PhilippinesDecomp
Revised March 10, 2006

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

Robert Retherford
East-West Center
Honolulu

Naohiro Ogawa
Nihon University
Tokyo

Rikiya Matsukura
Nihon University
Tokyo

Paper presented at the annual meeting of the Population Association of America,
30 March – 1 April 2006, Los Angeles

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 to the transitions from birth to first marriage, first marriage to first birth, first birth to second birth, and so on. The methodology also encompasses the total marital fertility rate calculated from PPRs, mean and median ages at first marriage, and mean and median closed birth intervals at each parity. The methods are applicable to either period measures or cohort measures. The methods are illustrated by application to data from the 1993, 1998, and 2003 DHS surveys in the Philippines.

This paper develops multivariate methods to estimate effects of socioeconomic predictor variables on the total fertility rate (TFR) and on the trend in the TFR. The methods are applied to individual-level survey data. The analysis of effects 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 developed in this paper to be applicable, however, the TFR must be calculated from parity progression ratios (PPRs), where the concept of a woman's parity is extended from its usual definition as the number of children that she has ever borne to include the state of being single (i.e., never-married) with no children ever born and the state of being ever-married with no children ever born. The parity progression ratios (PPRs) considered here then pertain to 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, from second birth to third birth, from third birth to fourth birth, and from fourth or higher-order birth to next higher-order birth. The PPRs so obtained are aggregated to a TFR and a total marital fertility rate (TMFR). The multivariate methods are applicable to either period or cohort measures of the TFR and TMFR as well as each PPPR individually. Although it is assumed in this paper that all births occur within marriage, the methods can easily be modified to handle premarital births, as is explained in the concluding discussion.

Effects of predictor variables on the trend in the TFR are evaluated in two ways: (1) by recalculating trends with one or more predictor variables controlled and (2) by decomposing the change in TFR between two periods or two cohorts into components due to change in each predictor with the other predictors controlled.

We focus on the TFR calculated from PPRs instead of the TFR calculated from ASFRs (TFR_{asfr}) for two major 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. As far as we know, a multivariate method for analyzing factors affecting TFR calculated from PPRs using individual data has not been developed until now. 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 whether she is married, 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.

Another advantage of the TFR calculated from PPRs is that PPRs are unaffected by temporary distortions in the parity composition of women in each age group. The TFR calculated from ASFRs (TFR_{asfr}), on the other hand, is affected by these temporary distortions (Kohler and Ortega 2002a and 2002b; Kohler, Billari, and Ortega 2002). PPRs and the TFR calculated from them are also influenced to some extent by temporary distortions in the age composition of women at each starting parity, but these distortions usually have a smaller effect on the estimate of the TFR. Largely for this reason, external shocks, such as economic booms or recessions, result in fertility fluctuations that tend to be smaller for TFR calculated from PPRs than for TFR calculated from ASFRs.

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.

By way of illustration, we apply our methods 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. Cohort measures are based on the earlier reproductive experience of women age 40–49 at the time of each survey. In the Philippines surveys, urban areas were over-sampled, so weights must be used to restore representativeness to the sample. Failure to use the weights would substantially bias estimates of population composition, which plays an important role in our analysis. The three surveys are described in more detail in the basic survey reports, which include questionnaires and 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 first marriage (B–M)
p_M PPR for transition from first marriage to first birth (M–1)
p_1 PPR for transition from first birth to second birth (1–2)
p_2 PPR for transition from second birth to third birth (2–3)
p_3 PPR for transition from third to fourth birth (3–4)
p_{4+} PPR for transition from fourth or higher-order birth to next higher-order birth (4+ to 5+)

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 the case of our Philippines surveys, a cutoff at 4+ is appropriate.

PPRs are calculated from life tables. In general, the life table method is appropriate when the input data include time elapsed between a starting event and a terminal event. The generic term for a terminal event is "failure," and we use this term throughout this paper. In the case of p_B, the starting event is the woman's own birth and "failure" is her first marriage if a first marriage occurs. In the case of p_M, p_1, p_2, p_3, and p_{4+}, 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 shall refer to a birth-to-first-marriage life table also as a nuptiality table.

Because the number of first marriages that occur before age 15 or after age 40 is negligible in the Philippines, we start our nuptiality tables at age 15 and end them at age 40. Thus time in the nuptiality table ranges from 0 years (corresponding to age 15) to 25 years (corresponding to age 40). In the analysis of progression to first marriage, we shall consider that time in the nuptiality table ranges from 0 to 25 years. In the case of subsequent parity transitions, the number of births that occur after 10 years of duration in parity is also negligible, so we terminate life tables for these transitions at 10 years. Thus time in these life tables 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 15, the age at the start of the nuptiality table) are measures of the mean and median age at first marriage. In the case of the life tables for subsequent parity transitions, the mean and median failure times (in years) are measures of mean and median closed birth interval. (The medians so calculated are true medians, based on all failures that occur over the course of the life table. Typically in DHS surveys, medians are calculated differently, as the age by which half of the starting cohort experience failure, yielding a somewhat higher estimate of the median age at failure.)

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

$$TFR = p_B p_M + p_B p_M\, p_1 + p_B p_M\, p_1\, p_2 + p_B\, p_M\, p_1\, p_2\, p_3 + p_B\, p_M\, p_1\, p_2\, p_3\, p_{4+} /(1–p_{4+}) \qquad (1)$$

As explained by Feeney (1986), the term $p_{4+} /(1–p_{4+})$ on the right side of this equation is obtained by assuming that p_4 and all higher-order PPRs equal p_{4+} and pulling out a geometric series. (Recall that if r is a positive number less than one, the geometric series $r+r^2+r^3+... = r/(1–r)$.) This turns out to be quite a good approximation, because PPRs at parities beyond 4 usually decline rather gradually as parity increases.

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, one alternative is to replace p_B and p_M with p_0, defined as the fraction of women who eventually proceed 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. In the three DHS surveys that we examine, the first non-formalized union is treated as a first marriage. We therefore retain p_B and p_M in our analysis of these surveys.

MULTIVARIATE ANALYSIS OF THE TFR AND ITS COMPONENTS

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 in this context 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 are calculated from PPRs, TFR and TMFR can also be modeled in a multivariate way.

A number of multivariate survival models are available. We would like a model that handles time-varying predictor variables and time-varying effects of predictor variables, because predictors are always time-varying in our period life tables and even to some extent in our cohort life tables (as will be explained shortly), and because the effects of predictors on parity progression ratios tend to vary systematically over time (i.e., over time in the life table). For example, in a multivariate cohort life table, the effect of higher education, relative to lower education, is usually to lower the risk of first marriage at the younger reproductive ages and raise it at the older reproductive ages, with a relatively small effect on overall progression to first marriage and a relatively large effect on mean and median age at first marriage. If the effect of education varies with time in this way, a proportional hazards model is inadequate, because in a proportional hazards model the multiplicative effect of education on the risk of first marriage is constant over time in the life table.

Effects may be time-varying not only for progression to first marriage but also for progression to higher-order parities. One reason is that birth intervals (except for the interval between first marriage and first birth) tend not to change much as fertility falls (Pathak et al. 1998), implying that, in our models, the effect of education on birth intervals, as measured by mean or median failure time, tends to be relatively small while its effect on PPRs tends to be relatively large. This is impossible to model with a proportional hazards model of parity progression, because in a proportional hazards model, mean failure time and the probability of failure by the end of the life table cannot vary independently. In a proportional hazards model, if mean failure time rises with education, the probability of failure by the end of the life table must fall, and if mean failure time falls with education, the probability of failure by the end of the life table must rise.

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). We also need a survival model which, when fitted to data, yields a baseline hazard function, so that we can estimate not only the effect of a predictor variable on the risk of failure (as measured by the coefficients of the predictor variables), but also the risk of failure itself (i.e., the hazard function) predicted by the model. Only then can we can calculate predicted values of life table parameters such as PPRs and mean and median failure times. This point will become clearer in the following paragraphs.

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. Cox's proportional hazards model is conceptualized as

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

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 (unobservable) 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:

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

where $a_t = \log h_0(t)$. As always in statistical applications, 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 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 positive for all values of x, and ranges over all positive values of x. 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.

A Cox model in which a predictor is time-varying but its effect, as measured by its coefficient, is time-invariant might seem to be proportional but is not. To show this, we consider a model with only one predictor, urban/rural residence, specified as a time-varying dummy variable $U(t)$ (1 if urban, 0 if rural) with a fitted coefficient b that is time-invariant. $U(t)$ is time-varying because some individuals move from rural to urban or *vice versa* as time progresses. In the case of this simple model, equation (2) reduces to $h_i(t) = h_0(t) \exp[b\, U_i(t)]$. For an individual person who moves from rural to urban at time t_0, the effect of residence on $h_i(t)$ is to multiply $h_0(t)$ by exp(0) = 1 when the person is rural (i.e., when $U(t) = 0$ for $t<t_0$) and by $\exp(b)$ when the person is urban (i.e., when $U(t) = 1$ for $t \geq t_0$). Since the effect of residence on the hazard is not the same at all values of t for the individual, the effect is not proportional. Looked at in a more aggregate way, however, the effect of "being urban" relative to "being rural" in this simple model is always to multiply the baseline hazard $h_0(t)$ by $\exp(b)$, which is time-invariant. Because the model seems to be proportional even though it is not, we refer to it as "quasi-proportional." These rather subtle distinctions are relevant to the multivariate period life tables of parity progression that we shall consider later.

The continuous-time Cox model is fitted by the method of partial likelihood. The baseline hazard function $h_0(t)$ does not appear in the partial likelihood equations—hence the word "partial" (Allison 1995, pp. 114–115). 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). Unfortunately, this second procedure does not work when one or more predictor variables or their effects are time-varying, in which case the Cox model does not yield a baseline hazard function. Because we need the baseline hazard function in order to calculate the hazard function for specified values of the predictors (the necessity of this baseline 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 model, which we consider next.

The complementary log-log (CLL) model

Basic form of the model

The general form of the discrete-time CLL model is

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

where i denotes the i^{th} observation, t is a counter variable denoting life table time interval (t = 0, 1, ...), P_{it} is the discrete probability of failure during the t^{th} life table time interval, a_t is an unspecified function of t (unspecified in the sense of not having a particular mathematical form), and predictors and coefficients are as defined in the Cox model in equations (2) and (3). Life table time intervals may be of variable length, but if the intervals are uniformly one time unit in length (as is assumed henceforth in this paper), then t can also be interpreted as time at the start of the interval. Equation (4) can be written more compactly as

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

where **b** is a row vector of coefficients and **x** is a column vector of predictor variables. The model is fitted by the method of maximum likelihood (Prentice and Gloeckler 1978).

