

Revised March 26, 2007

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 23rd Population Census Conference, Christchurch, New Zealand, 16–18
April 2007.

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). Components of the TFR include PPRs, the total marital fertility rate (TMFR), and the TFR itself as measures of the quantum of fertility, and mean and median ages at first marriage and mean and median closed birth intervals by birth order as measures of the tempo or timing of fertility. The methods are illustrated by application to data from the 1993, 1998, and 2003 Demographic and Health Surveys (DHS) in the Philippines.

This paper develops multivariate methods to estimate effects of socioeconomic variables on the total fertility rate and on the trend in the total fertility rate (TFR). The methods are applicable to individual-level data. The methods are the same for period measures and cohort measures of the TFR, the only difference being the way that the underlying data sets for the period and cohort analyses are constructed. Because the application to period measures is more innovative than the application to cohort measures, the paper focuses on period measures.

For the multivariate methods to be applicable, the TFR must be calculated from parity progression ratios (PPRs). In our illustrative examples, a woman's parity is defined in the usual way as the number of children she has ever borne, but with parity zero subdivided into two states: never-married with no children and ever-married with no children. Parity progression ratios (PPRs) are the fractions of women who progress from their own birth to first marriage, from first marriage to first birth, from first birth to second birth, and so on. The PPRs so obtained are aggregated to a TFR and a total marital fertility rate (TMFR). TFR, TMFR, and PPRs are measures of the quantum of fertility. Mean and median ages at first marriage and mean and median closed birth intervals by birth order are measures of the timing or tempo of fertility. Except for the TFR itself, these measures of quantum and tempo are what we mean here by the components of the TFR. The multivariate methods are applicable to not only the TFR but also each of its components.

We focus on the TFR calculated from PPRs instead of the TFR calculated from ASFRs (TFR_{asfr}) for two reasons: The first is that a multivariate method for analyzing factors affecting TFR_{asfr} calculated from individual data has already been developed and applied by Schoumaker (2004), who used Poisson regression for this purpose. The second reason is that, from an explanatory point of view, age-specific fertility rates are not ideal measures of the components of the TFR. A woman's decision about whether to have a next birth does not depend primarily on her age. More important considerations are her marital status, time elapsed since marriage if she is married but does not yet have any children, time elapsed since her last birth if she already has children, and the number of children that she already has. The TFR calculated from PPRs takes all these considerations into account. Henceforth in this paper, "TFR" and "TMFR" refer to the total fertility rate and the total marital fertility rate calculated from PPRs.

By way of illustration, we apply the methodology to data from three demographic and health surveys (DHS) undertaken in the Philippines in 1993, 1998, and 2003. We use a multivariate survival model—the complementary log-log (CLL) model—to model parity progression. Period measures of the TFR and its components are estimated for the 5-year period before each survey. The design of these three surveys is described in more detail in the basic survey reports, which include questionnaires and detailed information about sampling procedures (Philippines National Statistics Office and Macro International 1994; Philippines National Statistics Office, Philippines Department of Health, and Macro International 1999; Philippines National Statistics Office and ORC Macro 2004).

METHODOLOGY AND ILLUSTRATIVE RESULTS FOR A CROSS-SECTIONAL ANALYSIS

By a cross-sectional analysis, we mean an analysis based on a single time period using data from a single survey. Illustrative results are actually shown for three surveys, each analyzed independently of the other two.

Methodology

A discrete-time survival model, the complementary log-log (CLL) model (Prentice and Gloeckler 1978), is used as the multivariate survival model for purposes of modeling parity progression. As noted by Allison (1982; 1995), the CLL model is preferable to the more commonly used discrete-time logit model, because the CLL model is derived from the continuous-time Cox proportional hazards model (Cox 1972). Therefore, coefficients of predictors in the CLL model, but not in the discrete-time logit model, have the same interpretation as in the continuous-time Cox proportional hazards model, namely that a one-unit increase in a predictor variable multiplies the underlying continuous-time hazard $h_i(t)$ by $\exp(b)$, where b is the coefficient of the predictor variable and $\exp(b)$ is the relative risk. Although the CLL model is derived from the Cox proportional hazards model, it can easily be tricked into handling non-proportionality (i.e., time-varying predictor variables and time-varying effects of predictor variables). Indeed, this is one of its principal advantages over the original Cox model (Allison 1982, 1995).

