 **	21.0 **
High	PPR (p_B)	0.80 **	0.88	0.88 **	0.90 **	0.88 **	0.91 **
	Mean A_m	25.8 **	25.6 **	25.2 **	25.1 **	25.0 **	25.2 **
	Median A_m	25.4 **	25.0 **	24.6 **	24.7 **	24.4 **	24.6 **
		ADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_B)	0.83 **	0.82 **	0.89 **	0.92 *	0.90 *	0.94
	Mean A_m	24.9 **	24.8 **	24.1 **	22.5	22.6	22.9
	Median A_m	24.1 **	24.2 **	23.4 **	21.8	21.9	22.1
Rural†	PPR (p_B)	0.89	0.95	0.95	0.95	0.94	0.95
	Mean A_m	23.6	23.1	23.0	22.1	22.8	22.6
	Median A_m	22.7	22.5	22.2	21.4	22.0	21.8
Education							
Low†	PPR (p_B)	0.92	0.88	0.96	0.95	0.94	0.96
	Mean A_m	22.4	22.0	20.8	20.5	20.3	20.5
	Median A_m	20.8	20.7	19.6	19.6	19.6	19.5
Medium	PPR (p_B)	0.87	0.88	0.94	0.94	0.93	0.97
	Mean A_m	23.2 *	23.3 **	22.5 **	21.8 **	21.8 **	21.7 **
	Median A_m	22.2 **	22.4 **	21.5 **	21.1 **	21.0 **	21.0 **
High	PPR (p_B)	0.81 **	0.88	0.88 **	0.91 **	0.89 **	0.91 **
	Mean A_m	25.8 **	25.6 **	25.1 **	25.0 **	25.0 **	25.1 **
	Median A_m	25.4 **	25.1 **	24.6 **	24.6 **	24.4 **	24.6 **

† Reference Category. * $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Note: In this table and also in Tables 5-14, one or more asterisks after a quantity indicate that the quantity differs significantly from the corresponding quantity in the reference category.

Table 5: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals (CBI) for progression from first marriage to first birth (M-1): 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
		UNADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_M)	0.96	0.96	0.94	0.97	0.97	0.96
	Mean CBI	1.3	1.3	1.5	1.4	1.4	1.4
	Median CBI	1.0	1.0	1.0	1.0	1.0	1.0
Rural†	PPR (p_M)	0.97	0.97	0.94	0.97	0.97	0.97
	Mean CBI	1.3	1.4	1.4	1.4	1.4	1.4
	Median CBI	1.0	1.0	1.1	1.1	1.1	1.0
Education							
Low†	PPR (p_M)	0.97	0.96	0.93	0.97	0.97	0.96
	Mean CBI	1.4	1.4	1.5	1.5	1.5	1.5
	Median CBI	1.1	1.1	1.3	1.1	1.2	1.1
Medium	PPR (p_M)	0.97	0.96	0.95	0.96	0.98	0.97
	Mean CBI	1.3	1.3	1.5	1.4	1.2 **	1.3 *
	Median CBI	1.0	1.0	1.1 *	1.1	0.9 **	1.0 *
High	PPR (p_M)	0.95	0.96	0.92	0.96	0.97	0.96
	Mean CBI	1.2	1.3	1.3 *	1.3 **	1.4	1.3 **
	Median CBI	0.9	0.9 **	0.9 **	0.9 *	1.0 **	0.9 **
		ADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_M)	0.96	0.96	0.93	0.97	0.97	0.96
	Mean CBI	1.3	1.3	1.5 *	1.4	1.4	1.4
	Median CBI	1.0	1.0	1.0	1.1	1.0	1.0
Rural†	PPR (p_M)	0.97	0.97	0.94	0.97	0.97	0.97
	Mean CBI	1.3	1.4	1.3	1.4	1.4	1.4
	Median CBI	1.0	1.0	1.0	1.0	1.1	1.0
Education							
Low†	PPR (p_M)	0.97	0.96	0.93	0.97	0.97	0.96
	Mean CBI	1.4	1.4	1.6	1.5	1.5	1.5
	Median CBI	1.1	1.1	1.3	1.1	1.2	1.1
Medium	PPR (p_M)	0.97	0.96	0.95	0.96	0.98	0.97
	Mean CBI	1.3	1.3	1.5	1.4	1.2 **	1.3
	Median CBI	1.0	1.0	1.1 *	1.1	0.9 **	1.0 *
High	PPR (p_M)	0.95	0.96	0.92	0.96	0.96	0.96
	Mean CBI	1.2	1.3	1.3 *	1.3 **	1.4	1.3 **
	Median CBI	0.9	0.9 *	0.9 **	0.9 *	1.0 *	0.9 **

† Reference Category. * $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Table 6: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals (CBI) for progression from first birth to second birth (1-2): 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
		UNADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_1)	0.88	0.81 **	0.81 **	0.93 **	0.89 **	0.90 **
	Mean CBI	3.0 *	3.1 *	3.2	2.3	2.6 **	2.6 *
	Median CBI	2.4	2.5 *	2.6	2.0	2.1	2.1
Rural†	PPR (p_1)	0.90	0.90	0.87	0.97	0.96	0.94
	Mean CBI	2.7	2.8	3.1	2.4	2.3	2.5
	Median CBI	2.3	2.3	2.5	2.0	2.0	2.0
Education							
Low†	PPR (p_1)	0.92	0.91	0.88	0.96	0.94	0.94
	Mean CBI	2.6	2.5	2.9	2.3	2.4	2.5
	Median CBI	2.2	2.1	2.5	2.0	2.0	2.0
Medium	PPR (p_1)	0.92	0.85 *	0.84	0.96	0.93	0.92
	Mean CBI	2.9 **	3.0 **	3.1	2.4	2.4	2.6
	Median CBI	2.3	2.4 *	2.5	2.0	2.0	2.1 *
High	PPR (p_1)	0.83 **	0.82 **	0.80 **	0.89 **	0.87 **	0.87 **
	Mean CBI	3.1 **	3.2 **	3.4 **	2.5 *	2.7 **	2.8 **
	Median CBI	2.4 *	2.6 **	2.7	2.0	2.1	2.2 **
		ADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_1)	0.88	0.82 **	0.81 *	0.93 **	0.89 **	0.90 *
	Mean CBI	2.9	3.0	3.2	2.3	2.5 *	2.6
	Median CBI	2.3	2.5	2.6	2.0	2.1	2.1
Rural†	PPR (p_1)	0.89	0.90	0.86	0.96	0.95	0.93
	Mean CBI	2.8	2.9	3.1	2.4	2.4	2.5
	Median CBI	2.3	2.3	2.5	2.0	2.0	2.1
Education							
Low†	PPR (p_1)	0.92	0.90	0.87	0.95	0.93	0.93
	Mean CBI	2.6	2.6	3.0	2.3	2.4	2.5
	Median CBI	2.2	2.2	2.5	2.0	2.0	2.0
Medium	PPR (p_1)	0.91	0.85	0.84	0.96	0.93	0.92
	Mean CBI	2.9 **	3.0 **	3.1	2.4	2.4	2.6
	Median CBI	2.3	2.4	2.5	2.0	2.0	2.1 *
High	PPR (p_1)	0.83 **	0.82 **	0.80 *	0.90 **	0.88 *	0.88 **
	Mean CBI	3.0 **	3.2 **	3.4 **	2.5 *	2.6 *	2.8 **
	Median CBI	2.4 *	2.6 **	2.7	2.0	2.1	2.2 *

† Reference Category. * $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Table 7: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals (CBI) for progression from second birth to third birth (2-3): 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
		UNADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_2)	0.78 **	0.73 **	0.68 **	0.87 **	0.81 **	0.83 **
	Mean CBI	3.1	3.4 *	3.6	2.7	2.8	3.0
	Median CBI	2.6	2.6	2.9	2.3	2.4	2.4
Rural†	PPR (p_2)	0.85	0.83	0.79	0.94	0.92	0.89
	Mean CBI	3.1	3.0	3.5	2.6	2.7	2.8
	Median CBI	2.6	2.6	2.8	2.3	2.3	2.4
Education							
Low†	PPR (p_2)	0.89	0.82	0.84	0.95	0.93	0.91
	Mean CBI	2.9	2.9	3.3	2.5	2.6	2.7
	Median CBI	2.5	2.5	2.7	2.3	2.3	2.3
Medium	PPR (p_2)	0.79 **	0.81	0.72 **	0.90 **	0.86 **	0.86 **
	Mean CBI	3.0	3.0	3.3	2.7 *	2.8	2.9
	Median CBI	2.6	2.6	2.7	2.3	2.4	2.4
High	PPR (p_2)	0.72 **	0.72 **	0.64 **	0.80 **	0.74 **	0.78 **
	Mean CBI	3.6 **	3.8 **	4.1 **	2.9 **	3.1 **	3.3 **
	Median CBI	2.8 *	2.9 **	3.5 **	2.4	2.5 *	2.6 **
		ADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_2)	0.78	0.73 **	0.69 *	0.88 **	0.83 **	0.84
	Mean CBI	3.1	3.3	3.6	2.7	2.8	2.9
	Median CBI	2.6	2.6	2.9	2.3	2.3	2.4
Rural†	PPR (p_2)	0.83	0.82	0.76	0.92	0.90	0.88
	Mean CBI	3.1	3.1	3.5	2.7	2.7	2.9
	Median CBI	2.7	2.6	2.9	2.3	2.3	2.4
Education							
Low†	PPR (p_2)	0.88	0.80	0.83	0.94	0.92	0.91
	Mean CBI	2.9	2.9	3.3	2.5	2.6	2.7
	Median CBI	2.5	2.5	2.7	2.3	2.3	2.3
Medium	PPR (p_2)	0.79 **	0.80	0.72 **	0.89 **	0.86 **	0.86 **
	Mean CBI	3.0	3.0	3.3	2.7 *	2.8	2.9
	Median CBI	2.6	2.6	2.7	2.3	2.3	2.4
High	PPR (p_2)	0.73 **	0.73	0.65 **	0.81 **	0.75 **	0.78 **
	Mean CBI	3.6 **	3.8 **	4.1 **	2.9 **	3.1 **	3.3 **
	Median CBI	2.8 **	2.9 **	3.5 **	2.4	2.5 *	2.6 **