P_t is often called the discrete hazard, but it should be noted that P_t is defined quite differently from the hazard $h(t)$ in the continuous-time 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 (not per unit time) that failure will occur in the 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 but not identical. If the interval is more than one time unit in length, P_t and the average value of $h(t)$ over the interval will be very different.

The derivation of equations (4) and (5) illustrates that the CLL model is a proportional hazards model. We consider a simplified derivation pertaining to one-time-unit intervals (years), where t denotes exact time at the start of a life table time interval ($t = 0, 1, 2, \ldots$). The derivation begins with $\log[-\log(1-P_t)]$ and makes the substitutions $P_t = [S(t)-S(t+1)]/S(t)$ and $S(t) = [S_0(t)]^{\exp(\mathbf{bx})}$, where P_t denotes the probability of failure between exact times t and t+1 conditional on survival to age t, $S(t)$ 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 (i.e., when $\exp(\mathbf{bx}) = 1$). After these two substitutions and some algebraic manipulation, one obtains

$$\log[-\log(1-P_t)] = \log[-\log(1-P_{0,t})] + \mathbf{bx} \tag{6}$$

where $P_{0,t}$ denotes the baseline P_t function defined when all predictors equal zero. Equation (6) is the same as equation (5), in which $a_t = \log[-\log(1-P_{0,t})]$. 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). It is possible, however, to “trick” the CLL model in equation (5) to handle non-proportionality in the form of time-varying predictor variables and time-varying effects of predictor variables, as will be explained shortly.

A major advantage of the discrete-time CLL model over the continuous-time Cox model is that the former model yields a baseline P_t function, even when the CLL model in equation (5) is tricked to include time-varying predictors and time-varying effects of predictors. The reason why the CLL model yields this additional information is that the CLL model is estimated using maximum likelihood instead of partial likelihood, so that the terms a_t (actually the terms from which a_t is calculated, as explained below) in equation (5) remain in the likelihood equations and can therefore be estimated.

The CLL model is superior to another discrete-time survival model, namely the discrete-time logit model, because coefficients of predictors in the CLL model, but not in the discrete-time logit model, have the same interpretation as coefficients of predictors in the Cox model, namely that a one-unit increase in a predictor multiplies the underlying continuous hazard $h_i(t)$ by $\exp(b)$, where b is the coefficient of the predictor. This is evident from the substitution $S(t) = [S_0(t)]^{\exp(\mathbf{bx})}$ (which is derived from the continuous-time Cox model) in the derivation of equation (5) above. 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, yield estimates of the coefficient vector **b** that are not quite identical.

Dummy variable specification of life table time interval

The terms a_t in equation (5) require further explanation. The regression model that is actually fitted (which is a single equation, not one equation for each value of t) includes, in place of a_t, a set of variables representing life table time intervals. Life table time intervals are represented in two ways. The first way is in terms of a variable that we shall call YEAR, which is simply life table life table time t (t = 0, 1, 2, ...). But equation (5) does not use this variable. Instead, the variable YEAR is classified into one-year categories, and life table time intervals are specified by a set of dummy variables, each representing a particular life table time interval. In the case of our multivariate nuptiality tables for the Philippines, there are 25 time intervals, ranging from the 0^{th} time interval to the 24^{th} time interval. (Note that, in order to be consistent with our definition of life table time t, we do not refer to the initial time interval as the first time interval or the last time interval as the 25^{th} time interval.) We denote these dummy variables as $T_0, T_1, ..., T_{23}$, with the last interval (for which t = 24) as the reference category. It follows that time interval 0 is specified by $T_0 = 1$ and $T_1 = T_2 = ... = T_{23} = 0$; time interval 1 is specified by $T_0 = 0$, $T_1 = 1$, and $T_2 = ... = T_{23} = 0$; and time interval 24 (the last interval) is specified by $T_0 = T_1 = ... = T_{23} = 0$. We denote the fitted coefficients of these dummy variables as c_0, c_1, ..., c_{23}.

In equation (5), the intercept of the fitted model (which is the predicted value of $\log[-\log(1-P_{24})]$ when all predictors—including the dummy variables $T_0, T_1, ..., T_{23}$— are set to zero) is a_{24}, corresponding to the last time interval. The predicted value of $a_{23} = \log[-\log(1-P_{23})]$, with all the x_j and $T_0, ..., T_{22}$ set to zero and T_{23} set to one, is $a_{24}+c_{23}$. Thus the value of a_{23} in equation (5) must be calculated as $a_{23}=a_{24}+c_{23}$. More generally, $a_t=a_{24}+c_t$. As this example shows, there are actually many more variables and coefficients on the right side of equation (5) than meet the eye. An important point to note in this regard is that 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 is referred to as an *unrestricted* specification.

Expanded data set of person-year observations

The sample of observations to which the CLL model in equation (4) is fitted also requires further explanation. Each individual's survival history is broken down into a set of discrete time segments, which in our analysis are person-years, up to the year of failure (person-years after the year of failure are excluded). These person-years are then treated as distinct observations, and the new sample of person-year observations is referred to as the "expanded sample." Variables in the original woman record are carried over into the person-year records created from that woman record. Additional variables assigned to the person-year records are YEAR (life table time *t*), a variable that we call CALTIME that indicates the calendar year in which the person-year observation is located, and a dummy variable FAILURE indicating whether failure occurred during that person-year of exposure. In our nuptiality example, the value of YEAR for a particular person-year record is calculated as the difference between CALTIME and the calendar year in which the person reached age 15. The values of these variables for each person-year observation are the input data for fitting the model. Note that 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. (See Allison 1995, ch. 7, for details on how to set up the person-year data set.)

A separate expanded data set of person-year observations is created for each parity transition in the period analysis and in the cohort analysis. Table 1 shows the distribution of the original survey samples for 1993, 1998, and 2003 by residence and education. Distributions are shown for women age 15–49 and women age 40–49. Expanded data sets for the period analysis and the cohort analysis, shown in Table 2, are created from these two groups of women. The sample sizes in Table 2 indicate number of person-year observations in the Philippines 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 the six parity transitions, for a total of 36 data sets.

- Tables 1 and 2 about here -

The person-year observations created from a person record are not independent observations. It might seem that these observations should be treated as a cluster, and that the model fitting procedure should include an adjustment for clustering. It has been shown, however, that adjustments for clustering are unnecessary for this type of survival model (see Allison (1982; 1995, ch. 7).

Because the CLL model is applied to a person-year data set, it easily handles censoring—both right-censoring and left-censoring. 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 period, and left-censoring pertains to that part of an individual's exposure that occurs before the calendar period. The CLL model's way of handling censoring is quite simple: If a person is censored in a particular year (either right-censored or left-censored), the corresponding person-year is not included in the expanded data set for the particular parity transition under consideration. In the period analysis, a person-year is censored if the person-year does not fall in the calendar period under consideration.

More generally in the period analysis, a person-year is included in the expanded data set only if all of the following conditions are met: (1) the person-year falls within the specified calendar time period and within the time limits of the life table (e.g., in our nuptiality tables, the value of t for the person-year record cannot be less than 0 or greater than 24); (2) the person corresponding to the person-year record experienced the starting event either in a previous year or in the particular year under consideration; (3) the person did not experience failure in a previous year and may or may not have experienced failure in the particular year under consideration. Person-years not meeting all three of these conditions are coded as censored and not included in the expanded data set for a given parity transition. The censoring criteria are illustrated diagrammatically in Figure 1 in the case of progression from 15th birthday to first marriage.

- Figure 1 about here -

Time-varying predictors

The CLL model also easily handles time-varying predictor variables. One simply assigns, where appropriate, different values of the predictor to different person-year observations created from a particular person record. Although the value of a predictor can vary from one person-year to the next for a person, the CLL model assumes 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 never changes. This assumption is reflected in the form of equation (4), where predictor variables for person-year observations do not have t subscripts. In effect, the expansion of the person sample into a person-year sample converts time-varying predictors into time-invariant predictors.

Our illustrative models for the Philippines include only urban/rural residence and education as predictors. Both variables are defined at the time of survey. Residence is defined as a categorical variable with two categories, namely urban and rural. Education is defined as a categorical variable with three categories, namely less than secondary, some or completed secondary, and more than secondary. Henceforth we refer to these three education categories as low, medium, and high.

We define residence and education as time-invariant predictors, because the surveys do not provide enough information to model them as time-varying. For example, if a woman was age 45 and urban at the time of survey, we assume (incorrectly in many cases) that she was also urban at all earlier ages. Despite this, in the application of the CLL model to the Philippines data, residence and education must be viewed as time-varying predictors. In the period analysis, this is so because the group of women who reach a particular age during the period is not the same group of women who reach some other age during the period (although the two groups may overlap). 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 40th 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. Thus, although residence and education are time-invariant variables for particular women, they are time-varying predictors in our period life tables.

Even in our cohort analysis, the residence and education variables are somewhat time-varying, because the person-year sample design leaves room for the effects of "frailty." "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 nuptiality table, less-educated persons have higher interval-specific risks of failure (i.e., fourth birth) than more-educated persons. This means that in the cohort data set for progression to fourth birth, in the absence of a control for education, the proportion of person-year observations with low education will be higher at the beginning than at the end of a cohort life table, and *vice versa* for high education.

Time-varying effects (unrestricted case)

The CLL model can also incorporate time-varying effects of predictors. Suppose that, in our example of progression to first marriage, the predictor is urban/rural residence, specified by the dummy variable U (1 if urban, 0 if rural). 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 not time-varying. 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. In our example of progression to first marriage, this results in an additional set of predictor variables that we denote here as $W_0 = U\,T_0$, $W_1 = U\,T_1$, ..., $W_{23} = U\,T_{23}$.