Another advantage of the CLL model over the Cox model is that the CLL model is estimated by maximum likelihood and therefore yields a baseline hazard function. By contrast, the Cox model is derived by partial likelihood and does not yield a baseline hazard function. A baseline hazard function is necessary for the methods developed in this paper.

The CLL model is applied not to the original “person sample” but instead to an “expanded sample” of person-year observations created from the original person observations. The expanded sample makes it easy to include time-varying predictor variables in the CLL model. For example, if a person moves from urban to rural, some of the person-year observations created for that person are coded as rural and some are coded as urban. The CLL model can also handle time-varying effects of predictor variables, by interacting socioeconomic variables with life table time or some function of life table time such as time-squared.

The CLL model also handles left-censoring as well as right-censoring (Allison 1995). This enables application of the model to period data as well as cohort data. Previous studies have applied the CLL model and other discrete-time survival models to cohort data. An innovative aspect of the present paper is the application of the CLL model to period data. This is done by treating person-year observations before or after the period under consideration as censored.

The form of the CLL model is basically the same for each parity transition. In our illustrative application to Philippines data, the main difference is that, in the case of the transition from birth to first marriage (B–M), the multivariate life tables are truncated at 25 years of duration in parity (the difference between the beginning and ending life table ages of 15 and 40), whereas in the case of higher-order transitions they are truncated at 10 years of duration in parity. First marriages after age 40 and next births after 10 years duration in parity are rare and are

ignored.

Two socioeconomic predictor variables are included in the illustrative analysis using Philippines DHS survey data: urban-rural residence (specified by a dummy variable U) and education (specified by dummy variables M and H , representing medium and high education with low education as the reference category), as assessed at the time of the survey. These variables are treated as time-invariant in the absence of adequate information on their values in earlier years before the survey.

In the case of the birth-to-first-marriage (B–M) transition, the CLL model is

$$P = 1 - \exp\{-\exp[a + b_0T_0 + b_1T_1 + \dots + b_{23}T_{23} + U(c+dt+et^2) + M(f+gt+ht^2) + H(j+kt+mt^2)]\} \quad (1)$$

where P is the predicted value of the probability of failure (also called the discrete hazard) in a life table time interval; T_0, \dots, T_{23} are 24 dummy variables representing the first 24 of 25 life table time intervals; t is a counter variable (equal to 1, 2, ..., 25) that also denotes life table time interval; a is an intercept term (implying that $P = 1 - \exp[-\exp(a)]$ for the 25th life table time interval when all predictors equal zero), and $b_0, \dots, b_{23}, c, d, e, f, g, h, j, k,$ and m are coefficients to be fitted (along with the intercept a) to the data.

In the context of equation (1), a “failure” is a first marriage. In the case of higher-order parity transitions, a failure is a next birth. Fitting the CLL model to the data is accomplished by maximum likelihood methods using computational procedures recommended by Allison (1995).

Although, for higher-order transitions (i.e, higher than B–M), the birth histories in the Philippines DHS surveys are specified by month, we aggregate months into years. This is done because monthly data sometimes result in empty cells, in which case the maximum likelihood estimation procedure for fitting a CLL model does not converge to a solution.

Time enters the analysis in three different ways: two measures of life table time, as shown in equation (1), and one measure of calendar time (e.g., the year 1999), which is used when creating the expanded person-year data set for a particular calendar time period. Each person-year record in the expanded data set for the B–M transition contains values of $T_0, \dots, T_{23}, t,$ and CALTIME (representing the calendar year in which the person-year observation is located), as well as values of the socioeconomic predictors $U, M,$ and H .

In equation (1), effects of socioeconomic predictor variables are specified as time-varying, where the time variation refers to life table time. For example, the effect of a one-unit increase in U (from 0 to 1) — i.e., the effect of urban relative to rural on the hazard — is to multiply the underlying continuous-time hazard of a first marriage for rural by $\exp(c+dt+et^2)$, which is the relative risk. Preliminary analysis showed that a quadratic specification of relative risk best describes time variation in the effects of $U, M,$ and H on P (Retherford et al. 2006).