† Reference Category. * $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Table 8: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals (CBI) for progression from third birth to fourth birth (3-4): 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
				UNADJUSTED ESTIMATES			
Residence							
Urban	PPR (p_3)	0.71 **	0.61 **	0.59 **	0.80 **	0.72 **	0.71 **
	Mean CBI	3.3	3.3	3.5	2.9	3.0	3.2 *
	Median CBI	2.7	2.6	2.8	2.5	2.5	2.6 *
Rural†	PPR (p_3)	0.78	0.73	0.70	0.87	0.85	0.84
	Mean CBI	3.1	3.1	3.6	2.8	2.8	2.9
	Median CBI	2.6	2.6	2.8	2.4	2.5	2.5
Education							
Low†	PPR (p_3)	0.82	0.76	0.74	0.91	0.88	0.87
	Mean CBI	3.1	3.1	3.5	2.7	2.7	2.8
	Median CBI	2.6	2.6	2.8	2.4	2.4	2.4
Medium	PPR (p_3)	0.72 **	0.67 *	0.64 **	0.79 **	0.76 **	0.75 **
	Mean CBI	3.3	3.0	3.5	3.0 **	3.0 **	3.1 **
	Median CBI	2.6	2.5	2.8	2.6 **	2.6 *	2.6 **
High	PPR (p_3)	0.64 **	0.55 **	0.53 **	0.66 **	0.61 **	0.59 **
	Mean CBI	3.4	3.6	3.7	3.2 **	3.5 **	3.3 **
	Median CBI	2.7	2.8	2.8	2.6 **	2.8 **	2.7 **
				ADJUSTED ESTIMATES			
Residence							
Urban	PPR (p_3)	0.71	0.63 *	0.61	0.82	0.74 **	0.72 **
	Mean CBI	3.3	3.3	3.5	2.8	2.9	3.1
	Median CBI	2.7	2.6	2.8	2.4	2.5	2.6
Rural†	PPR (p_3)	0.76	0.70	0.67	0.84	0.81	0.80
	Mean CBI	3.1	3.2	3.6	2.9	3.0	2.9
	Median CBI	2.6	2.6	2.8	2.5	2.5	2.5
Education							
Low†	PPR (p_3)	0.81	0.74	0.73	0.91	0.87	0.86
	Mean CBI	3.1	3.2	3.5	2.7	2.7	2.9
	Median CBI	2.6	2.6	2.8	2.3	2.4	2.4
Medium	PPR (p_3)	0.72 **	0.67	0.64 *	0.79 **	0.76 **	0.75 **
	Mean CBI	3.3	3.0	3.5	3.0 **	3.0 **	3.1 *
	Median CBI	2.6	2.5	2.8	2.6 **	2.6	2.6 **
High	PPR (p_3)	0.65 **	0.56 **	0.54 **	0.67 **	0.63 **	0.60 **
	Mean CBI	3.4	3.6	3.8	3.3 **	3.5 **	3.3 *
	Median CBI	2.7	2.8	2.9	2.7 **	2.8 **	2.7 *

† Reference Category. * $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Table 9: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals (CBI) for progression from fourth birth to fifth birth (4-5): 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
				UNADJUSTED ESTIMATES			
Residence							
Urban	PPR (p_4)	0.63 **	0.61 *	0.52 **	0.72 **	0.70 **	0.63 **
	Mean CBI	3.2	3.4	3.5	3.0 *	3.2 *	3.1
	Median CBI	2.7	2.9	2.9	2.5	2.6	2.6
Rural†	PPR (p_4)	0.73	0.70	0.66	0.82	0.80	0.79
	Mean CBI	2.9	3.1	3.4	2.8	2.9	2.9
	Median CBI	2.6	2.7	3.0	2.4	2.5	2.5
Education							
Low†	PPR (p_4)	0.75	0.73	0.70	0.85	0.81	0.80
	Mean CBI	3.0	3.0	3.3	2.8	3.0	2.9
	Median CBI	2.6	2.6	2.9	2.4	2.6	2.5
Medium	PPR (p_4)	0.69	0.65	0.57 **	0.69 **	0.74 *	0.67 **
	Mean CBI	3.1	3.5 **	3.5	3.0	3.0	3.1
	Median CBI	2.7	2.9 *	3.0	2.5 *	2.5	2.6
High	PPR (p_4)	0.47 **	0.53 **	0.42 **	0.54 **	0.57 **	0.51 **
	Mean CBI	3.2	3.3	3.4	3.2	3.1	3.3
	Median CBI	2.5	2.8	2.8	2.6	2.5	2.7
				ADJUSTED ESTIMATES			
Residence							
Urban	PPR (p_4)	0.63	0.62	0.52 **	0.73 *	0.70 **	0.63 **
	Mean CBI	3.2	3.4	3.5	2.9	3.2 *	3.1
	Median CBI	2.7	2.9	2.9	2.4	2.6	2.6
Rural†	PPR (p_4)	0.69	0.68	0.63	0.79	0.78	0.76
	Mean CBI	2.9	3.2	3.4	2.8	2.9	3.0
	Median CBI	2.6	2.7	3.0	2.4	2.5	2.5
Education							
Low†	PPR (p_4)	0.74	0.72	0.69	0.84	0.80	0.78
	Mean CBI	3.0	3.0	3.3	2.8	3.1	2.9
	Median CBI	2.6	2.7	2.9	2.4	2.6	2.5
Medium	PPR (p_4)	0.69	0.65	0.56 **	0.70 **	0.74	0.67 **
	Mean CBI	3.1	3.5 *	3.5	3.0	3.0	3.1
	Median CBI	2.7	2.9 *	3.0	2.5	2.5	2.6
High	PPR (p_4)	0.48 **	0.54 **	0.43 **	0.55 **	0.58 **	0.53 **
	Mean CBI	3.1	3.3	3.4	3.2	3.0	3.3
	Median CBI	2.5	2.7	2.8	2.6	2.4	2.7

† Reference Category. * $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Table 10: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals (CBI) for progression from fifth birth to sixth birth (5-6): 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
				UNADJUSTED ESTIMATES			
Residence							
Urban	PPR (p_5)	0.63 **	0.57 **	0.51 **	0.70 **	0.67 **	0.67 **
	Mean CBI	3.0	3.2	3.3	3.0 *	2.9	3.0
	Median CBI	2.5	2.7	2.8	2.6 *	2.6	2.5
Rural†	PPR (p_5)	0.78	0.71	0.67	0.82	0.81	0.78
	Mean CBI	3.1	3.2	3.4	2.8	3.0	3.0
	Median CBI	2.8	2.7	2.9	2.5	2.5	2.6
Education							
Low†	PPR (p_5)	0.77	0.73	0.71	0.83	0.80	0.80
	Mean CBI	3.1	3.1	3.3	2.8	2.8	2.9
	Median CBI	2.6	2.8	2.9	2.5	2.5	2.5
Medium	PPR (p_5)	0.68 *	0.58 **	0.54 **	0.70 **	0.71 **	0.68 **
	Mean CBI	3.2	3.3	3.2	3.0	3.1	3.1
	Median CBI	2.8	2.6	2.8	2.7 *	2.7	2.6
High	PPR (p_5)	0.43 **	0.53 **	0.40 **	0.47 **	0.61 **	0.49 **
	Mean CBI	3.0	3.1	3.5	3.4	3.4	3.0
	Median CBI	2.5	2.8	3.0	2.7	2.8	2.5
				ADJUSTED ESTIMATES			
Residence							
Urban	PPR (p_5)	0.63 **	0.58 *	0.52 *	0.71 **	0.69 **	0.68 *
	Mean CBI	3.0	3.1	3.3	3.0	2.9	2.9
	Median CBI	2.5 *	2.7	2.8	2.6	2.5	2.5
Rural†	PPR (p_5)	0.75	0.68	0.64	0.79	0.80	0.75
	Mean CBI	3.2	3.2	3.3	2.8	3.0	3.0
	Median CBI	2.8	2.7	2.9	2.5	2.6	2.6
Education							
Low†	PPR (p_5)	0.75	0.71	0.69	0.82	0.79	0.80
	Mean CBI	3.1	3.1	3.3	2.8	2.8	2.9
	Median CBI	2.6	2.8	2.8	2.5	2.5	2.5
Medium	PPR (p_5)	0.68	0.59 *	0.54 **	0.71 **	0.71 *	0.68 **
	Mean CBI	3.2	3.3	3.2	3.0	3.1 *	3.1
	Median CBI	2.8	2.6	2.8	2.7 *	2.7	2.6
High	PPR (p_5)	0.44 **	0.55 *	0.42 **	0.48 **	0.63 **	0.50 **
	Mean CBI	3.0	3.2	3.5	3.3	3.4	3.0
	Median CBI	2.5	2.8	3.0	2.7	2.8	2.5

† Reference Category. * $0.01 < p \le 0.05$. ** $p \le 0.01$. All tests of statistical signficance are 2-tailed tests.