The effect of education is specified as time-varying in a similar fashion, but this time two dummy variables, M and H, must be used to represent the three categories of high, medium, and low education. The three categories are represented as $(M, H) = (0, 0)$ for low education, (1, 0) for medium education, and (0, 1) for high education. Both M and H must be interacted with the dummy variables T_0, T_1, ..., T_{23}. The specification of this interaction requires the creation the of the new variables $X_0 = MT_0$, $X_1 = M\,T_1$, ..., $X_{23}= M\,T_{23}$ with coefficients u_0, u_1, ..., u_{23}, and $Y_0 = H\,T_0$, $Y_1 = H\,T_1$, ..., $Y_{23} = H\,T_{23}$ with coefficients v_0, v_1, ..., v_{23}.

With the effects of residence and education specified in this way, the model is

$$\log[-\log(1-P_t)] = a_t + bU + d_0W_0 + ... + d_{23}W_{23} + g_MM + u_0X_0 + ... + u_{23}X_{23}$$
$$+ g_HH + v_0Y_0 + ... + v_{23}Y_{23} \qquad (7)$$

In equation (7), the terms containing U can be written as $bU+d_0W_0+d_1W_1+...+d_{23}W_{23} = bU+d_0UT_0+d_1UT_1+...+d_{23}UT_{23} = U(b+d_0T_0 +d_1T_1+...+d_{23}T_{23})$. The effect of a one-unit change in U (i.e., a change from 0 to 1, representing a change from rural to urban) on $\log[-\log(1-P_t)]$ is then $b+d_0$ for the 0^{th} time interval, because for this interval $T_0 = 1$ and $T_1 = T_2 = ... = T_{23} = 0$, so that the above sum of four terms containing U reduces to $(b+d_0)U$. By similar reasoning, the effect of a one-unit change in U is $b+d_1$ for the 1^{st} time interval, $b+d_2$ for the 2^{nd} time interval, ..., $b+d_{23}$ for the 23^{rd} time interval, and b for the 24^{th} time interval. Thus, as long as $d_0, d_1, ..., d_{23}$ 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 g_M+u_0 for the 0^{th} time interval, g_M+u_1 for the 1^{st} time interval, ..., g_M+u_{23} for the 23^{rd} time interval, and g_M for the 24^{th} time interval; the effect of a change from low to high education is g_H+v_0 for the 0^{th} time interval, g_H+v_1 for the first time interval, ..., g_H+v_{23} for the 23^{rd} time interval and g_H for the 24^{th} 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 $W_0, W_1, ..., W_{23}$—that is time-varying, not the coefficients themselves. The estimated coefficients are pure numbers and therefore 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*.

When we specify time-varying predictors and time-varying effects of predictors, the CLL model is still specified as in equations (4) and (5), which at first blush 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 $W_0, W_1, ..., W_{23}, X_0, X_1, ..., X_{23}, Y_0, Y_1, ..., Y_{23}$,. 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.

Time varying effects: linear, quadratic, or cubic specification

There is a problem with the unrestricted specification of time-varying effects of predictor variables: 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-month observation. The problem arises in the case of the dummy variables $W_0, W_1, ..., W_{23}, X_0, X_1, ..., X_{23}, Y_0, Y_1, ..., Y_{23}$. For example, consider the variable $Y_{23} = H\,T_{23}$ in the multivariate nuptiality analysis. This variable is 1 if the person-year observation has high education and is still single at age 38, and 0 otherwise. In the cross-classification of FAILURE against Y_{23}, it could easily be the case that there are no person-year observations for which both of these variables are 1. In this case one has to combine time intervals until the cross-classification has cases in all four cells. The problem becomes more acute at higher-order parity transitions, where the number of cases is in the data set is smaller to begin with. Because of this problem, we have used an alternative specification of time-varying effects that avoids this problem and requires many fewer predictor variables.

We again use progression to first marriage as an example, with residence as the sole predictor and its effect modeled as time-varying. Instead of interacting U with the 24 dummy variables representing life table time intervals to model time-varying effects, we create the variables $Z_1 = Ut$, $Z_2 = Ut^2$, and $Z_3 = Ut^3$, 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. If we want to use a cubic specification, we include all three variables Z_1, Z_2, and Z_3.

Suppose that we use a cubic specification, and suppose that the fitted coefficients of Z_1, Z_2, and Z_3 are c, d, and e. The right side of the model equation then includes the terms $bU+cZ_1+dZ_2+eZ_3 = bU + cUt+dUt^2+eUt^3 = U(b+ct+dt^2+et^3)$. Thus, at any given value of t, the effect of a one-unit change in U is to increase $\log[-\log(1-P_t)]$ in equation (5) by $b+ct+dt^2+et^3$ units.

The same approach is used to simplify the specification of the time-varying effects of M and H. In the case of the cubic specification, the complete model equation is

$$\log[-\log(1-P_t)] = a_t + U(b+ct+dt^2+et^3) + M(f+gt+ht^2+it^3) + H(j+kt+mt^2+nt^3) \qquad (8)$$

In the previous section that used dummy variable specifications throughout, we needed 75 coefficients to model the time-varying effects of residence and education (25 for residence and 50 for education). In the alternative approach using a cubic specification, we need only twelve coefficients (four for residence and eight for education).

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 when using the full DHS data set with no selection. In other words, the weights sum to the total survey sample size.

In our analysis, when the expanded data sets are created, the weight for a woman carries over to the person-year observations created for that woman. In other words, the same weight is attached both to the original woman record and to each person-year record created from that original woman record.

Each time a CLL model is fitted to an expanded data set (recall from Table 2 that our analysis is based on 36 different expanded data sets), it is important that the weights attached to the person-year records be normalized so that they sum to the number of person-year observations in the particular expanded data set. If this is not done, estimates are biased and likelihood ratio tests of the difference between two nested models are invalid.

For calculating normalized weights, we have the following definitions pertaining to the particular expanded data set to which a CLL model is fitted:

N	The number of person-year observations in the expanded data set
w_i	The original weight attached to the i^{th} person-year record in the expanded data set
W	The sum of the w_i over the person-year observations in the expanded data set
w_i*	Normalized weight to be used when fitting the CLL model to the expanded data set

Normalized weights are calculated as

$$w_i^* = (\sum w_i)(\text{ N/W}) \qquad (9)$$

where the summation is over all person-year observations in the expanded data set.

Output of the CLL model

The output from the fitted CLL model includes estimates of the intercept, the coefficients c_t of the dummy variables representing life table time intervals (t = 0, 1, ..., 23 in the case of progression to first marriage), and the coefficients b_j of the predictors (j = 1, 2, ..., k in equation (4)). Values of a_t are calculated from the intercept (a_{24} in the case of progression to first marriage) and the coefficients c_t (t = 0, ...,23). Standard errors of the estimates of the intercept a_{24} and the coefficients c_t and b_j, as well as various other statistics useful for hypothesis testing, are also part of the output.

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

Once we have fitted values of a_t and the coefficients b_j in equation (4), we can predict values of $\log[-\log(1-P_{it})]$ on the left side of the equation for specified values of the predictors. We can then solve for the value of P_{it}, which is assumed to be the same for all persons with the specified values of the predictors, so we drop the subscript i and just write P_t as in equation (5) . If there are 25 time intervals, as in the case of our nuptiality tables, there are 25 such equations (all derived from a single-equation CLL model), which are solved for 25 values of P_t, for $t = 0, 1, ..., 24$. These 25 values of P_t constitute the discrete-time hazard function for the specified values of the predictors.

A baseline $P_{0,t}$ function is obtained by setting all predictors (i.e., all socioeconomic predictors) equal to zero in the fitted model in equation (4). Equation (4) then reduces to

$$\log[-\log(1-P_{0,t})] = a_t \tag{10}$$

If t ranges from 0 to 24, equation (5) represents 25 separate equations, one for each value of t. Each equation can be solved for P_t, yielding

$$P_{0,t} = 1- \exp[-\exp(a_t)] \tag{11}$$

where the subscript 0 in $P_{0,t}$ denotes the baseline value of P_t with all predictors set to zero. From equations (10) and (11) 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

$$P_t = 1- \exp[-\exp(a_t + \mathbf{bx})] \tag{12}$$

The P_t function (evaluated either with all predictors set to zero or at other specified values of the predictors) determines an entire life table. In presenting calculation formulae for the life table measures of interest, we interpret t as exact time at the start of a life table time interval.

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

$$S(0) = 1$$
$$S(t+1) = S(t)\,(1-P_t),\ \ t=0, 1, ..., 24\ . \tag{13}$$

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

$$f_t = S(t)\,P_t\ . \tag{14}$$

The unconditional probability of failure by time t is calculated as

$$F(t) = 1-S(t)\ . \tag{15}$$

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

$$\mathrm{PPR} = F(25)\ . \tag{16}$$

The mean age at failure is calculated as

$$\text{Mean failure time} = \sum [f_t/F(25)](t + 0.5) \,, \tag{17}$$

where the summation ranges from t=0 to t=24.

The median failure time is calculated as

$$\text{Median failure time} = t, \text{ such that } F(t)/F(25) = 0.5. \tag{18}$$

P_t functions and life tables for progression from first marriage to first birth, first birth to second birth, second birth to third birth, third birth to fourth birth, and fourth or higher-order birth to next higher-order birth, 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 25 one-year time intervals. PPRs and mean and median failure times are calculated from these life tables.

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).

One has to be careful in constructing the data set for progression from fourth or higher-order birth to next higher-order birth (4+ to 5+). In the cohort analysis (based on women age 40–49 at the time of the survey), a woman of, say, parity 7 at the time of the survey gets counted four times in the 4+ category, just like four separate women—the first time from fourth to fifth birth, the second time from fifth to sixth birth, the third time from sixth to seventh birth, and the fourth time from seventh to eighth birth (even though she did not have an eighth birth up to the time of the survey). In the period analysis, the only difference is that if, for example, a parity-7 woman had her seventh birth before the 5-calendar-year period before the survey, then she gets counted only once, namely for the transition between 7 and 8. And even then we include only those person-years that fall within the 5-year calendar period before the survey. Suppose, as another example, that she had her fifth birth just before the 5-year period. We would then count her three times—for the 5–6 transition, the 6–7 transition, and the 7–8 transition, but not the 4–5 transition. The general strategy for constructing the expanded data set for, say, the 4+ to 5+ transition is to form the expanded data set for each transition separately for 4–5, 5–6, ..., n to n+1 (where n+1 is the highest parity attained by anyone in the original person sample), and then merge these data sets.

Once the CLL model is fitted to the data for a particular parity transition, the calculation of predicted values of the P_t function and derived life table for specified values of the predictor variables is easily done in a spreadsheet program such as EXCEL. The spreadsheet format is the same in the period analysis and the cohort analysis.