A time-varying specification of the effect of U on the probability of first marriage is necessary because the effect of urban residence, relative to rural residence, is to lower the

probability of first marriage at younger ages and increase it at older ages (because urban marriages tend to be postponed to later ages, relative to rural marriages). Thus the effect of urban residence on the risk of progression to first marriage is not constant over life table time; i.e., the effect is not proportional. Similarly, the effect of education is modeled as time-varying, because the effect of more education is also to lower the probability of first marriage at younger ages and raise it at older ages. The effects of U , M , and H on the probability of next birth are also modeled as time-varying, again with a quadratic specification, for similar reasons.

The set of estimated values of P (one for each life table time interval) is called the discrete hazard function, which in equation (1) is a multivariate hazard function specified for particular values of the socioeconomic predictor variables. From the discrete hazard function, it is a simple matter to calculate the rest of the life table, which, therefore, is itself multivariate. PPR is calculated from this life table as one minus the proportion “surviving” (still single, in the case of the B–M transition) at the end of the life table. It is a simple matter also to calculate mean and median failure times from the life table. In the case of the B–M transition, the mean and median failure times, when added to 15 (age at the start of the life table), are mean and median ages at first marriage. In the case of higher-order transitions, mean and median failure times are mean and median closed birth intervals. Because the life table is itself multivariate, every measure calculated from it is also multivariate, including mean and median failure times. (The medians 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.)

Unadjusted and adjusted estimates of TFR (or one of its components) by categories of a predictor variable are calculated using the logic of what is known in the demographic literature as multiple classification analysis (MCA) (Andrews, Morgan, and Sonquist 1969; Retherford and Choe 1993). In MCA, “unadjusted” means “without controls”, and “adjusted” means “with controls”.

For a particular parity transition, unadjusted estimates of the discrete hazard function by urban/rural residence, for example, are calculated from a CLL model that includes U as the sole socioeconomic predictor variable with time-varying effects, with a quadratic specification of the time variation. Thus, in the case of equation (1) for the B–M transition, one drops the last two terms containing M and H . The discrete hazard function for urban is then calculated by setting $U = 1$ in the fitted equation, and the discrete hazard function for rural is calculated by setting $U = 0$ in the fitted equation. Unadjusted urban and rural life tables are then calculated from the unadjusted urban and rural discrete hazard functions, and unadjusted urban and rural values of PPR and mean and median ages at first marriage are calculated from the unadjusted urban and rural life tables.

Adjusted estimates of the discrete hazard function by urban/rural residence for the B–M transition are calculated from equation (1) with all of the socioeconomic predictor variables U , M , and H included. Education, as indicated by M and H , is viewed as the control variable. To obtain the adjusted discrete hazard function for urban, one sets $U = 1$ and M and H equal to their interval-specific mean values in the B–M person-year data set to which the CLL model is fitted. In this context, “interval” means life table time interval. To obtain the adjusted discrete hazard function for rural, one sets $U = 0$ and M and H equal to the same interval-specific mean values

that were used to calculate the adjusted discrete hazard function for urban. In this way M and H are held constant or “controlled” at their interval-specific mean values when U is varied from 0 to 1. Adjusted urban and rural life tables are then calculated from the adjusted urban and rural discrete hazard functions, and adjusted urban and rural values of PPR and mean and median ages at first marriage are calculated from the adjusted urban and rural life tables.

The procedure for calculating unadjusted and adjusted estimates for each higher-order parity transition is similar. In these calculations, each higher-order transition has its own set of interval-specific mean values of M and H derived from the person-year data set for that parity transition.

Because the units of analysis for CLL models are person-years instead of persons (where “year” in person-year pertains to a particular life table time interval), and because control variables are held constant at their *interval-specific* mean values, the MCA approach to calculating adjusted estimates is straightforward, even when control variables or their effects are time-varying.

Basic notation for PPRs and parity transitions is:

p_B	PPR for transition from a woman’s own birth to her first marriage (B–M)
p_M	PPR for transition from first marriage to first birth (M–1)
p_1	PPR for transition from first 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 to fifth birth (4–5)
p_5	PPR for transition from fifth to sixth birth (5–6)
p_{6+}	PPR for transition from sixth or higher-order birth to next higher-order birth (6+ to 7+)

In our illustrative analysis using Philippines DHS data, life tables for the B–M transition actually start at age 15 rather than birth, because first marriages or first unions below age 15 are rare. A parity transition cutoff at 6+ to 7+ is also found to be appropriate. It is assumed that no births occur before first marriage. This works, because in the Philippines DHS surveys, informal unions are treated as first marriages. In the case of the small number of premarital or pre-union births, we recoded date of first marriage or date of first non-formalized union back to date of first premarital birth.