Table 11: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals (CBI) for progression from sixth to seventh birth (6-7): 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
				UNADJUSTED ESTIMATES			
Residence							
Urban	PPR (p_6)	0.66	0.61	0.64	0.73 **	0.66 **	0.66 **
	Mean CBI	3.0	3.3	3.0 *	2.8	3.0	3.1
	Median CBI	2.5	2.9	2.6	2.5	2.6	2.7
Rural†	PPR (p_6)	0.75	0.69	0.73	0.83	0.80	0.77
	Mean CBI	3.1	3.2	3.6	2.8	3.0	2.9
	Median CBI	2.6	2.7	2.9	2.4	2.6	2.5
Education							
Low†	PPR (p_6)	0.74	0.71	0.72	0.83	0.81	0.79
	Mean CBI	3.0	3.2	3.3	2.8	3.0	2.8
	Median CBI	2.6	2.8	2.8	2.4	2.6	2.5
Medium	PPR (p_6)	0.69	0.62	0.67	0.68 **	0.66 **	0.64 **
	Mean CBI	3.2	3.3	3.6	2.9	3.1	3.2
	Median CBI	2.6	2.8	2.9	2.7 *	2.6	2.7
High	PPR (p_6)	0.46 **	0.43 **	0.56	0.51 **	0.47 **	0.57 **
	Mean CBI	2.7	2.7	3.3	2.2 **	2.8	3.5
	Median CBI	2.3	2.5	2.6	2.0 *	2.4	2.8
				ADJUSTED ESTIMATES			
Residence							
Urban	PPR (p_6)	0.67	0.62	0.64	0.73 *	0.67 *	0.67 *
	Mean CBI	2.9	3.2	3.0 *	2.7	3.0	3.0
	Median CBI	2.5	2.9	2.6	2.4	2.5	2.6
Rural†	PPR (p_6)	0.73	0.66	0.71	0.80	0.77	0.75
	Mean CBI	3.0	3.1	3.7	2.7	3.0	3.0
	Median CBI	2.6	2.6	2.9	2.4	2.6	2.5
Education							
Low†	PPR (p_6)	0.74	0.70	0.70	0.82	0.80	0.78
	Mean CBI	3.0	3.2	3.2	2.8	3.0	2.8
	Median CBI	2.6	2.8	2.7	2.4	2.6	2.5
Medium	PPR (p_6)	0.70	0.62	0.66	0.69 **	0.66 **	0.64 **
	Mean CBI	3.2	3.3	3.6	2.9	3.0	3.2
	Median CBI	2.6	2.8	2.9	2.7 *	2.6	2.7
High	PPR (p_6)	0.47 **	0.44 **	0.56	0.53 **	0.49 **	0.58 *
	Mean CBI	2.8	2.7	3.4	2.2 *	2.8	3.4
	Median CBI	2.3	2.4	2.6	2.0	2.4	2.7

† Reference Category. * $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Table 12: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals (CBI) for progression from seventh to eighth birth (7+-8+): 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
		UNADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_{7+})	0.65	0.47 **	0.58	0.73	0.63 *	0.65 **
	Mean CBI	3.0	2.9	3.2	2.9	2.7	3.0
	Median CBI	2.5	2.5	2.7	2.4	2.4	2.5
Rural†	PPR (p_{7+})	0.66	0.71	0.61	0.73	0.71	0.72
	Mean CBI	2.9	3.0	3.0	2.8	2.8	2.9
	Median CBI	2.5	2.5	2.5	2.5	2.5	2.5
Education							
Low†	PPR (p_{7+})	0.66	0.65	0.61	0.74	0.71	0.70
	Mean CBI	2.9	2.9	2.9	2.8	2.7	2.8
	Median CBI	2.6	2.5	2.5	2.4	2.4	2.5
Medium	PPR (p_{7+})	0.64	0.54	0.61	0.69	0.66	0.69
	Mean CBI	2.6 *	3.7	3.7 *	2.9	3.1	3.4 **
	Median CBI	2.3 *	2.7	2.9	2.5	2.6	2.8 *
High	PPR (p_{7+})	0.60	0.27 **	0.58	0.67	0.40 **	0.66
	Mean CBI	4.3	2.6	2.3	3.6	2.5	2.6
	Median CBI	3.5	2.3	2.1	2.7	2.0	2.4
		ADJUSTED ESTIMATES					
Residence							
Urban	PPR (p_{7+})	0.65	0.53	0.57	0.73	0.64	0.65 **
	Mean CBI	3.0	2.8	3.1	2.8	2.7	2.9
	Median CBI	2.6	2.5	2.7	2.4	2.4	2.5
Rural†	PPR (p_{7+})	0.65	0.62	0.60	0.73	0.70	0.72
	Mean CBI	2.8	3.0	2.9	2.8	2.8	2.9
	Median CBI	2.5	2.5	2.5	2.5	2.5	2.5
Education							
Low†	PPR (p_{7+})	0.66	0.64	0.61	0.74	0.70	0.70
	Mean CBI	2.9	2.8	2.9	2.8	2.7	2.8
	Median CBI	2.6	2.5	2.6	2.4	2.4	2.5
Medium	PPR (p_{7+})	0.64	0.55	0.61	0.69	0.67	0.69
	Mean CBI	2.6 *	3.7	3.7 *	2.9	3.1	3.4 **
	Median CBI	2.3 *	2.7	2.8	2.5	2.6	2.8 *
High	PPR (p_{7+})	0.59	0.28 **	0.59	0.65	0.40 **	0.67
	Mean CBI	4.2	2.6	2.3 *	3.5	2.5	2.6
	Median CBI	3.4	2.3	2.0	2.6	2.1	2.4

† Reference Category. * $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Table 13 shows unadjusted and adjusted estimates of TFR and TMFR, calculated from unadjusted and adjusted PPRs, where the cutoff for the open parity interval is at as high a parity as possible. TFR and TMFR are always higher for rural than for urban, and always lower for women with more education. In each of the three surveys, as expected, adjustment trends to reduce residence differentials and education differentials in TFR and TMFR.

In Tables 4–12 the effects of residence and education on PPRs and mean and median failure times are not always statistically significant. But in Table 13, the effects of residence and education on TFR and TMFR are always statistically significant at the 1 percent level.

MULTIVARIATE ANALYSIS OF TRENDS OVER THE THREE SURVEYS

The first step in the multivariate analysis of trend in a PPR, mean or median failure time, TFR, or TMFR is to pool data from the three surveys. There are a number of ways in which the data could be pooled. For example, before pooling, one could re-normalize the weights from each survey to add to the average sample size over the three surveys, so that each survey is weighted equally when the three data sets are pooled. Alternatively, one could take population growth into account, so that later surveys are weighted more than earlier surveys if the population is growing, as it has been in the Philippines. In our illustrative application to the three Philippines surveys, sample sizes are roughly equal, and we do not incorporate these refinements. We simply merge person-year data sets, weighted as described earlier, over the three surveys.

In the case of period estimates, the merge involves collapsing the first three columns of Table 2 into one column, resulting in 16 pooled period data sets in place of the original 48 period data sets. The shaded area in the Lexis diagram in Figure 3, which again takes progression to first marriage as an example, indicates the person-year observations to be included in the pooled data sets for the B-M transition. Before merging the three period data sets, we create two new variables in each data set, PERIOD2 and PERIOD3, indicating the second and third five-year periods within the 15-calendar-year time period, with the first (earliest) five-year period as the reference category. Thus (PERIOD2, PERIOD3) = (0,0) for all person-year observations in the earliest five-year period, (1,0) for all person-year observations in the second five-year period, and (0,1) for all person-year observations in the third five-year period. The effects of PERIOD2 and PERIOD 3 are modeled as time-varying, again with a quadratic specification of the time variation (same specification as that used for the time-varying effects of residence and education in equations (11) and (12)).

To calculate unadjusted period estimates of p_B and mean and median ages at first marriage, based on the pooled data set, we estimate a CLL model for progression to first marriage that includes (in addition to the 29 dummy variables indicating the 30 life table time intervals) only PERIOD2 and PERIOD3 as predictor variables, with quadratic specifications of their time-varying effects. Unadjusted estimates of p_B and mean and median ages at first marriage are then estimated for period 1, period 2, and period 3 by setting (PERIOD2, PERIOD3) alternatively to (0, 0), (1, 0), and (0, 1) in the fitted model in order to calculate the discrete hazard function and, from that, the life table for each of the three periods. Values of p_B and mean and median ages at first marriage are then calculated from each of the three life tables.