Note that the CLL model can be run without any predictors except the dummy variables representing life table time intervals. In this case one obtains a basic life table for either period data or cohort data, pertaining to all persons regardless of their characteristics. One can calculate PPRs, mean and median failure times, TFR, and TMFR from these basic life tables. These are shown in Table 3.

- Table 3 about here -

In Table 3, it is noteworthy that the period estimate p_B (PPR for progression to first marriage) increased between the two surveys. The cohort estimates of p_M and p_1 did not change over ten years. In all other cases, the PPR declined consistently over time. In all case, mean age at marriage exceeds median age at marriage, and mean birth intervals exceed median birth intervals, 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 births). Birth intervals between first marriage and first birth are very short, reflecting the fact that in the Philippines many first births are conceived shortly before marriage or non-formalized union.

Calculating unadjusted and adjusted values of the P_t function and derived life table

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 predictor variable, without time-varying effects. 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, urban and rural values of the mean failure time, and urban and rural values of the median failure time. 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, one runs 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 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 25 equations (one for each value of *t*), 25 separate means of *M* and *H*, representing education, are used as controls when computing adjusted values of the P_t function and derived life tables for urban and rural. These interval-specific means are calculated from the expanded data set 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 predictors are controlled by holding them constant at their interval-specific mean values when residence is varied from urban to rural—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 can result in a substantial upward bias for younger women and a substantial downward bias for older women in the period estimates of adjusted values of P_t by urban and rural residence. We use interval-specific means in both the period analysis and the cohort analysis.

The above is 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 only needs to change the way in which the predictor variables are set to particular values.

In the case of adjusted estimates, 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.

The calculation of adjusted P_t functions and derived life tables again is easily carried out in a spreadsheet program such as EXCEL.

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 15th 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 categories of each predictor variable. The unadjusted and adjusted PPRs for the various parity transitions are substituted into equation (1) to yield unadjusted and adjusted values of TFR and (with p_B set to one) TMFR for categories of each predictor variable.

A final point concerns how to handle time-varying effects of predictors when calculating adjusted life tables. We consider the time-varying effect of residence as an example. When the effect of residence on $\log[-\log(1-P_t)]$ is modeled as time-varying, the way in which U is set alternatively to 1 and 0 requires further explanation. The term containing U on the right side of equation (8) is $U(b+ct+dt^2+et^3)$. In the case of urban, this term equals b for the 0^{th} interval, $b+c+d+e$ for the 1^{st} interval, $b+2c+4d+9e$ for the 2^{nd} interval, and so on. In the case of rural, the term $U(b+ct+dt^2+et^3)$ is always zero because U is zero. It should be noted that the value of e can be very small because t^3 gets very large for intervals at the end of the life table. One has to be sure to instruct the computer program that estimates the CLL model (we used GENMOD in SAS) to format the estimates to enough decimal places (we used ten decimal places) to produce at least four significant digits in the estimate of the coefficient e.

Before proceeding to calculate unadjusted and adjusted values of PPRs, mean and median ages at marriage, and mean and median closed birth intervals, it is necessary to choose a a CLL model specification for calculating adjusted values. Table 4 compares four specifications, all of which include both residence and education as predictor variables: (1) quasi-proportional (time-varying predictors without time-varying effects), (2) time-varying predictors with a linear specification of time-varying effects, (3) time-varying predictors with a quadratic specification of time-varying effects, and (4) time-varying predictors with a cubic specification of time-varying effects. For reasons of space, the table is illustrative, pertaining only to progression to first marriage. It confirms the expectation that inclusion of time-varying effects in the model usually has a greater effect on mean and median ages at marriage than on the PPR. Likelihood ratio tests comparing nested models show that the quadratic specification fits the data much better than either the quasi-proportional specification or the linear specification. The cubic specification is significantly better than the quadratic specification in the period analysis but not in the cohort analysis. Similar tables for higher-order transitions (not shown) also indicate that the quadratic specification is significantly better than either the quasi-proportional specification or the linear specification, and that the cubic specification is sometimes, but not always, significantly better than the quadratic specification. On the basis of these tables, we decided always to use the cubic specification when calculating adjusted estimates, in order to get the best possible fit to the data.

- Table 4 about here -

Ideally, model results should be accompanied by tables of estimated coefficients and their standard errors. Because the number of coefficients is large, however, we do not present this information. Instead, we have limited tests of statistical significance to comparison of two CLL models where one model is nested in the other, as in Table 4. The likelihood ratio test is then used to test whether the more-elaborate model is significantly better than the less-elaborate model.

Tables 5–11 show unadjusted and adjusted estimates of PPRs, mean and median ages at marriage and closed birth intervals, and TFR and TMFR for the three surveys. In Table 5, pertaining to progression to first marriage, PPR tends to rise over time for both periods and cohorts in both the unadjusted and adjusted cases. PPR tends to be higher for rural than for urban, and higher for those with less education. Mean and median ages at marriage tend to be lower for rural than for urban, and lower for those with less education.

Controlling for education does not have much effect on urban-rural differences, but controlling for residence usually reduces the differences by education fairly substantially. The controls tend to have a greater effect on mean and median ages at marriage than on the PPR, except in the case of high education, where the control for residence has a substantial effect on both PPR and mean and median ages at marriage.

- Table 5 about here -

The adjusted period estimates in Table 6, pertaining to the transition from marriage to first birth, show high PPRs that declined only slightly over time, mainly between the second and third surveys. As noted earlier, birth intervals are very short, apparently due to the high frequency of births conceived shortly before first marriage or non-formalized union. Birth intervals increased somewhat over time, again mainly between the second and third surveys. By residence, birth intervals increased slightly in both urban and rural areas. By education, they increased for low and medium education but changed little for high education. The cohort pattern was somewhat different. PPRs hardly changed, but mean and median age at marriage fell slightly for medium education.

- Table 6 about here -

The adjusted period estimates in Table 7, pertaining to the transition from first to second birth, show declines in PPRs that are much greater for urban than for rural, and much greater for low and medium education than for high education. Mean and median birth intervals tend to increase fairly substantially over time in all residence and education categories, except for medium education, where the increase was very small. In the case of cohorts, PPRs decline slightly over time, whereas birth intervals tend to increase more substantially, in most cases by about 0.2 year.

- Table 7 about here -

The adjusted period estimates in Table 8, 9, and 10, pertaining to the transitions from second to third birth, third to fourth birth, and fourth 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. By education, PPRs tend to decrease and birth intervals to increase with education. Over time, PPRs tend to fall and birth intervals to increase. Interestingly, in the transition from third to fourth birth, the period estimates of birth intervals tend to dip downward between the first and second surveys and then increase substantially by 0.4–0.5 year between the second a third surveys. The dip in birth intervals for 1993–97 is not reflected by an upward spike in PPRs, however, as one might expect. This unexpected result may have something to do with the fact that the decision on whether to have a fourth birth is a key decision for many couples, given that the TFR for 1993–97 was 3.1 children per woman, as seen earlier in Table 3.

- Tables 8–10 about here -

Table 11 shows unadjusted and adjusted estimates of TFR and TMFR, calculated from the unadjusted and adjusted estimates of PPRs in Tables 5–10. TFR and TMFR are always higher for rural than for urban, and always lower with more education. As expected, differentials by residence and differentials by education tend to be smaller for adjusted than for unadjusted. The table also shows that TFR and TMFR fall consistently over time. The only exception is the unadjusted cohort TMFR for medium education, which rises sharply between 1993 and 1998 and then falls even more sharply between 1998 and 2003.

- Table 11 about here -

The mode of presentation of regression results in Tables 5–11 is known in the literature as multiple classification analysis (MCA) (Andrews, Morgan, and Sonquist 1969; Retherford and Choe 1993). 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. The MCA approach to presenting results focuses on predicted values of the dependent variable (e.g., PPR, mean or median failure time, TFR, or TMFR) classified by categories of each predictor variable, with other predictor variables held constant.

MULTIVARIATE ANALYSIS OF TRENDS IN THE TFR AND ITS COMPONENTS

Multivariate analysis of trends over several surveys

Until now, our multivariate analysis has been cross-sectional, in the sense that our CLL models have been applied separately to one survey at a time. We now consider how this multivariate cross-sectional analysis can be extended into a multivariate longitudinal analysis. By this we mean an analysis of how much each socioeconomic predictor variable contributes to the trend in a nuptiality or fertility measure while holding the other predictor variables constant.

When we analyze how a predictor contributes to a trend, we distinguish between (1) the contribution of changes in population composition by categories of that predictor and (2) the contribution of changes in the cross-sectional effect of that predictor on the nuptiality or fertility measure under consideration. Population composition is measured by the means of the dummy variables representing a predictor, and cross-sectional effects are measured by coefficients of those dummy variables. Regarding population composition, in the case of residence the mean of U (1 if urban, 0 if rural) is simply the proportion urban in the sample of person-years pertaining to the parity transition under consideration. In the case of education, the mean of H is the proportion with high education, the mean of M is the proportion with medium education, and the difference between one and the sum of the high and medium proportions is the proportion with low education. The analysis employs interval-specific means rather than overall means to measure population composition.

The logic of the multivariate longitudinal approach is the following: First we calculate the trend in a particular measure, such as p_B (the PPR for progression to first marriage), based on model-predicted values of p_B with the predictors U, M, and H all set at their interval-specific mean values in the underlying CLL model. This is done separately for each survey. In the case of residence, it is convenient to conceptualize each set of interval-specific means as a 25×1 column array **U**. Similarly, in the case of education, we define the column arrays **M** and **H**.

Then we recalculate the trend, this time substituting the mean of **U** over the three surveys in place of the observed value of **U** in each individual survey. If changes in population composition by urban/rural residence explain part of the trend, the trend will become flatter once population composition by urban/rural residence is controlled in this way.

Then we recalculate the trend a second time, this time with not only a common value of **U** but also common values of **M** and **H** representing education. Then we recalculate the trend a third time, additionally using common values of the coefficients of U and its time-varying components (coefficients b, c, d, and e in equation (8)), where again the common values are obtained by averaging over the three surveys. Then we recalculate the trend a fourth time, additionally using a common set of values of the coefficients of M, H, and their time-varying components (coefficients f, g, h, i, j, k, m, and n in equation (8)).