TFR is calculated from the PPRs as

$$\text{TFR} = p_B p_M + p_B p_M p_1 + p_B p_M p_1 p_2 + p_B p_M p_1 p_2 p_3 + p_B p_M p_1 p_2 p_3 p_4 + p_B p_M p_1 p_2 p_3 p_4 p_5 + p_B p_M p_1 p_2 p_3 p_4 p_5 p_{6+} / (1 - p_{6+}) \quad (2)$$

The term $p_B p_M$ is the expected number of first births, the term $p_B p_M p_1$ is the expected number of second births, and so on. As explained by Feeney (1986), the term $p_{6+} / (1 - p_{6+})$ is obtained by assuming that p_6 and all higher-order PPRs equal p_{6+} and pulling out a geometric series. (Recall that if r is a positive number less than one, the geometric series $r + r^2 + r^3 + \dots = r / (1 - r)$.) The formula for TMFR is the same as the formula for TFR in equation (1), except that p_B is set equal to one.

Unadjusted values of TFR by urban/rural residence are calculated from equation (2) using unadjusted PPRs by urban/rural residence. Adjusted values of TFR by urban/rural residence are calculated from equation (2) using adjusted PPRs by urban/rural residence. Unadjusted and adjusted values of TMFR are similarly calculated from equation (2), but with p_B set equal to one.

Further details of methodology may be found in an earlier working paper, which elaborates on the illustrative results shown below and includes, in addition to the period analysis presented here, a cohort analysis based on women age 40–49 in the same three Philippines DHS surveys (Retherford et al. 2006).

Illustrative results

Illustrative results are shown in Figure 1 and Tables 1–5 for the Philippines. In these tables, periods are defined as the 5-year period before each of the three surveys. Table 1 shows the distributions of the original samples over residence and education categories, and Table 2 shows the expanded samples of person-year observations derived from the original person observations.

- Tables 1 and 2 about here -

Figure 1 illustrates non-proportionality of the effect of education on the hazard of first marriage (B–M transition), based on equation (1) above. The figure uses data from the Philippines 1993 DHS to graph the relative risks $\exp(f+gt+ht^2)$ and $\exp(j+kt+mt^2)$ against t to show the extent to which the time-varying specifications of the effects of medium and high education on the hazard of first marriage, relative to the effect of low education, depart from the time-invariant proportional specification. In the proportional case, the graphs would be horizontal lines. The figure indicates increasing postponement of marriage with more education, inasmuch as (1) each of the two curves starts out far below one, rises above one, and then falls to far below one, and (2) the curve for high education is shifted to the right, relative to the curve for medium education.

- Figure 1 about here -

Table 3 shows, again for the B–M transition, unadjusted and adjusted estimates of p_B and mean and median ages at first marriage for each of the three surveys. Estimates are based on a separate analysis for each survey. In the case of adjusted estimates, separate analyses mean that a separate set of interval-specific mean values of the control variables is used for each survey, so that control variables are held constant within surveys but not between surveys. The table shows, quite unexpectedly, that p_B rises over time in every category of residence and education in both the unadjusted and adjusted cases. In any particular time period, p_B 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. Urban/rural differences in p_B and mean and median ages at first marriage are reduced by the introduction of controls for education, and education group differences in these measures are reduced by the introduction of controls for residence.

- Table 3 about here -

Table 4 shows unadjusted and adjusted estimates of p_3 and mean and median closed birth intervals for the transition from third to fourth birth, which is a key transition because many Filipino women stop childbearing after the third birth. For each period, p_3 tends to be higher and birth intervals shorter for rural than for urban. p_3 likewise tends to decrease and birth intervals to increase as education increases. Over time, p_3 tends to fall and birth intervals to increase in each of the various residence and education categories. Adjustment of the estimates tends to reduce residence differentials and education differentials in both p_3 and birth intervals.

- Table 4 about here -

Tables for parity transitions 1-2, 2-3, 4-5, 5-6, and 6+ to 7+ show patterns similar to those in Table 4 for the 3-4 transition. The table for the M-1 transition, on the other hand, shows small socioeconomic differentials and small changes over time, reflecting the fact that almost everyone who gets married wants to have at least one child. Because these tables are voluminous, they are not shown; but they can be found in the above-mentioned working paper (Retherford et al. 2006).