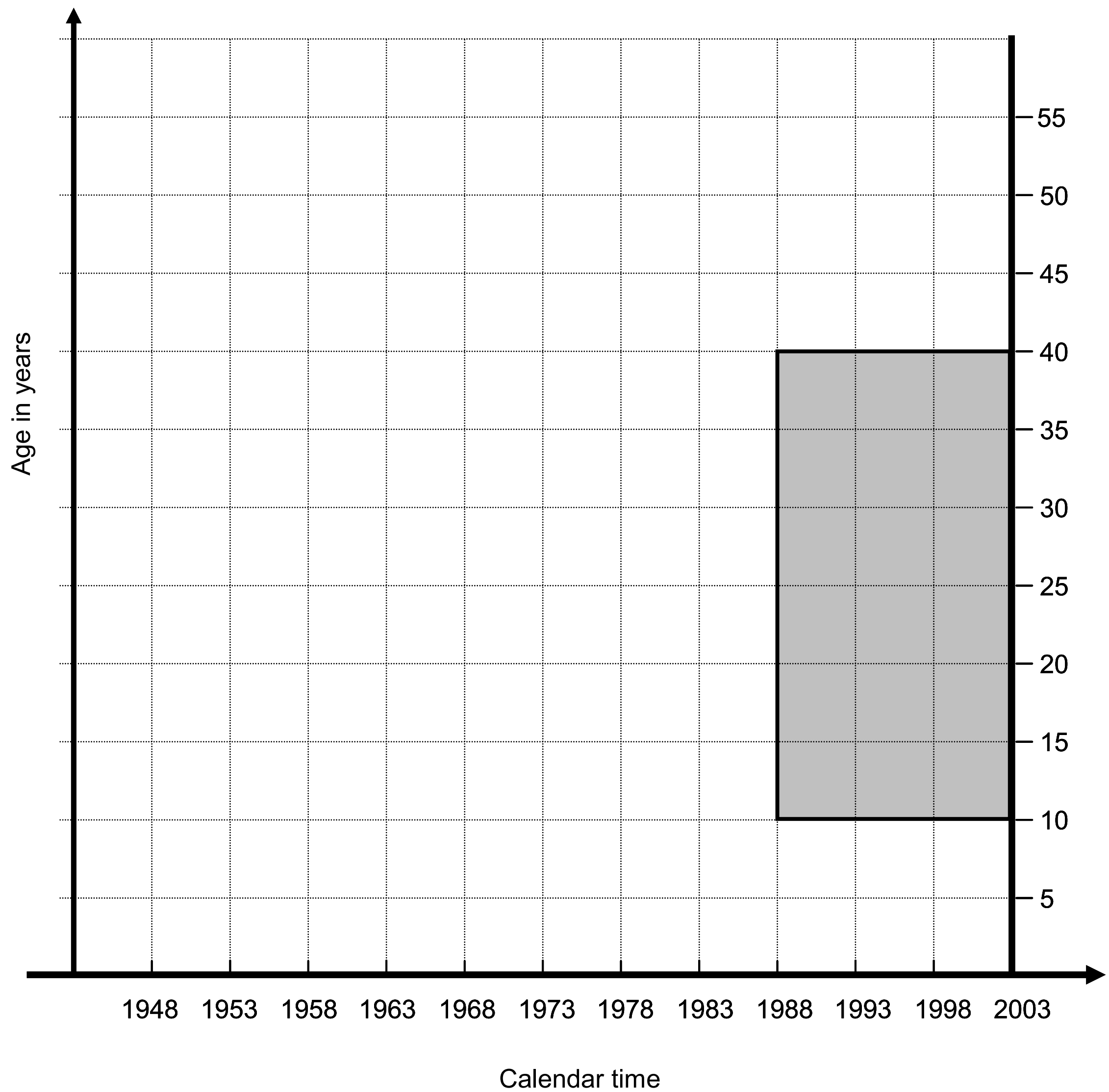
Table 13: Unadjusted and adjusted values of the total fertility rate and the total marital fertility rate, calculated from unadjusted and adjusted parity progression ratios: 1993, 1998, and 2003 DHS surveys, Philippines

		Period analysis			Cohort analysis		
		1988-92	1993-97	1998-02	1993	1998	2003
		TOTAL FERTILITY RATES					
Residence							
Urban	Unadjusted	3.11 **	2.60 **	2.58 **	4.47 **	3.66 **	3.73 **
	Adjusted	3.13 **	2.61 **	2.59 **	4.57 **	3.77 **	3.77 **
Rural†	Unadjusted	4.22	4.05	3.57	5.94	5.46	5.21
	Adjusted	3.86	3.80	3.30	5.41	4.98	4.73
Education							
Low†	Unadjusted	4.64	3.99	3.89	6.22	5.59	5.46
	Adjusted	4.48	3.67	3.69	6.08	5.32	5.21
Medium	Unadjusted	3.60 **	3.18 **	3.07 **	4.54 **	4.29 **	4.25 **
	Adjusted	3.60 **	3.13 **	3.04 **	4.57 **	4.31 **	4.25 **
High	Unadjusted	2.42 **	2.56 **	2.28 **	3.12 **	2.87 **	2.97 **
	Adjusted	2.45 **	2.60 **	2.32 **	3.18 **	2.99 **	3.05 **
		TOTAL MARITAL FERTILITY RATES					
Residence							
Urban	Unadjusted	3.72 **	3.13 **	2.88 **	4.84 **	4.04 **	3.97 **
	Adjusted	3.79 **	3.18 **	2.90 **	4.96 **	4.19 **	4.02 **
Rural†	Unadjusted	4.64	4.24	3.72	6.18	5.74	5.40
	Adjusted	4.34	4.00	3.47	5.71	5.32	4.97
Education							
Low†	Unadjusted	5.01	4.39	4.01	6.50	5.89	5.66
	Adjusted	4.90	4.16	3.84	6.37	5.64	5.41
Medium	Unadjusted	4.10 **	3.57 **	3.25 **	4.83 **	4.59 **	4.37 **
	Adjusted	4.11 **	3.56 **	3.23 **	4.85 **	4.63 **	4.37 **
High	Unadjusted	3.01 **	2.91 **	2.59 **	3.46 **	3.25 **	3.27 **
	Adjusted	3.04 **	2.97 **	2.63 **	3.52 **	3.36 **	3.35 **

† Reference Category. * 0.01 < $p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Note: TFRs and TMFRs in this table are calculated from PPRs in Tables 4-11 as well as PPRs for higher-order parity transitions that are not shown in Tables 4-11 (see note to Table 3). The PPRs in Table 12, pertaining to the open parity interval 7+ to 8+, are not used in calculating the TFRs and TMFRs in Table 13.

Figure 3: Lexis diagram illustrating censoring when setting up the expanded data set for a multivariate period analysis of the trend in progression from 10th birthday to first marriage, based on pooled data from surveys in 1993, 1998, and 2003



Note: See note to Figure 1.

Adjusted estimates of p_B and mean and median ages at first marriage are similarly calculated, the only difference being that the underlying CLL model is expanded to include residence and education in the set of predictor variables, again with quadratic specifications of their time-varying effects. Estimates of p_B and mean and median ages at marriage are then calculated from the model in the same way as in the unadjusted case, but this time with U, M, and H set to their interval-specific mean values in the pooled data set of person-year observations. The logic of the procedure is that, if the predicted values of p_B (or mean or median age at first marriage) are found to be the same for all three periods, we would provisionally conclude that changes in population composition by residence and education and changes in the effects of residence and education explain the trend in p_B (provisionally because there are other predictor variables that affect marriage but are not included in the model).

Unadjusted and adjusted values of PPRs and mean and median failure times for higher-order transitions for the three periods are similarly calculated, and unadjusted and adjusted values of TFR and TMFR for the three periods are calculated from the unadjusted and adjusted PPRs. In this way, one can also analyze the extent to which residence and education explain the trends in PPR and mean and median failure times for higher-order progressions and for TFR and TMFR.

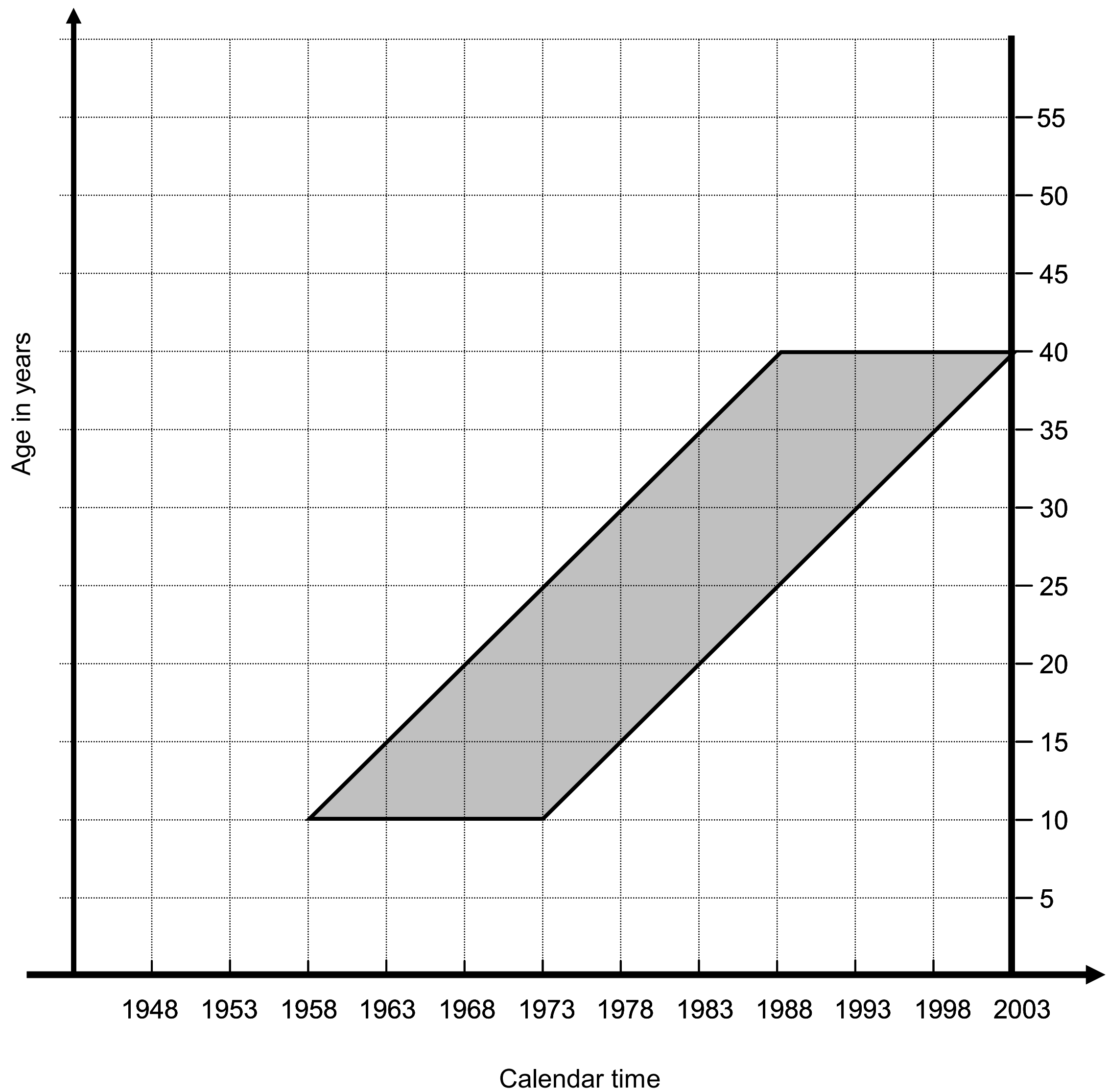
The approach is similar in the cohort analysis. We start by reconstructing the original cohort data sets in Table 2 for women age 40–44 instead of 40–49. This results in cohorts that are five years wide instead of ten years wide. The 10-year-wide cohorts from the three surveys overlap, as mentioned earlier, whereas the 5-year-wide cohorts do not. For purposes of pooling, we want to start with non-overlapping cohorts, and the 5-year-wide cohorts are suitable for this purpose. The shaded area in the Lexis diagram in Figure 4, which again takes progression to first marriage as an example, indicates the person-year observations to be included in the pooled cohort data set. Except for this shift to 5-year-wide cohorts, the pooling of the data sets from each survey proceeds in the same way as in the period case.