Equalizing means of predictors across the three surveys allows us to look at the effects on the trend stemming from changes in population composition by residence and education. Equalizing coefficients of predictors across the three surveys allows us to look at the effects on the trend stemming from changes in the cross-sectional effects of these predictors. Somewhat arbitrarily, we have specified the order of introduction of changes as residence composition, education composition, residence effects, education effects. In general, the order of introduction will make a difference in the results.

Finally, the original trend and the four modified trends are plotted on the same graph in order to see visually how each set of additional controls affects the trend. If changes in population composition by residence and education and changes in the cross-sectional effects of residence and education on the P_t function are all that matter, the trend in the particular fertility or nuptiality measure considered will disappear when means and coefficients are equalized over the three surveys. Any residual trend is a result of changes in the baseline P_t function with all predictors set to zero. This residual trend reflects the effects of changes in unobserved variables, plus across-the-board effects of economic and social change that affect all individuals regardless of their values on the predictor variables. We expect these across-the-board effects to be larger for trends in period measures of fertility and nuptiality than for trends in cohort measures of fertility and nuptiality, because population-level shocks such as economic booms and busts are period-specific to a much greater extent than they are cohort-specific. Thus we expect the residual trend to be flatter in the case of cohort measures of fertility and nuptiality than in the case of period measures of fertility and nuptiality. Results for the Philippines are shown in Figure 2.

- Figure 2 about here -

[*It was not possible to complete this part of the empirical analysis before the PAA meeting. Therefore, Figure 2 is omitted from this version of the paper.*]

Decomposition of change in the TFR and its components between two surveys

Method of varying one thing at a time. Decomposing fertility change into components due to change in each socioeconomic predictor variable while holding the other socioeconomic predictor variables constant is another way of analyzing the trend in fertility. The first step in the decomposition analysis is to compute the change that is to be decomposed. By way of example, we consider the change in p_B (the PPR for progression to first marriage) between the 1993 and 2003 DHS surveys in the Philippines. We compute this change using values of p_B computed from CLL models that include both residence and education as time-varying predictors with a cubic specification of time-varying effects, fitted to each survey separately and evaluated at the interval-specific means of the predictors for each survey separately. Let D denote the change in p_B between the two surveys (p_B in the later survey minus p_B in the earlier survey), calculated in this way. We wish to decompose D into components.

The components of D to be calculated include components due to changes in population composition by residence and education and components due to changes in the effects of residence and education. Changes in population composition are again indicated by changes in the interval-specific means of the dummy variables representing residence and education, and changes in effects are indicated by changes in the coefficients of these variables and the variables representing the time-varying components of these variables. In this approach to calculating decompositions, the order in which components are calculated makes no difference in the estimated size of the components, as will become clear shortly.

We consider components due to changes in population composition first. As in the previous section, population composition is measured by the set of interval-specific means of the of the dummy variables representing the predictor variables. The first step in calculating the component of D due to changes in population composition by urban/rural residence is to recalculate the value of p_B from the first survey by substituting the interval-specific means **U** in the second survey in place of the interval-specific means of **U** in the first survey. The means of the education-related predictors are left unchanged at their mean values **M** and **H** in the first survey. The coefficients of the various predictors are also left unchanged at their values in the first survey.

The second step is to calculate a new value of D, denoted by D*, using the recalculated value of p_B from the first survey and the original value of p_B from the second survey. The difference *D* – *D** is interpreted as the component of *D* due to changes in the interval-specific means of the residence-related variables between the two surveys.

The component of *D* due to changes in the interval-specific means of the education-related variables is calculated in a similar manner.

The component of *D* due to changes in the coefficients of the residence-related variables and the component of *D* due to changes in the coefficients of the education-related variables are also calculated in a similar manner. The component of *D* due to change in the baseline P_t function (for which the values of *U*, *M*, and *H* are all zero) is also calculated in a similar manner.

Decompositions of change in the other PPRs, in mean and median failure time for each PPR, in TFR, and in TMFR are also calculated using this procedure of varying one thing at a time. In the case of TFR and TMFR, components due to changes in means of predictors, coefficients of predictors, and baseline P_t functions can be further broken down into subcomponents due to changes in means, coefficients, and baseline P_t functions pertaining to each PPR separately. Varying means and coefficients is simply done for one PPR at a time.

A residual component in the decomposition is obtained by summing the components of *D* as calculated above, and then subtracting this sum from *D*.

Results of applying the method of varying one thing at a time to Philippines data are shown in Table 12. In all cases, PPRs, TFR, and TMFR decomposed into five components due to changes in residence composition, education composition, residence effects, education effects, and baseline hazard function, plus a sixth residual component. (We have not calculated the more detailed decomposition of changes in TFR and TMFR into additional parity-specific subcomponents, although that could be done.) Changes in residence composition explain little of the change in PPRs, TFR, or TMFR. Changes in education composition explain more, especially in the cohort case. The components due to changes in the effects of residence and education are opposite in sign and offset each other to some extent, with the residence components tending to be somewhat larger in absolute magnitude than the education components. In the case of individual PPRs, the largest components are those stemming from change in the baseline hazard function (except for p_B and p_M in the cohort case, where the changes being decomposed are very small). The components due to changes in the baseline hazard function tend to be larger in the period case than in the cohort case, as expected. This is also true of the decomposition of change in TMFR. In the case of TFR, however, the component due to changes in baseline hazard functions is about the same for the period TFR and the cohort TFR. This rather different result probably occurs because p_B rose instead of declined over time, and because the period p_B rose much more than the cohort p_B. In the case of changes in PPRs, residual components in the decompositions tend to be uniformly small in the period case, but somewhat larger and offsetting in the cohort case. In the case of TFR and TMFR, residual components are small in both the period case and the cohort case.

- Table 12 about here -

Method of cumulatively varying one thing at a time. This method is similar to that described in the previous section, except that changes in interval-specific means, coefficients, and baseline P_t function are introduced in a cumulative manner. In this case, the order of introduction, which is somewhat arbitrary, makes a difference. We use the following order: interval-specific means of residence, interval-specific means of education, coefficients of residence-related variables, coefficients of education-related variables, baseline P_t function. Components are calculated as incremental changes in D. In this case there is no residual component in the decomposition. Results for the method of cumulatively varying one thing at a time are shown in Table 13. Qualitative conclusions from this table are broadly similar to those from Table 12.

- Table 13 about here -

CONCLUSION

Our preferred survival model for multivariate analysis of parity progression-based measures of the total fertility rate (TFR) and its components using individual-level data is the complementary log-log (CLL) model with time-varying predictors and a cubic specification of time-varying effects. "Survival" in this case means "not having a first marriage by age 40" or "not having a next birth after 10 years of duration in parity." Conversely, "failure" means "having a first marriage by age 40" or "having a next birth before 10 years of duration in parity." The analysis using Philippines data shows that time-varying-effects models are necessary for an adequate analysis of the effects of socioeconomic variables on the probability of "failure," because these effects are far from proportional. The methodology is applicable to both period and cohort measures of the TFR and its components.

For the methodology to realize its full potential in terms of applicability to survey data, demographic and health surveys and fertility surveys need to start collecting more information on single (i.e., never-married) women. Currently, most of these surveys collect a great deal of demographic, socioeconomic, and health information from ever-married women but very little from single women. Given the increasingly large effect of later marriage and less marriage on fertility in most countries of the world (though not in the Philippines over the time periods we have examined), the need for more information on single women is increasingly apparent.

The potential applicability of the methodology would also be greatly enhanced if fertility surveys devoted more effort to the collection of integrated individual event histories covering marriages, education, work, and migration as well as pregnancies and births.

Our methodology, as applied to Philippines data in this paper, assumes that no births occur before first marriage or non-formalized union. The methodology is easily modified, however, when many births occur before first marriage or non-formalized union. In this case, one uses p_0 (PPR for the transition from parity 0 to parity 1, regardless of marital status) in place of p_B and p_M and includes in the set of predictor variables dummy variables to represent whatever marital status categories are deemed appropriate (e.g., three dummy variables to represent the four categories of single, currently married, divorced or separated, and widowed). One then calculates PPRs, mean and median ages at marriage and closed birth intervals, TFR, and TMFR for each marital status. Of course, this modified approach requires detailed event-history information for single women as well as ever-married women, and marital status categories must be defined broadly enough to include sufficient numbers of cases.

Finally, it should be noted that the multivariate period and cohort life table approach developed here can be applied not only to parity progression but also to any measure involving time elapsed between a starting event and a terminating event (e.g., birth and death, entering and exiting the formal education system, entering and exiting the labor force, giving birth and stopping breastfeeding).

ACKNOWLEDGMENTS

We thank Paul Allison, Minja Choe, and Tom Pullum for helpful advice and Hassan Eini-Zinab for valuable help in developing computer programs for the analysis. Support for this research was provided by National Institute of Child Health and Human Development grant 1 R03 HD045508-02.

PhilippinesDecomp
Revised March 10, 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.

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

Kohler, H.-P., F. C. Billari, and J. A. Ortega. 2002. The emergence of lowest-low fertility in Europe during the 1990s. PDR 28(4): 641-680.

Kohler, H.-P., and J. A. Ortega. 2002a. Tempo-adjusted period parity progression measures, fertility postponement and completed cohort fertility. *Demographic Research* 6: 91–144.

Kohler, H.-P., and J. A. Ortega. 2002b. Tempo-adjusted period parity progression measures: assessing the implications of delayed childbearing for cohort fertility in Sweden, the Netherlands and Spain. *Demographic Research* 6: 145–190.

Pathak, K. P., Griffith Feeney, and Norman 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.

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

Revised March 10, 2006

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

		Women age 15-49			Women age 40-49		
Survey yea	Education	Urban	Rural	Total	Urban	Rural	Total
1993	Low	13.5	19.9	33.3	22.3	31.8	54.1
	Medium	23.7	16.0	39.7	15.2	9.1	24.3
	High	19.3	7.6	26.9	16.2	5.5	21.6
	Total	56.5	43.4	100.0	53.7	46.3	100.0
			(N = 15029)			(N = 2707)	
1998	Low	10.2	17.5	27.8	16.4	28.8	45.2
	Medium	24.7	17.6	42.3	16.7	11.1	27.8
	High	21.7	8.3	29.9	21.2	5.8	27.0
	Total	56.6	43.4	100.0	54.3	45.7	100.0
			(N = 13983)			(N =2651)	
2003	Low	9.4	15.0	24.5	16.3	24.6	40.9
	Medium	26.0	18.9	44.8	19.3	12.8	32.2
	High	22.4	8.3	30.8	19.8	7.2	27.0
	Total	57.8	42.2	100.0	55.4	44.6	100.0
			(N = 13633)			(N = 2915)	

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.