Table 5 shows unadjusted and adjusted estimates of TFR and TMFR, calculated from unadjusted and adjusted estimates of PPRs. TFR and TMFR are always higher for rural than for urban, and always lower for women with more education. Adjustment substantially reduces residence differentials and education differentials in TFR and TMFR. Both TFR and TMFR tend to fall over time. TMFR tends to fall more steeply than TFR, which is expected, because, as seen earlier, p_B tends to rise over time, thereby offsetting to some extent the declines in the higher-order PPRs from which TMFR is calculated.

- Table 5 about here -

The various CLL models that underlie Tables 3-5 have a large number of coefficients. The SAS program that we have used to estimate the CLL models indicates the standard error of each estimated coefficient, but the program does not provide standard errors for the PPRs, mean and median failure times, TFRs, and TMFRs that we have calculated from the fitted models. Estimation of these standard errors would require bootstrap methods, but we have not done this. Instead, we have employed an alternative approach that compares nested models that progressively include more variables. For each parity transition, we have compared three models: the first with no socioeconomic predictor variables, the second with residence only, and the third with both residence and education. For every transition except M-1 and 6+ to 7+, likelihood-ratio tests show that, at the 5 percent level of significance, the second model differs significantly from the first, and the third model differs significantly from the second at the 5 percent level. In the case of the transition from 6+ to 7+, the first model does not differ significantly from the second, but the second model does differ significantly from the third.

METHODOLOGY AND ILLUSTRATIVE RESULTS FOR ANALYSIS OF TRENDS

Methodology

The multivariate analysis of trend in the TFR (or one of its components) is rather similar to the

cross-sectional single-survey analyses described above. Only a few simple changes in methodology are needed. We again use the three Philippines DHS surveys to illustrate the methodology.

The first step in the trend analysis is to pool the three person-year samples (for the 5-year period preceding each of the three surveys) pertaining to each parity transition. As already mentioned, each person-year observation in each of the original three person-year samples and in the pooled person-year sample has a value of CALTIME (calendar year) attached to it. The value of CALTIME indicates the calendar year in which the person-year observation is located. New variables PERIOD2 and PERIOD3 are defined, based on each person-year observation's value of CALTIME. In the pooled sample, (PERIOD2, PERIOD3) = (0,0) for all person-year observations in the earliest five-year period, (1,0) for all person-year observations in the second five-year period, and (0,1) for all person-year observations in the third five-year period.

To calculate unadjusted estimates of the trends in p_B and mean and median ages at first marriage, based on the pooled data set, one estimates a CLL model for progression to first marriage that includes (in addition to the 24 dummy variables indicating the 25 life table time intervals) only PERIOD2 and PERIOD3 as predictor variables. An unadjusted discrete hazard function is then estimated for each of periods 1, 2, and 3 by setting (PERIOD2, PERIOD3) alternatively to (0, 0), (1, 0), and (0, 1) in the fitted model. Unadjusted life tables for the three periods are then calculated from the three unadjusted discrete hazard functions. Unadjusted values of p_B and mean and median ages at first marriage are then calculated from the three unadjusted life tables. These are the unadjusted trends.

Adjusted trends in p_B and mean and median ages at first marriage are similarly calculated, the only difference being that the underlying CLL model is expanded to include residence and education in the set of predictor variables, with quadratic specifications of their time-varying effects. Estimates of p_B and mean and median ages at marriage are then calculated from the model in the same way as in the unadjusted case, but this time with U , M , and H held constant at their interval-specific mean values in the pooled data set of person-year observations for the B–M transition. Coefficients of the socioeconomic predictors are already the same (i.e., held constant) for all three five-year time periods, because only one set of coefficients of these predictors is estimated when using the pooled data set. If the adjusted values of p_B (or mean or median age at first marriage) are found to be the same for all three periods, we would provisionally conclude that changes in population composition by residence and education and changes in the effects of residence and education explain the trend in p_B (provisionally because there may be other predictor variables affecting marriage that are correlated with residence and education but not included in the model).