In the cohort analysis based on the pooled cohort data set, the three 5-year-wide cohorts are specified by two dummy variables, COHORT2 and COHORT3, representing the second and third five-year cohorts. The analysis then proceeds in the same way as in the period case, except that COHORT2 and COHORT3 are used in place of PERIOD2 and PERIOD3.

Results of the multivariate trend analysis based on pooled data from the three Philippines surveys are shown in Tables 14 and 15. The unexpected upward trend in p_B and downward trend in age at first marriage in the period analysis, observed earlier in Table 3, persist after adjustment for residence and education, as shown in Table 14. The trends in higher-order PPRs are mostly downward, the trends in mean and median closed birth intervals are mostly upward, and the trends in TFR and TMFR are both downward. In most cases, adjusting for residence and education reduces—i.e., partially explains—the trend.

Not all the changes in PPRs and mean and median failure times in Table 14 are statistically significant. Unadjusted and adjusted changes in both period and cohort estimates of TFR and TMFR, however, are always statistically significant, not only between the first and third periods and the first and third cohorts, but also between the first and second periods and the

Figure 4: Lexis diagram illustrating censoring when setting up the expanded data set for a multivariate cohort analysis of the trend in progression from 10th birthday to first marriage, based on pooled data from surveys in 1993, 1998, and 2003



Note: See note to Figure 2.

Table 14: Unadjusted and adjusted trends in TFR and its components (pooled data analysis)

		Period analysis			Cohort analysis		
Transition	Measure	1988-92 †	1993-97	1998-02	1993 †	1998	2003
B-M	p_B						
	Unadjusted	0.86	0.89	0.92 **	0.95	0.92 *	0.95
	Adjusted	0.85	0.88	0.91 **	0.93	0.91	0.94
	Mean A_m						
	Unadjusted	23.9	23.9	23.2 **	21.8	21.8	22.0
	Adjusted	24.5	24.3	23.6 **	22.7	22.5	22.5
	Median A_m						
	Unadjusted	23.1	23.0	22.4 **	21.1	21.1	21.2
	Adjusted	23.8	23.6	22.8 **	22.0	21.7	21.7
M-1	p_M						
	Unadjusted	0.96	0.96	0.94 **	0.97	0.97	0.96
	Adjusted	0.96	0.96	0.94 **	0.97	0.97	0.96
	Mean CBI						
	Unadjusted	1.3	1.3	1.4 *	1.4	1.4	1.4
	Adjusted	1.3	1.3	1.4 **	1.4	1.4	1.4
	Median CBI						
	Unadjusted	1.0	1.0	1.0	1.1	1.0	1.0
	Adjusted	1.0	1.0	1.0	1.1	1.0	1.0
1-2	p_1						
	Unadjusted	0.89	0.85 *	0.83 **	0.95	0.92 *	0.91 **
	Adjusted	0.88	0.85 *	0.83 **	0.94	0.92 *	0.91 **
	Mean CBI						
	Unadjusted	2.9	2.9	3.2 **	2.4	2.5	2.6 **
	Adjusted	2.9	2.9	3.2 **	2.4	2.5	2.6 **
	Median CBI						
	Unadjusted	2.3	2.4	2.5 **	2.0	2.0	2.1 *
	Adjusted	2.3	2.4	2.5 **	2.0	2.0	2.1 *
2-3	p_2						
	Unadjusted	0.81	0.78	0.73 **	0.90	0.87 *	0.87 *
	Adjusted	0.80	0.78	0.72 **	0.89	0.86	0.86 *
	Mean CBI						
	Unadjusted	3.1	3.2	3.6 **	2.7	2.8 *	3.0 **
	Adjusted	3.1	3.2	3.5 **	2.7	2.8	3.0 **
	Median CBI						
	Unadjusted	2.6	2.6	2.9 **	2.3	2.4	2.5 **
	Adjusted	2.6	2.6	2.8 **	2.3	2.4	2.5 **

Table 14, continued: Unadjusted and adjusted trends in TFR and its components (pooled data analysis)

Transition	Measure	Period analysis 1988-92 †	1993-97	1998-02	Cohort analysis 1993 †	1998	2003
3-4	p_3						
	Unadjusted	0.74	0.67 **	0.64 **	0.83	0.76 **	0.77 **
	Adjusted	0.73	0.66 **	0.64 **	0.81	0.75 **	0.76 **
	Mean CBI						
	Unadjusted	3.2	3.2	3.5 **	2.9	3.0	3.1
	Adjusted	3.2	3.2	3.5 **	3.0	3.0	3.1
	Median CBI						
	Unadjusted	2.6	2.6	2.8 **	2.5	2.5	2.6
	Adjusted	2.6	2.6	2.8 *	2.5	2.5	2.6
4-5	p_4						
	Unadjusted	0.68	0.66	0.59 **	0.74	0.74	0.69 *
	Adjusted	0.66	0.64	0.58 **	0.72	0.72	0.68
	Mean CBI						
	Unadjusted	3.0	3.2	3.4 **	2.8	3.1 *	3.0
	Adjusted	3.1	3.2	3.4 **	2.8	3.1 *	3.0
	Median CBI						
	Unadjusted	2.6	2.8	2.9 **	2.4	2.6 *	2.5
	Adjusted	2.6	2.7	2.9 **	2.4	2.6 *	2.5
5-6	p_5						
	Unadjusted	0.71	0.65 *	0.61 **	0.77	0.71	0.73
	Adjusted	0.69	0.64	0.59 **	0.75	0.69 *	0.72
	Mean CBI						
	Unadjusted	3.1	3.2	3.3	2.9	2.9	2.9
	Adjusted	3.1	3.1	3.3	2.9	2.9	2.9
	Median CBI						
	Unadjusted	2.7	2.7	2.8	2.5	2.6	2.5
	Adjusted	2.7	2.7	2.8	2.5	2.6	2.5
6-7	p_6						
	Unadjusted	0.71	0.66	0.69	0.80	0.74	0.78
	Adjusted	0.70	0.65	0.67	0.77	0.73	0.75
	Mean CBI						
	Unadjusted	3.0	3.2	3.4 *	2.9	2.9	3.4 *
	Adjusted	3.0	3.2	3.3 *	2.9	2.9	3.3 *
	Median CBI						
	Unadjusted	2.6	2.7	2.8	2.5	2.5	2.8 *
	Adjusted	2.6	2.7	2.7	2.5	2.5	2.7 *

Table 14, continued: Unadjusted and adjusted trends in TFR and its components (pooled data analysis)

Transition	Measure	Period analysis 1988-92 †	1993-97	1998-02	Cohort analysis 1993 †	1998	2003
7+-8+	p_{7+}						
	Unadjusted	0.65	0.61	0.60 *	0.80	0.76	0.75 *
	Adjusted	0.65	0.61	0.60	0.80	0.75	0.74 *
	Mean CBI						
	Unadjusted	2.9	3.0	3.1	3.1	2.8	3.1
	Adjusted	2.9	2.9	3.0	3.1	2.8	3.1
	Median CBI						
	Unadjusted	2.5	2.5	2.6	2.6	2.5	2.6
	Adjusted	2.5	2.5	2.6	2.6	2.5	2.6
TFR	Unadjusted	3.58	3.23 **	2.99 **	5.31	4.58 *	4.45 **
	Adjusted	3.36	3.11 *	2.90 **	5.00	4.61	4.28 **
TMFR	Unadjusted	4.15	3.64 **	3.23 **	5.62	4.96 *	4.69 **
	Adjusted	3.97	3.55 **	3.18 **	5.36	5.05	4.55 **

† Reference Category. * $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Note: A_m denotes age at first marriage, and CBI denotes closed birth interval.The calculation of TFR and TMFR utilizes not only the PPRs shown in the table (except for p_{7+}, which is not used) but also higher-order PPRs that are not shown (see note to Table 3).

first and second cohorts, except for changes in the adjusted cohort estimates of TFR and TMFR between the first and second cohorts.