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

Analysis type Survey year	Parity transition B-M	 M-1	 1-2	 2-3	 3-4	 4+ - 5+
Period analysis						
1993	24117	4492	7866	8098	7054	16615
1998	21867	4577	7502	7453	6872	15568
2003	20097	4880	8313	7866	6381	12972
Cohort analysis						
1993	22687	6434	9208	9897	9667	26841
1998	22619	6607	9324	10110	9980	25706
2003	24686	7050	10598	11256	10968	24252

Notes: Expanded sample sizes are numbers of person-year observations. Each cell in the table corresponds to a separate data set, for which the number of person-year observations is shown. 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.

Table 3: Period and cohort estimates of parity progression ratios, mean and median ages at first marriage, and mean and median closed birth intervals, derived from CLL models in which the only predictor variables are the set of dummy 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	0.86	0.88	0.91	0.94	0.92	0.94
Mean age at marriage	24.0	23.9	23.3	21.9	22.1	22.3
Median age at marriage	23.0	23.2	22.5	21.1	21.3	21.4
M-1						
PPR	0.96	0.96	0.93	0.96	0.96	0.96
Mean interval	1.7	1.8	1.8	1.8	1.8	1.8
Median interval	1.5	1.6	1.6	1.6	1.6	1.6
1-2						
PPR	0.87	0.84	0.81	0.94	0.91	0.91
Mean interval	3.2	3.2	3.4	2.8	2.8	3.0
Median interval	2.8	2.8	2.9	2.5	2.5	2.6
2-3						
PPR	0.79	0.76	0.70	0.90	0.85	0.84
Mean interval	3.4	3.4	3.9	3.0	3.1	3.2
Median interval	3.0	3.0	3.3	2.7	2.7	2.8
3-4						
PPR	0.72	0.64	0.62	0.82	0.78	0.75
Mean interval	3.5	3.5	3.8	3.2	3.2	3.3
Median interval	3.0	3.0	3.3	2.8	2.8	2.9
4+ to 5+						
PPR	0.69	0.65	0.63	0.76	0.74	0.72
Mean interval	3.2	3.3	3.4	2.9	3.1	3.1
Median interval	2.9	2.9	3.0	2.7	2.8	2.8
TFR	3.44	3.10	2.84	5.15	4.29	4.25
TMFR	4.02	3.50	3.10	5.50	4.85	4.50

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, a cohort is defined by the cohort's age at the time of survey. Three cohorts are defined as women age 40-49 at the time of each of the three surveys. A separate CLL models is calculated for the cohort from each survey, using data from only that survey. Unadjusted values are based on CLL models with time-varying predictors (either residence or education but not both) without time-varying effects. Adjusted values are based on CLL models with time-varying predictors (both residence and education) with a cubic specification of time-varying effects.

Table 4: Adjusted values of parity progression ratios and mean and median age at first marriage, as calculated from four alternative models: progression from birth to first marriage (B-M): 2003 DHS, Philippines

		Model			
		1	2	3	4
		PERIOD ANALYSIS			
Residence					
Urban	PPR	0.87	0.88	0.88	0.88
	Mean age at marriage	23.9	24.2	24.0	24.3
	Median age at marriage	23.0	23.5	23.3	23.5
Rural	PPR	0.95	0.94	0.95	0.95
	Mean age at marriage	22.7	22.8	22.9	23.1
	Median age at marriage	21.9	22.1	22.0	22.4
Education					
Low	PPR	0.99	0.95	0.96	0.96
	Mean age at marriage	21.6	20.6	20.8	20.9
	Median age at marriage	20.8	19.8	19.5	19.6
Medium	PPR	0.95	0.94	0.94	0.94
	Mean age at marriage	22.8	22.7	22.6	22.6
	Median age at marriage	22.1	21.9	21.7	21.7
High	PPR	0.82	0.88	0.87	0.87
	Mean age at marriage	24.3	25.5	25.2	25.3
	Median age at marriage	23.6	24.8	24.7	24.7
Likelihood ratio test (p-values)			0.00	0.00	0.04
		COHORT ANALYSIS			
Residence					
Urban	PPR	0.93	0.93	0.93	0.93
	Mean age at marriage	22.6	22.9	23.1	23.1
	Median age at marriage	21.7	22.1	22.3	22.4
Rural	PPR	0.94	0.94	0.94	0.94
	Mean age at marriage	22.3	22.5	22.9	22.9
	Median age at marriage	21.4	21.7	21.9	21.9
Education					
Low	PPR	0.98	0.95	0.95	0.95
	Mean age at marriage	21.4	20.6	20.8	20.8
	Median age at marriage	20.6	19.8	19.7	19.7
Medium	PPR	0.97	0.96	0.96	0.96
	Mean age at marriage	21.9	21.9	21.9	21.9
	Median age at marriage	21.1	21.2	21.1	21.1

High	PPR	0.85	0.91	0.90	0.90
	Mean age at marriage	23.9	25.5	25.3	25.4
	Median age at marriage	22.9	24.8	24.7	24.6
Likelihood ratio test (p-values)			0.00	0.00	0.22

Notes:

Life tables of progression from birth to first marriage start at age 15 and end at age 40.

Period estimates pertain to the 5-year period before the survey. Cohort estimates are based on the marriage and birth histories of women age 40-49 at the time of the survey.

Model 1: Predictors are U, M, and H. Model is quasi-proportional, with time-varying predictors without time-varying effects.
Model 2: Predictors are U, M, and H, with time-varying predictors and a linear specification of the time-varying effects of U, M, and H.
Model 3: Predictors are U, M, and H, with time-varying predictors and a quadratic specification of the time-varying effects of U, M, and H.
Model 4: Predictors are U, M, and H, with time-varying predictors and a cubic specification of the time-varying effects of U, M, and H.

Adjusted estimates of PPR and mean and median age at first marriage, classified by urban-rural residence, are based on CLL models that contain both residence and education as predictor variables. Adjusted estimates are calculated from the fitted CLL model by varying the value of U while holding M and H (representing medium and high education) constant at their interval-specific mean values in the sample to which the model is fitted. Likewise, adjusted estimates classified by low, medium, and high education are calculated from the same CLL model by varying the values of M and H while holding U (representing urban residence) constant at its interval-specific mean values in the sample to which the model is fitted.

The likelihood ratio test compares two models (where one model is nested in the other) in order to see whether the more elaborate model is significantly better than the less elaborate model. A p-value is the probability that the more elaborate model is no better than the less elaborate model. A p-value in a particular column compares the model in that column with the model in the previous column.

Table 5: Unadjusted and adjusted estimates of parity progression ratios and mean and median ages at marriage 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	0.80	0.82	0.87	0.92	0.90	0.93
	Mean age at marriage	24.5	24.6	23.9	22.3	22.5	22.7
	Median age at marriage	23.5	24.1	21.6	21.5	21.6	21.6
Rural	PPR	0.93	0.96	0.96	0.96	0.95	0.96
	Mean age at marriage	23.2	22.8	22.4	21.4	21.6	21.8
	Median age at marriage	22.2	22.1	20.8	20.6	20.8	20.8
Education							
Low	PPR	0.95	0.96	0.99	0.97	0.97	0.98
	Mean age at marriage	22.6	22.7	21.3	21.1	21.2	21.4
	Median age at marriage	21.7	22.0	20.6	20.4	20.5	20.6
Medium	PPR	0.90	0.91	0.95	0.94	0.94	0.97
	Mean age at marriage	23.5	23.6	22.8	22.0	22.0	21.9
	Median age at marriage	22.6	22.9	22.0	21.3	21.3	22.0
High	PPR	0.73	0.81	0.82	0.83	0.82	0.84
	Mean age at marriage	24.8	24.5	24.3	23.4	23.4	23.9
	Median age at marriage	23.9	23.9	23.7	22.6	22.5	23.7
				ADJUSTED ESTIMATES			
Residence							
Urban	PPR	0.81	0.82	0.88	0.92	0.90	0.93
	Mean age at marriage	24.9	25.0	24.3	22.6	22.8	23.1
	Median age at marriage	24.1	24.4	23.5	21.9	22.0	22.4
Rural	PPR	0.89	0.95	0.95	0.95	0.93	0.94
	Mean age at marriage	23.7	23.3	23.1	22.3	22.9	22.9
	Median age at marriage	22.7	22.6	22.4	21.5	22.0	21.9
Education							
Low	PPR	0.90	0.88	0.96	0.95	0.94	0.95
	Mean age at marriage	22.3	22.0	20.9	20.7	20.6	20.8
	Median age at marriage	20.8	20.7	19.6	19.7	19.7	19.7
Medium	PPR	0.87	0.88	0.94	0.94	0.93	0.96
	Mean age at marriage	23.3	23.5	22.6	22.0	21.9	21.9
	Median age at marriage	22.1	22.5	21.7	21.3	21.1	21.1
High	PPR	0.80	0.88	0.87	0.90	0.89	0.90
	Mean age at marriage	25.9	25.7	25.3	25.1	25.1	25.4
	Median age at marriage	25.4	25.2	24.7	24.8	24.3	24.6