The extent to which residence and education explain the trends in the various measures is examined in more detail in Table 15, which is calculated from Table 14 using values more exact than shown in Table 14. By "trend" in Table 15 is meant the change in TFR or one of its components between the first and third surveys. "Explanation" is measured by the percentage by which the introduction of cross-survey controls for residence and education reduces the change. This "percentage explained" is calculated as {[(unadjusted change) − (adjusted change)]/(unadjusted change)} x 100.

None of the "percentages explained" pertaining to the B–M and M–1 transitions are statistically significant at the 5 percent level. In the period analysis, however, the "percentages explained" for the B–M transition are not far off from statistical significance. Observed levels of significance of "percentage explained" are 6 percent for p_B, 8 percent for mean age at first marriage, and 9 percent for median age at first marriage.

In the case of the B–M and M–1 transitions, some of the "percentages explained" are negative. Negative percentages occur when p_B or p_M increases or when mean or median failure time decreases. In these cases, urbanization and rising levels of education partially offset the change by reducing p_B or p_M or by increasing mean or median failure time. Removing these offsetting effects of urbanization and rising levels of education by controlling for residence and

Table 15: Percentages of the unadjusted changes in PPR, mean and median failure times, TFR, and TMFR between the 1993 and 2003 surveys that are explained by residence and education (pooled data analysis)

Transition	Measure	Period analysis (5-year period before each survey)	Cohort analysis (women 40-44 at each survey)
B-M	p_B	-13.7	-130.4
	Mean A_m	-31.1	235.4
	Median A_m	-46.6	333.9
M-1	p_M	3.9	-1.7
	Mean CBI	-5.9	23.9
	Median CBI	-68.9	31.6
1-2	p_1	7.5	2.6
	Mean CBI	13.5 *	10.6
	Median CBI	10.6 *	13.9
2-3	p_2	9.4	5.4
	Mean CBI	13.8 **	18.9 **
	Median CBI	16.1 **	18.8 *
3-4	p_3	9.4	26.2 *
	Mean CBI	6.8	37.3
	Median CBI	4.6	31.0
4-5	p_4	8.8	18.8
	Mean CBI	9.4	16.7
	Median CBI	8.8	17.2
5-6	p_5	9.0	16.4
	Mean CBI	11.6	30.4
	Median CBI	13.9	26.2
6-7	p_6	-12.4	29.9
	Mean CBI	15.8	5.4
	Median CBI	11.0	3.9
7+-8+	p_{7+}	6.5	-3.3
	Mean CBI	25.7	158.9
	Median CBI	6.3	44.2
TFR		23.1 **	16.1
TMFR		12.7 **	12.4

* $0.01 < p \leq 0.05$. ** $p \leq 0.01$. All tests of statistical signficance are 2-tailed tests.

Note: Calculated using more exact values than shown in Table 14. Standard errors were derived by the jackknife method (see Appendix B). One or more asterisks after a percentage indicate that the percentage differs significantly from zero.

education across surveys causes p_B or p_M to rise even more and mean or median failure time to fall even more in the adjusted case than in the unadjusted case.

In the cohort analysis in Table 15, the unadjusted changes in PPRs and mean and median failure times for the B–M and M–1 transitions between the first and third surveys are very small, so that, in the expression {[(unadjusted change) − (adjusted change)]/(unadjusted change)} x 100, the numerator is percentaged on a very small denominator, sometimes resulting in a very large "percentage explained". A small denominator occurs because of offsetting effects (effects of residence and education and effects of other factors operating in opposite directions) that are both large compared with the total change. In such cases, the very large "percentage explained" is not statistically significant.

In the case of transitions 1–2 and higher, very few of the "percentages explained" are statistically significant, the exceptions being those pertaining to mean and median closed birth intervals for the 1–2 and 2–3 transitions.

In the 1–2, 2–3, and higher-order transitions, the percentage of change that is accounted for by residence and education is usually greater for mean and median closed birth intervals than it is for PPRs. A possible explanation of this pattern is that declines in PPRs reflect not only effects of residence and education at the individual level but also across-the-board effects of other factors, such as promotion of smaller families by the family planning program. Quite plausibly, most of these other factors have a larger effect on PPRs than on birth intervals. If so, the additional effects of these other factors on PPRs tend to reduce the percentage of the downward change in a PPR that is due solely to urbanization and rising levels of education.

Table 15 also shows that, overall, residence and education account for 23 percent of the change in the period TFR and 16 percent of the change in the cohort TFR between the first and third surveys. In the case of TMFR the percentages are lower, at 13 and 12 percent. They are lower because the effects of residence and education on PPRs are always to lower fertility, regardless of parity transition. But the overall increase in p_B tends to reduce the decline in TFR so that the unpercentaged contribution from residence and education to change in the TFR is percentaged on a smaller denominator (i.e., a smaller decline in the TFR), thereby tending to increase the percentaged contribution from residence and education. This does not happen in the case of change in TMFR, which is unaffected by what happens to p_B. The table also shows that the "percentages explained" pertaining to changes in the period TFR and TMFR are statistically significant, but the "percentages explained" pertaining to changes in the cohort TFR and TMFR are not significant.

Another feature of Table 15 is that the percentage contribution of residence and education to change in TFR and TMFR is greater than the percentage contribution of residence and education to the change in any of the individual PPRs from which TFR and TMFR are calculated. This occurs because of the way that TFR and TMFR are calculated from PPRs in equation (2), where, within each term on the right-hand side of the equation, a number of PPRs are multiplied together. Because of the cumulative multiplicative nature of the calculation, small percentaged changes in individual PPRs within a term, if all such changes are in the same direction, can result in a much larger percentaged change in the term as a whole and, ultimately, in TFR and TMFR.

DISCUSSION AND CONCLUSION

The multivariate methodology developed in this paper, though somewhat complicated in its internal details, ultimately results in simple bivariate tables that are much more easily understood than the multiplicity of coefficients in the underlying CLL models. The methodology has the added advantage of being applicable to not only the cohort TFR and its components but also the period TFR and its components. The application to the period TFR is of particular interest, because the period TFR is the fertility measure most commonly used by demographers and policy makers.

Another important feature of the methodology is its analytical flexibility. This flexibility, involving separate specifications of calendar time and life table time in the same statistical modeling procedure, offers great potential for solving difficult problems of two-way causation by means of lagged predictors, time-varying predictors, and time-varying effects of predictors, where time lags and time variation in predictors and their effects refer to life table time, not calendar time. For example, had our surveys included work histories as well as birth histories, a variable representing woman's work status could be lagged one year behind the time in the life table when she is at risk of a birth, so that causation runs from work to fertility but not from fertility to work (except to the extent that future plans for both work and fertility are simultaneously decided earlier in the life cycle), and this could be done without changing the values of the calendar-time variables representing time periods or cohorts.

The methodology can also handle integrated event histories, which are much talked about but rarely collected. Examples of event histories that can be integrated include marriage histories, birth histories, education histories, work histories, and migration histories. Regarding event histories, the application of the methodology to period data is especially significant because, if a survey sample is large enough, the period analysis can be restricted to the year before the survey, during which many time-varying characteristics (e.g., urban/rural residence) can realistically be viewed as time-invariant predictor variables in the multivariate models. This approach obviates the need for retrospective information on every time-varying characteristic of interest, as would be necessary when analyzing a real cohort whose experience extends further back in calendar time. The result is that more predictor variables can be included in the analysis when using data from surveys that ask about current status but not historical status on many predictor variables of interest.

Finally, it should be noted that the methodology is applicable not only to parity progression but also to any measure involving time elapsed between a starting event and a terminating event, such as birth and death, birth and first sexual intercourse, childbirth and cessation of breastfeeding, entering and exiting the formal education system, and entering and exiting the labor force.

APPENDIX A

CONSISTENCY CHECKS

This appendix presents a few simple consistency checks that show how the estimates of PPRs and TFR derived by the above methodology compare with estimates of PPRs and TFR derived by more conventional methods. The CLL models used for the checks include only the predictor variables representing life table time. Residence and education are omitted from the models.

Effect of the cutoff parity for the open parity interval on CLL model-based estimates of the TFR

The cutoff parity for the open parity interval can make a difference in the CLL model-based estimate of the TFR, as shown in Table A1. In the case of the cohort TFR, the higher the cutoff, the lower the TFR estimate. Cutoffs higher than 6+ make very little difference in the estimate of the TFR, however. In the case of the period analysis, the choice of cutoff makes no difference in the estimate of the TFR, at least out to tenths of a child.

Clearly it is desirable to use as high a cutoff for the open parity interval as possible, and that is what is done in this paper.

Comparison of CLL model-based estimates with birth history-based estimates of PPRs and TFR

This consistency check compares CLL model-based estimates with birth history-based estimates of PPRs and TFR using cohort data for women age 40–49 in the 1993 survey. Comparisons are first done without weights (i.e., all weights in the original person sample are re-set to one), so that potential errors in our weighting procedures cannot enter into the comparisons. Of course, when weights are not used, the estimates of PPRs and TFR are biased upward, because, as mentioned earlier, rural women are over-represented in the surveys.

The comparisons are then re-done with weights. The birth history-based PPRs are based on persons (i.e., women) as the units of analysis, while the CLL model-based PPRs are based on person-years as the units of analysis, implying different weighting procedures, as described earlier. When making comparisons, births of order 16 or higher are ignored.

PPRs are calculated by single parities out to 13+. The CLL model-based PPRs are calculated using the dummy variable specification of life table time intervals out to the 9−10 transition, and the quadratic specification of life table time for transitions 10−11, 11−12, 12−13, and 13+.

Results are shown in Table A2. In each of the two cases, without weights and with weights, the CLL model-based PPR and the birth history-based PPR agree closely at lower-order transitions but increasingly less closely at higher-order transitions. The direction of the discrepancies is systematic, whereby the CLL model-based PPR increasingly exceeds the birth history-based PPR as parity increases.

Table A1: Effect of the cutoff parity for the open parity interval on CLL model-based estimates of the cohort TFR (based on women age 40-49 at time of survey) and the period TFR (pertaining to the 5-year period immediately preceding the survey): 1993 DHS, Philippines

Open parity interval	Period TFR	Cohort TFR
2+	3.59	5.48
3+	3.61	5.38
4+	3.60	5.31
5+	3.60	5.29
6+	3.60	5.27
7+	3.60	5.24
8+	3.59	5.19
9+	3.59	5.17
10+	3.59	5.16
11+	3.59	5.16
12+	3.59	5.14
13+		5.15

Note: The category 13+ includes 13-14 and 14-15. Births of order 16 and higher are ignored. A dummy-variable specification of life table time was used up to but not including the parity transition where the CLL model no longer converged, after which a quadratic specification of life table time was used. In the case of the period analysis, the CLL model for 13+ did not converge even with a quadratic specification. The estimates in this table and all subsequent tables incorporate weights.

The reason for this pattern of discrepancies is that the CLL model-based PPRs take into account censoring, whereas the birth history-based PPRs do not. Because the birth history-based PPRs pertain to women age 40−49 at the time of the survey, not all of these women have completed their childbearing by the survey date. The impact of this censoring on the birth history-based PPRs increases as starting parity increases, because for a woman of any given age between 40 and 49 at time of survey, the likelihood of another birth occurring after the time of the survey increases as starting parity increases. Table A1 shows that the CLL model-based TFR exceeds the birth history-based CEB by 0.23 child in the unweighted case, and by 0.20 child in the weighted case. This result is consistent with the expected effects of censoring.

Comparison of CLL model-based estimates with Kaplan-Meier life table-based estimates of PPRs and TFR

The PPRs and TFR estimated from CLL models, based on person-year observations, should agree closely with PPRs derived from Kaplan-Meier (product-limit) life tables, based on person observations. In this case we expect close agreement, because both the CLL model and the Kaplan-Meier model take censoring into account.

Table A2: Comparison of CLL model-based estimates with birth history-based estimates of PPRs, TFR, and CEB (number of children ever born): Cohort estimates based on women age 40-49 in the 1993 DHS, Philippines

PPRs, TFR, and CEB	Unweighted CLL	Unweighted BH	Unweighted CLL/BH	Weighted CLL	Weighted BH	Weighted CLL/BH
PPR						
p_B	0.95	0.95	1.00	0.94	0.94	1.00
p_M	0.97	0.97	0.99	0.97	0.97	0.99
p_1	0.95	0.95	0.99	0.95	0.95	0.99
p_2	0.91	0.91	1.00	0.90	0.91	1.00
p_3	0.85	0.84	1.00	0.84	0.83	1.00
p_4	0.78	0.78	1.01	0.77	0.76	1.01
p_5	0.78	0.76	1.02	0.77	0.76	1.02
p_6	0.80	0.76	1.04	0.79	0.76	1.04
p_7	0.80	0.74	1.08	0.79	0.73	1.08
p_8	0.74	0.66	1.13	0.74	0.65	1.12
p_9	0.70	0.60	1.16	0.69	0.60	1.15
p_{10}	0.64	0.57	1.13	0.64	0.57	1.12
p_{11}	0.69	0.56	1.23	0.71	0.58	1.23
p_{12}	0.50	0.40	1.25	0.51	0.41	1.25
p_{13+}	0.62	0.38	1.62	0.57	0.37	1.52
TFR	5.30	5.07	1.05	5.15	4.95	1.04
CEB		5.07			4.95	

Notes: BH denotes birth history. The CLL estimates are based on person-year observations, whereas the BH estimates are based on person (i.e., woman) observations. Births of order 16 or higher are ignored in the calculations. CLL model-based estimates of PPRs for transitions B-M, M-1, ..., 9-10 are calculated from models that use a dummy variable specification of life table time, and CLL model-based estimates of PPRs for transitions 10-11 and higher are calculated from models that use a quadratic specification of life table time.

SAS does not provide an option to calculate Kaplan-Meier life tables using weights, so weights were incorporated in the following way: First, for the expanded data set for a specified parity transition, we multiplied the weights by 1,000 and truncated each of the resulting numbers to the nearest integer. Then, for any given woman, we created a number of duplicate women, the number of such women (including the original woman) being equal to the above integer. Using this data set, we then calculated the Kaplan-Meier life table for the specified parity transition.

Results are shown in Table A3. The CLL model-based estimates of PPRs and TFR agree closely with the Kaplan-Meier-based estimates of PPRs and TFR. The agreement is to four decimal places for each PPR up to p_9. For p_{10} and above the agreement is slightly less close, which is expected because, in the case of the CLL model-based estimates, a quadratic specification of the basic life table time dimension of the model is used instead of a dummy variable specification for parity transitions 10−11 and higher. TFR calculated from the CLL-based PPRs and TFR calculated from the Kaplan-Meier-based PPRs are identical out to hundredths of a child.

Table A3: Comparison of CLL model-based estimates with Kaplan-Meier estimates of PPRs and TFR: Cohort estimates based on women age 40-49 in the 1993 DHS, Philippines

PPR	CLL model-based estimates	Kaplan-Meier estimates
p_B	0.94	0.94
p_M	0.97	0.97
p_1	0.95	0.95
p_2	0.90	0.90
p_3	0.84	0.84
p_4	0.77	0.77
p_5	0.77	0.77
p_6	0.79	0.79
p_7	0.79	0.79
p_8	0.74	0.74
p_9	0.69	0.69
p_{10}	0.64	0.64
p_{11}	0.71	0.71
p_{12}	0.51	0.51
p_{13+}	0.57	0.60
TFR	5.15	5.15

Note: The CLL model-based estimates are based on person-year observations, whereas the Kaplan-Meier estimates are based on person (i.e., woman) observations. Weights are incorporated in both sets of estimates. The estimates are for all women age 40-49, regardless of their socioeconomic characteristics. The open-parity interval 13+ includes 13-14 and 14-15.

Comparison of PPRs and mean and median failure times, derived from CLL models that alternatively use a dummy variable specification and a quadratic specification of life table time

Table A4 shows that CLL models with a dummy variable specification of life table time yield slightly lower PPRs, slightly higher mean failure times (except for the B−M transition), and slightly lower median failure times than do CLL models with a quadratic specification of life table time. The discrepancies are larger for median failure times than for mean failure times.

Table A4: Comparison of PPRs and mean and median failure times, derived from CLL models that alternatively use a dummy variable specification and a quadratic specification of life table time: Cohort estimates based on women age 40-49 in the 1993 DHS survey, Philippines

	Dummy variable specification			Quadratic specification		
Parity transition	PPR	Mean failure time	Median failure time	PPR	Mean failure time	Median failure time
B-M	0.94	21.7	20.9	0.94	21.4	21.3
M-1	0.97	1.4	1.0	0.97	1.4	1.0
1-2	0.95	2.4	2.0	0.95	2.3	2.2
2-3	0.90	2.6	2.3	0.91	2.6	2.5
3-4	0.84	2.8	2.4	0.85	2.8	2.6
4-5	0.77	2.9	2.4	0.78	2.8	2.7
5-6	0.77	2.9	2.5	0.78	2.9	2.7
6-7	0.79	2.8	2.4	0.80	2.7	2.6
7-8	0.79	2.9	2.5	0.79	2.8	2.7
8-9	0.74	2.8	2.4	0.74	2.7	2.6
9-10	0.69	2.7	2.4	0.69	2.6	2.5

Note: Comparisons cannot be made for transitions higher than 9-10, because CLL models with a dummy variable specification of life table time do not converge for parity transitions higher than 9-10.

The discrepancies in Table A4 are small enough to use the quadratic specification in CLL models for higher-order parity transitions in instances where CLL models with the dummy variable specification do not converge.