Table 6: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals 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	0.96	0.96	0.93	0.96	0.97	0.96
	Mean closed interval	1.75	1.78	1.92	1.87	1.86	1.85
	Median closed interval	1.59	1.55	1.59	1.59	1.61	1.57
Rural	PPR	0.97	0.96	0.95	0.97	0.97	0.96
	Mean closed interval	1.69	1.81	1.81	1.87	1.88	1.85
	Median closed interval	1.54	1.56	1.54	1.54	1.62	1.57
Education							
Low	PPR	0.96	0.95	0.92	0.97	0.96	0.96
	Mean closed interval	1.73	1.84	1.95	1.87	1.93	1.89
	Median closed interval	1.61	1.58	1.61	1.60	1.65	1.59
Medium	PPR	0.94	0.96	0.94	0.96	0.98	0.97
	Mean closed interval	1.87	1.79	1.87	1.93	1.75	1.81
	Median closed interval	1.57	1.56	1.57	1.63	1.55	1.55
High	PPR	0.94	0.96	0.94	0.97	0.97	0.97
	Mean closed interval	1.83	1.77	1.83	1.80	1.88	1.82
	Median closed interval	1.55	1.55	1.55	1.57	1.62	1.56
		ADJUSTED ESTIMATES					
Residence							
Urban	PPR	0.96	0.95	0.93	0.96	0.97	0.96
	Mean closed interval	1.76	1.73	1.98	1.87	1.90	1.82
	Median closed interval	1.55	1.52	1.61	1.59	1.62	1.56
Rural	PPR	0.97	0.97	0.94	0.97	0.96	0.97
	Mean closed interval	1.68	1.86	1.74	1.87	1.83	1.87
	Median closed interval	1.50	1.60	1.53	1.61	1.60	1.59
Education							
Low	PPR	0.97	0.96	0.93	0.97	0.97	0.96
	Mean closed interval	1.85	1.86	2.00	1.90	1.96	1.92
	Median closed interval	1.62	1.60	1.71	1.59	1.67	1.61
Medium	PPR	0.97	0.96	0.95	0.96	0.97	0.96
	Mean closed interval	1.71	1.76	1.96	1.92	1.69	1.80
	Median closed interval	1.51	1.55	1.62	1.68	1.54	1.55
High	PPR	0.96	0.96	0.93	0.96	0.96	0.96
	Mean closed interval	1.65	1.79	1.69	1.72	1.89	1.78
	Median closed interval	1.48	1.54	1.44	1.53	1.60	1.53

Table 7: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals 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	0.87	0.81	0.79	0.94	0.89	0.90
	Mean closed interval	3.30	3.49	3.61	2.86	3.01	3.09
	Median closed interval	2.81	2.90	3.02	2.56	2.63	2.66
Rural	PPR	0.90	0.90	0.86	0.95	0.95	0.93
	Mean closed interval	3.18	3.20	3.40	2.75	2.74	2.93
	Median closed interval	2.73	2.72	2.86	2.49	2.45	2.56
Education							
Low	PPR	0.92	0.93	0.88	0.96	0.94	0.93
	Mean closed interval	3.06	3.06	3.35	2.75	2.78	2.92
	Median closed interval	2.66	2.64	2.82	2.49	2.48	2.55
Medium	PPR	0.89	0.85	0.84	0.96	0.93	0.92
	Mean closed interval	3.23	3.36	3.47	2.76	2.85	3.01
	Median closed interval	2.77	2.82	2.91	2.50	2.53	2.61
High	PPR	0.83	0.79	0.77	0.90	0.87	0.87
	Mean closed interval	3.43	3.53	3.65	3.03	3.08	3.19
	Median closed interval	2.90	2.93	3.07	2.66	2.67	2.72
		ADJUSTED ESTIMATES					
Residence							
Urban	PPR	0.87	0.82	0.79	0.93	0.89	0.90
	Mean closed interval	3.26	3.40	3.54	2.76	2.98	3.06
	Median closed interval	2.77	2.85	2.97	2.50	2.57	2.62
Rural	PPR	0.88	0.89	0.86	0.96	0.94	0.93
	Mean closed interval	3.22	3.28	3.51	2.86	2.77	2.97
	Median closed interval	2.78	2.78	2.92	2.55	2.50	2.59
Education							
Low	PPR	0.92	0.89	0.86	0.95	0.93	0.93
	Mean closed interval	3.03	2.96	3.30	2.75	2.83	2.95
	Median closed interval	2.68	2.63	2.86	2.51	2.53	2.58
Medium	PPR	0.90	0.85	0.82	0.96	0.93	0.92
	Mean closed interval	3.34	3.38	3.37	2.76	2.83	3.01
	Median closed interval	2.83	2.79	2.86	2.51	2.51	2.62
High	PPR	0.81	0.82	0.80	0.90	0.88	0.87
	Mean closed interval	3.33	3.61	3.80	2.96	3.03	3.14
	Median closed interval	2.79	3.00	3.19	2.59	2.59	2.65

Table 8: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals 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	0.77	0.73	0.68	0.88	0.82	0.83
	Mean closed interval	3.54	3.76	4.10	3.18	3.34	3.40
	Median closed interval	3.09	3.18	3.54	2.77	2.86	2.88
Rural	PPR	0.84	0.84	0.77	0.93	0.91	0.89
	Mean closed interval	3.38	3.50	3.89	2.97	3.05	3.21
	Median closed interval	2.94	2.95	3.30	2.65	2.68	2.77
Education							
Low	PPR	0.89	0.84	0.83	0.94	0.93	0.91
	Mean closed interval	3.24	3.48	3.74	2.95	3.01	3.14
	Median closed interval	2.86	2.94	3.12	2.64	2.67	2.73
Medium	PPR	0.80	0.82	0.74	0.89	0.86	0.85
	Mean closed interval	3.48	3.55	3.98	3.16	3.24	3.35
	Median closed interval	3.05	2.99	3.41	2.76	2.81	2.85
High	PPR	0.69	0.68	0.60	0.80	0.73	0.76
	Mean closed interval	3.70	3.84	4.22	3.42	3.55	3.58
	Median closed interval	3.27	3.28	3.70	2.93	3.02	3.01
		ADJUSTED ESTIMATES					
Residence							
Urban	PPR	0.77	0.74	0.68	0.88	0.82	0.84
	Mean closed interval	3.43	3.71	4.07	3.09	3.23	3.35
	Median closed interval	2.97	3.05	3.51	2.70	2.77	2.84
Rural	PPR	0.82	0.82	0.74	0.92	0.89	0.87
	Mean closed interval	3.52	3.51	3.99	3.06	3.17	3.29
	Median closed interval	3.07	3.03	3.40	2.71	2.77	2.81
Education							
Low	PPR	0.88	0.80	0.82	0.94	0.92	0.91
	Mean closed interval	3.24	3.35	3.80	2.97	3.06	3.15
	Median closed interval	2.88	2.87	3.28	2.68	2.70	2.76
Medium	PPR	0.79	0.81	0.71	0.89	0.86	0.85
	Mean closed interval	3.37	3.52	3.79	3.18	3.17	3.31
	Median closed interval	2.99	3.02	3.17	2.74	2.78	2.82
High	PPR	0.70	0.71	0.63	0.80	0.74	0.77
	Mean closed interval	3.93	4.03	4.48	3.28	3.53	3.62
	Median closed interval	3.36	3.36	3.99	2.77	2.94	3.01

Table 9: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals for progression from second birth to third 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	0.70	0.59	0.60	0.81	0.72	0.71
	Mean closed interval	3.70	3.66	4.02	3.34	3.42	3.55
	Median closed interval	3.13	3.13	3.42	2.90	2.96	3.01
Rural	PPR	0.77	0.71	0.68	0.86	0.84	0.83
	Mean closed interval	3.55	3.50	3.90	3.20	3.18	3.29
	Median closed interval	2.98	2.97	3.30	2.81	2.79	2.83
Education							
Low	PPR	0.80	0.73	0.73	0.84	0.88	0.73
	Mean closed interval	3.50	3.45	3.81	3.29	3.12	3.81
	Median closed interval	2.95	2.94	3.23	2.87	2.76	3.23
Medium	PPR	0.71	0.67	0.65	0.58	0.75	0.65
	Mean closed interval	3.69	3.55	3.96	3.76	3.41	3.96
	Median closed interval	3.12	3.02	3.37	3.29	2.96	3.37
High	PPR	0.64	0.53	0.52	0.84	0.59	0.52
	Mean closed interval	3.82	3.72	4.15	3.29	3.65	4.15
	Median closed interval	3.24	3.21	3.55	2.87	3.20	3.55
		ADJUSTED ESTIMATES					
Residence							
Urban	PPR	0.70	0.60	0.61	0.82	0.74	0.72
	Mean closed interval	3.74	3.62	3.96	3.24	3.27	3.49
	Median closed interval	3.10	3.02	3.35	2.81	2.85	2.94
Rural	PPR	0.75	0.68	0.65	0.83	0.81	0.79
	Mean closed interval	3.54	3.47	3.96	3.31	3.36	3.34
	Median closed interval	3.00	2.99	3.34	2.91	2.89	2.89
Education							
Low	PPR	0.81	0.72	0.73	0.90	0.86	0.86
	Mean closed interval	3.58	3.58	3.95	3.12	3.08	3.31
	Median closed interval	3.06	3.03	3.43	2.76	2.76	2.84
Medium	PPR	0.70	0.66	0.65	0.78	0.75	0.74
	Mean closed interval	3.58	3.46	3.93	3.44	3.41	3.49
	Median closed interval	2.97	3.01	3.28	3.02	2.96	2.95
High	PPR	0.64	0.53	0.52	0.66	0.61	0.59
	Mean closed interval	3.84	3.60	3.99	3.69	3.73	3.63
	Median closed interval	3.21	2.96	3.37	3.17	3.16	3.16

Table 10: Unadjusted and adjusted estimates of parity progression ratios and mean and median closed birth intervals for progression from fourth or higher order birth to next birth (4+ to 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	0.65	0.60	0.57	0.72	0.69	0.66
	Mean closed interval	3.31	3.47	3.59	3.12	3.30	3.35
	Median closed interval	2.94	3.03	3.19	2.79	2.91	2.95
Rural	PPR	0.73	0.70	0.68	0.81	0.79	0.77
	Mean closed interval	3.17	3.32	3.42	2.95	3.10	3.15
	Median closed interval	2.84	2.90	3.03	2.67	2.77	2.80
Education							
Low	PPR	0.74	0.73	0.71	0.82	0.80	0.79
	Mean closed interval	3.16	3.27	3.37	2.96	3.10	3.14
	Median closed interval	2.83	2.87	2.98	2.68	2.77	2.80
Medium	PPR	0.67	0.61	0.60	0.69	0.71	0.67
	Mean closed interval	3.28	3.46	3.55	3.20	3.27	3.35
	Median closed interval	2.92	3.02	3.15	2.85	2.89	2.95
High	PPR	0.45	0.50	0.47	0.52	0.57	0.53
	Mean closed interval	3.54	3.60	3.72	3.43	3.48	3.53
	Median closed interval	3.17	3.17	3.31	3.03	3.07	3.12
				ADJUSTED ESTIMATES			
Residence							
Urban	PPR	0.65	0.62	0.58	0.74	0.70	0.67
	Mean closed interval	3.23	3.43	3.37	3.11	3.24	3.26
	Median closed interval	2.86	3.04	3.01	2.76	2.87	2.89
Rural	PPR	0.71	0.68	0.67	0.78	0.78	0.76
	Mean closed interval	3.22	3.34	3.57	2.97	3.14	3.21
	Median closed interval	2.88	2.92	3.14	2.70	2.79	2.84
Education							
Low	PPR	0.73	0.71	0.69	0.81	0.79	0.78
	Mean closed interval	3.13	3.19	3.27	2.96	3.11	3.12
	Median closed interval	2.80	2.84	2.95	2.67	2.78	2.80
Medium	PPR	0.69	0.63	0.60	0.71	0.72	0.68
	Mean closed interval	3.37	3.68	3.69	3.24	3.27	3.37
	Median closed interval	3.03	3.14	3.22	2.90	2.89	2.96
High	PPR	0.46	0.51	0.48	0.53	0.58	0.55
	Mean closed interval	3.49	3.51	3.68	3.34	3.43	3.50
	Median closed interval	2.91	3.21	3.18	2.93	2.93	2.95