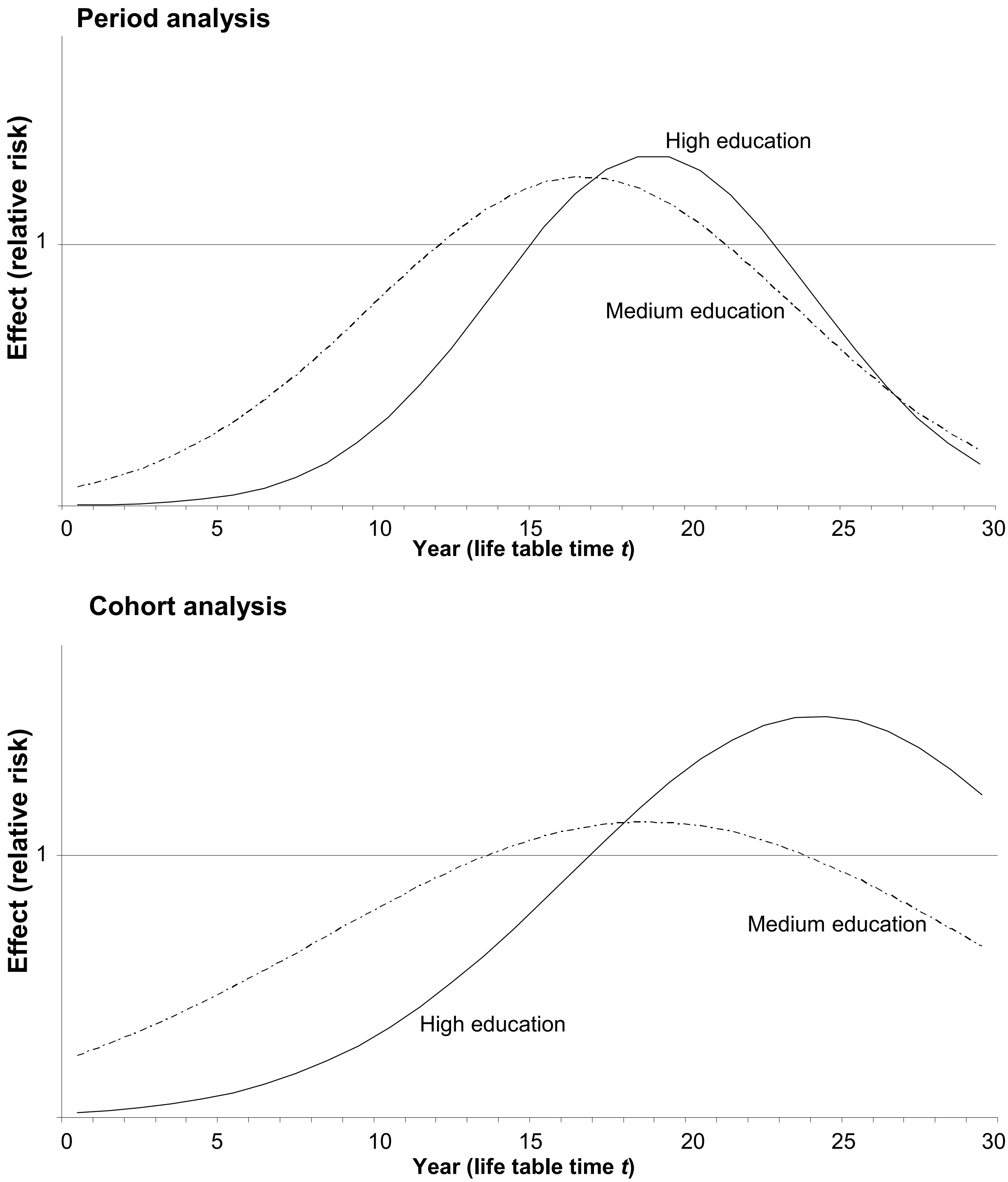
Comparison of proportional and time-varying specifications of effects of socioeconomic predictor variables on the risk of failure

In the text it was argued that the effects of residence and education were not proportional and had to be modeled as time-varying. Time-varying effects were accordingly modeled by interacting each of these predictor variables with a quadratic specification of life table time.

In the case of the B−M transition, the models with time-varying effects are as shown in equations (11) and (12) above. Figure A1 uses data from the 1993 survey to graph $\exp(f+gt+ht^2)$ and $\exp(j+kt+mt^2)$ against t to show how much the quadratic specifications of the time-varying effects of medium and high education on the continuous-time hazard of first marriage, relative to the effect of low education, depart from the time-invariant (i.e., proportional) specifications of these effects. In the proportional case, the graphs would be horizontal lines, so the comparison amounts to seeing the extent to which the graphs depart from horizontal lines.

The first graph in the figure is based on period data pertaining to the five-year period before the 1993 survey, and the second graph is based on cohort data pertaining to women age 40–49 at the time of the 1993 survey. Both graphs indicate postponement of marriage with more education, inasmuch as the relative-risk curves start out below one, rise above one, and then fall,

Figure A1: The effect of medium education, exp($f+gt+ht^2$), and the effect of high education, exp($j+kt+mt^2$), on progression from birth to first marriage, based on the 1993 survey



Note: Effects of medium and high education are relative to low education. See equations (11) and (12) in the text.

usually to values that are again below one, and inasmuch as the curve for high education is shifted to the right, relative to the curve for medium education. Both graphs show that the effects of education vary substantially as *t* increases, indicating major departures from proportionality. Similar graphs of the effect of urban/rural residence, which are not shown, also indicate the need for a time-varying specification of the effect of residence. Also not shown are similar graphs for higher-order parity transitions, which also indicate the need for time-varying specifications of the effects of residence and education on the hazard of a next birth.

APPENDIX B

JACKKNIFE ESTIMATES OF STANDARD ERRORS

Following the approach used in DHS surveys for calculating standard errors of complex measures such as the TFR, we use the jackknife method, which is recommended when the original sample is a multi-stage cluster sample, as is the case in all DHS surveys. DHS surveys apply the jackknife by taking repeated samples from the original sample, each time omitting one primary sampling unit (PSU) from the original sample. The number of repeated samples is the same as the number of PSUs.

PSUs typically are rural villages (or segments of villages in the case of large villages) and urban blocks. In the applications to Philippines data, the number of PSUs is 744 in the 1993 survey, 752 in the 1998 survey, and 819 in the 2003 survey. In the cross-sectional analysis, in which each of the three Philippines surveys is analyzed separately, the number of repeated samples and the number of jackknife iterations are the same as the number of PSUs in the original sample pertaining to the particular survey under consideration. In the trend analysis, based on pooled data, the number of iterations is the sum of the numbers of PSUs over all three surveys.

Jackknife estimates of standard errors are approximations that are more accurate for some measures than for others (Sarndal et al. 1992: 437–442). Our measures are unadjusted and adjusted PPRs, mean and median failure times, TFR, and TMFR by residence and education. The calculation of these measures is complex. The jackknife estimates of standard errors are approximations. The degree of bias in these approximations is unknown.

We did eight runs of our jackknife program—one for period estimates and one for cohort estimates for each of the three Philippines DHS surveys separately and for the pooled sample comprising all three surveys. In any given run of the program, each of N iterations (where N equals the number of PSUs) creates the various expanded samples, re-normalizes weights, and calculates unadjusted and adjusted estimates of PPRs, mean and median failure times, TFR, and TMFR by residence and education. In the case of the two runs based on the pooled sample, estimates of the percentages shown in Table 15 are also calculated for each jackknife iteration. The N iterations yield N estimates of each measure, from which a standard error of the estimate is calculated.

The standard error of any particular measure X derived by the jackknife method is calculated as

$$\mathrm{SE}(X) \;=\; [\mathrm{Var}(X)]^{0.5} \;=\; \{[(N-1)/N][\textstyle\sum(X_i - \overline{X})^2]]\}^{0.5} \qquad \text{(B.1)}$$

where X_i denotes the value of X in the i^{th} iteration (as calculated from the sample with one PSU removed), $\overline{X}$ denotes the mean of the X_i over the N iterations, and the summation ranges from 1 to N.

We also calculate standard errors of pairwise differences in the value of each measure between categories of a predictor variable. For example, in the case of the adjusted TFR by education (low, medium, high), we calculate $X = \mathrm{TFR_M} - \mathrm{TFR_L}$) for each of the N iterations and then use equation (B.1) to compute the standard error of $X = \mathrm{TFR_M} - \mathrm{TFR_L}$. The calculation is repeated for $X = \mathrm{TFR_H} - \mathrm{TFR_L}$. We then form the test statistics $z_M = (\mathrm{TFR_M} - \mathrm{TFR_L})/\mathrm{SE}(\mathrm{TFR_M} - \mathrm{TFR_L})$ and $z_H = (\mathrm{TFR_H} - \mathrm{TFR_L})/\mathrm{SE}(\mathrm{TFR_H} - \mathrm{TFR_L})$. z_M and z_H are assumed to be normally distributed, thereby enabling tests of whether $\mathrm{TFR_M}$ and $\mathrm{TFR_H}$ differ significantly from $\mathrm{TFR_L}$. In these comparisons, low education is considered as the reference category. All tests of significance are two-tailed tests.

ACKNOWLEDGMENTS

We thank Paul Allison, Minja Choe, and Tom Pullum for helpful advice. Support for this research was provided by grant R03 HD045508 from the U. S. National Institute of Child Health and Human Development, and by a grant to 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. An earlier version of this paper was presented at the annual meeting of the Population Association of America, Los Angeles, March 29-April 1, 2006.

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.

Andrews, F., J. Morgan, and J. Sonquist. 1969. *Multiple Classification Analysis.* Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan.

Bongaarts, J., and G. Feeney. 1998. On the quantum and tempo of fertility. *Population and Development Review* 24:271–291.

Cox, D. R. 1972. Regression models and life tables. *Journal of the Royal Statistical Society* B34: 187–220.

Feeney, G. 1986. *Period Parity Progression Measures of Fertility in Japan.* NUPRI Research Papers Series no. 35. Tokyo: Nihon University Population Research Institute.

Pathak, K. P., G. Feeney, and N. Y. Luther. 1998. *Alternative Contraceptive Methods and Fertility Decline in India*. National Family Health Survey Subject Report No. 7. Mumbai: International Institute for Population Sciences. Honolulu: East-West Center.

Philippines National Statistics Office and Macro International. 1994. *National Demographic Survey 1993*. Calverton, Maryland: Philippines National Statistical Office and Macro International.

Philippines National Statistics Office, Philippines Department of Health, and Macro International. 1999. *National Demographic and Health Survey 1998*. Manila: Philippines National Statistical Office and Macro International.

Philippines National Statistics Office and ORC Macro. 2004. *National Demographic and Health Survey*. Calverton, Maryland: Philippines National Statistical Office and ORC Macro.

Prentice, R. L., and L. A. Gloeckler. 1978. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 34: 57–67.

Retherford, R. D., and M. K. Choe. 1993. *Statistical Models for Causal Analysis*. New York: John Wiley.

Sarndol, C. E., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling.* New York: Springer-Verlag.

Schoumaker, B. 2004. A person-period approach to analyzing birth histories. *Population-E* 59: 689–702.