Table 11: 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	2.97	2.60	2.51	4.56	3.73	3.73
	Adjusted	3.03	2.62	2.55	4.67	3.87	3.79
Rural	Unadjusted	4.34	4.08	3.63	6.35	5.81	5.38
	Adjusted	3.88	3.79	3.34	5.66	5.20	4.90
Education							
Low	Unadjusted	4.85	4.39	4.03	6.60	6.12	5.34
	Adjusted	4.54	3.72	3.78	6.59	5.69	5.52
Medium	Unadjusted	3.58	3.26	3.09	3.98	4.40	3.97
	Adjusted	3.54	3.17	3.00	4.58	4.36	4.18
High	Unadjusted	2.14	2.28	2.09	3.14	2.65	2.68
	Adjusted	2.33	2.54	2.25	3.11	2.98	2.95
				TOTAL MARITAL FERTILITY RATES			
Residence							
Urban	Unadjusted	3.70	3.17	2.88	4.99	4.16	4.03
	Adjusted	3.73	3.20	2.90	5.09	4.31	4.08
Rural	Unadjusted	4.69	4.25	3.76	6.59	6.11	5.58
	Adjusted	4.36	4.00	3.53	5.98	5.59	5.19
Education							
Low	Unadjusted	5.10	4.59	4.07	6.77	6.29	5.44
	Adjusted	5.03	4.23	3.94	6.94	6.04	5.79
Medium	Unadjusted	4.00	3.60	3.25	4.23	4.71	4.11
	Adjusted	4.08	3.62	3.20	4.89	4.70	4.33
High	Unadjusted	2.94	2.80	2.55	3.79	3.25	3.17
	Adjusted	2.93	2.89	2.60	3.46	3.36	3.28

Note: TFRs and TMFRs in this table are calculated from PPRs in Tables 5-10.

Table 12: Decomposition of change in the total fertility rate and ts components between the 1993 and 2003 surveys: Method 1 (method of varying one thing at a time) (percent)

Measure in which change is decomposed	Component due to change in:						
	Composition by:		Effect of:		Baseline		
	Residence	Education	Residence	Education	hazard	Residual	Total
			PERIOD ANALYSIS				
pB	-1.0	-4.4	-14.6	-12.0	134.2	-2.2	100.0 (0.06)
pM	0.6	-0.3	-8.8	-29.1	142.7	-5.1	100.0 (-0.03)
p1	0.8	10.0	40.7	-27.7	72.8	3.4	100.0 (-0.06)
p2	-0.4	15.6	-1.0	1.4	83.4	1.0	100.0 (-0.08)
p3	0.2	13.0	-6.1	2.1	81.3	9.5	100.0 (-0.10)
p4+	-3.5	13.7	22.6	-0.3	60.3	7.2	100.0 (-0.06)
TFR	-0.9	21.3	26.4	-7.5	64.7	-4.2	100.0 (-0.51)
TMFR	-1.0	13.8	14.0	-9.7	85.7	-2.8	100.0 (-0.85)
							100.0
			COHORT ANALYSIS				
pB	6.9	-63.9	167.7	65.6	-53.2	-23.1	100.0 (0.01)
pM	5.4	99.9	210.6	-232.1	59.1	-42.9	100.0 (-0.00)
p1	1.9	2.9	-20.2	11.6	84.9	18.9	100.0 (-0.03)
p2	-0.2	20.3	-26.3	-15.7	101.3	20.5	100.0 (-0.04)
p3	1.1	27.8	38.7	-0.2	26.1	6.4	100.0 (-0.08)
p4+	-2.3	38.3	45.2	-21.2	42.1	-2.1	100.0 (-0.05)
TFR	-1.0	36.3	22.8	-20.2	63.1	-1.1	100.0 (-0.85)
TMFR	-0.7	32.6	28.5	-16.8	59.1	-2.7	100.0 (-0.94)

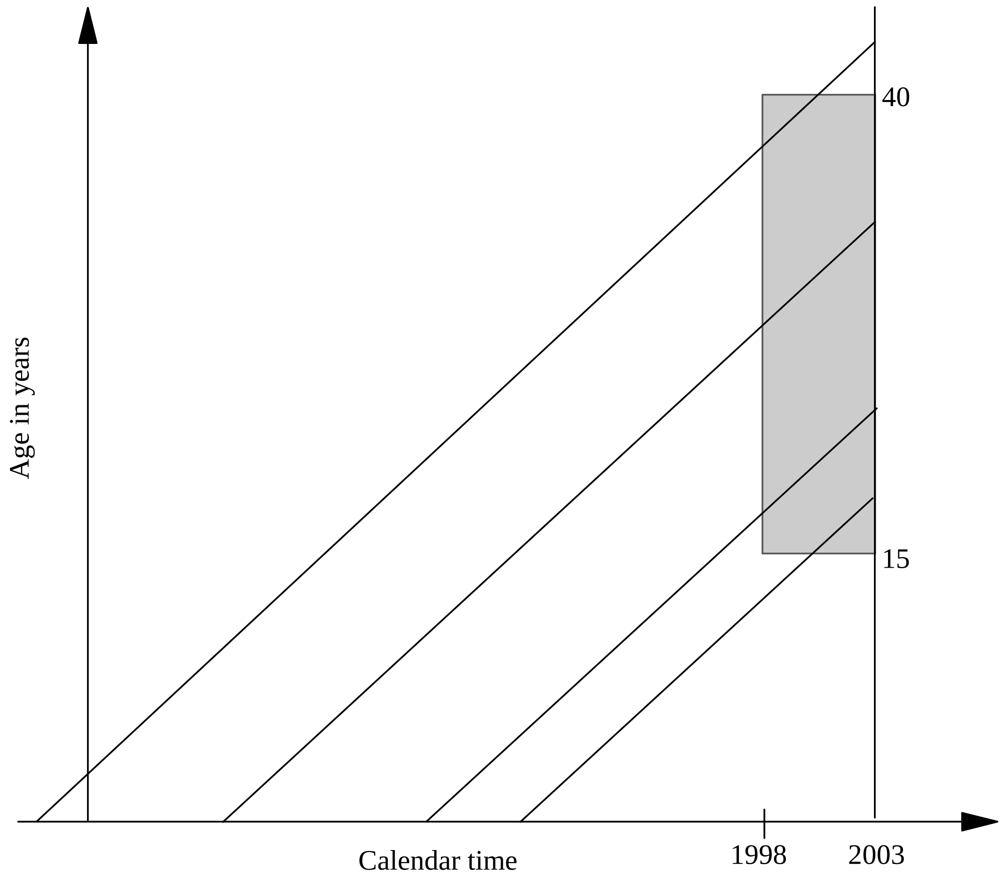
Notes: The underlying CLL models all have residence and education as time-varying predictor variables and a cubic specification of time-varying effects. Starting and ending values of the fertility measures to be decomposed were obtained by substituting interval-specific mean values of U (representing residence) and M and H (representing education) in the underlying CLL model equations. In the period analysis, fertility measures pertain to the 5-year period immediately preceding each survey. In the cohort analysis, cohorts are defined as women age 40-49 at the time of each of the two surveys. Numbers in parentheses in the last column indicate change in the PPR or TFR or TMFR between the two surveys.

Table 13: Decomposition of change in the total fertility rate and ts components between the 1993 and 2003 surveys: Method 2 (method of cumulatively varying one thing at a time) (percent)

Measure in which change is decomposed	Component due to change in:					
	Composition by:		Effect of:		Baseline	
	Residence	Education	Residence	Education	hazard	Total
			PERIOD ANALYSIS			
pB	-1.0	-14.1	-16.0	-12.3	143.3	100.0 (0.06)
pM	0.6	-0.2	-6.8	-34.4	140.8	100.0 (-0.03)
p1	0.8	10.0	43.3	-32.6	78.5	100.0 (-0.06)
p2	-0.4	15.5	0.2	5.3	79.4	100.0 (-0.08)
p3	0.2	13.0	-5.7	2.2	90.3	100.0 (-0.10)
p4+	-3.5	13.7	21.5	8.4	59.9	100.0 (-0.06)
TFR	-0.9	26.1	27.0	-4.5	52.3	100.0 (-0.51)
TMFR	-1.0	13.9	14.2	-7.6	80.5	100.0 (-0.85)
			COHORT ANALYSIS			
pB	6.9	-63.0	173.8	60.8	-78.6	100.0 (0.01)
pM	5.4	99.8	219.2	-310.7	86.3	100.0 (-0.00)
p1	1.9	3.0	-20.5	11.7	104.0	100.0 (-0.03)
p2	-0.2	20.2	-26.7	-16.6	123.3	100.0 (-0.04)
p3	1.1	27.9	43.2	-1.5	29.3	100.0 (-0.08)
p4+	-2.3	38.3	44.3	-26.8	46.6	100.0 (-0.05)
TFR	-1.0	36.4	19.4	-19.3	64.5	100.0 (-0.85)
TMFR	-0.7	32.8	25.1	-16.2	59.0	100.0 (-0.94)

Note: See notes to Table 12.

Figure 1: Lexis diagram illustrating censoring when setting up the expanded data file for calculating a multivariate period life table for progression from 15th birthday to first marriage, based on retrospective data from a survey in 2003



Notes: The shaded area, which is 5 years wide and 25 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-month segments, corresponding to person-months in the expanded data set. Person-months falling outside the shaded area are censored and not included in the expanded data set.