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

Illustrative results

Sample results for the trend in TFR and TMFR are shown in Table 6. The table shows that the adjusted trends in TFR and TMFR are not as steeply declining as the unadjusted trends, indicating that residence and education explain some of the unadjusted trend. Comparison of the first and third surveys indicates that the adjusted decline in TFR is 27 percent less than the unadjusted decline, indicating that 27 percent of the unadjusted decline in TFR is explained. Similarly, the adjusted decline in TMFR is 13 percent less than the unadjusted decline, indicating that 13 percent of the unadjusted decline in TMFR is explained.

- Table 6 about here -

DISCUSSION AND CONCLUSION

The methodology developed in this paper and applied illustratively to Philippines data assumes that no (or very few) 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 an additional subset of dummy variables to represent whatever marital status categories are deemed appropriate (e.g., a dummy variable for never-married (NEVMAR) with ever-married as the reference category). One then calculates PPRs, mean and median ages at marriage, and mean and median closed birth intervals, TFR, and TMFR for each marital status category by setting NEVMAR alternatively to 0 and 1. The variable NEVMAR can be interacted with the socioeconomic variables if it is thought that the effects of socioeconomic variables are not the same for never-married and ever-married. Thus the multivariate methodology developed here is more generally applicable than our Philippines example might seem to imply.

For the methodology to realize its full potential for fertility analysis, demographic and health surveys need to collect integrated event histories covering not only marriages and fertility but also, for example, education, work, and migration in order to take advantage of the flexibility of the models. This flexibility, involving separate specifications of calendar time and life table time, implies considerable potential for solving difficult problems of two-way causation by means of lagged predictors, time-varying predictors, and time-varying effects of predictors, where in this context lagged time and time variation in predictors and their effects refer to life table time, not calendar time. For example, a variable representing a woman's work status can be lagged one or two years behind the time in the life table where she is at risk of a birth, so that causation runs from work to fertility but not from fertility to work, and this can be done without changing the calendar time variables representing time periods or cohorts.

ACKNOWLEDGMENTS

We thank Paul Allison, Minja Choe, and Tom Pullum for helpful advice and Hassan Eini-Zinab for 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.

PhilippinesDecomp. Revised October 19, 2006

REFERENCES

Allison, P. 1982. Discrete-time methods for the analysis of event histories. In S. Leinhardt (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.

Allison, P. 1995. *Survival Analysis Using SAS: A Practical Guide*. Cary, N. C.: SAS Institute Inc.

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

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

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

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.

Retherford, R. D., N. Ogawa, and R. Matsukura. 2006. *Multivariate Analysis of Parity Progression-Based Measures of the Total Fertility Rate and its Components Using Individual-Level Data*. East-West Center Working Papers, Population and Health Series, No. 119. Honolulu: East-West Center. (Downloadable at <http://www.eastwestcenter.org/stored/pdfs/POPwp119.pdf>)

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

Revised 3/8/07

Table 1: Percent distribution of women age 15-49 by residence and education: 1993, 1998, and 2003 DHS surveys, Philippines

Survey year	Education	Residence		
		Urban	Rural	Total
1993	Low	13.1	21.8	34.9
	Medium	21.8	17.3	39.1
	High	17.7	8.3	26.0
	Total	52.6	47.4	100.0
		(N = 14949)		
1998	Low	9.4	21.9	31.3
	Medium	20.4	20.5	40.9
	High	18.3	9.5	27.8
	Total	48.1	51.9	100.0
		(N = 13983)		
2003	Low	9.0	16.5	25.5
	Medium	24.5	19.8	44.3
	High	21.1	9.2	30.3
	Total	54.5	45.5	100.0
		(N = 13633)		

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 for the period analysis (5-year period before each survey): 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	5-6	6+ - 7+
1993	23846	3753	7153	7501	6973	5709	4024	8749
1998	21601	3846	6838	6964	6526	5134	3697	7639
2003	19733	4210	7929	7507	6200	4699	3214	6297

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

		Unadjusted			Adjusted		
		1988-92	1993-97	1998-02	1988-92	1993-97	1998-02
Residence							
Urban	PPR (p_B)	0.83	0.83	0.90	0.82	0.82	0.89
	Mean age at marriage	24.7	24.6	23.8	24.9	24.8	24.0
	Median age at marriage	23.8	24.1	23.1	24.1	24.2	23.4
Rural	PPR (p_B)	0.91	0.96	0.96	0.89	0.95	0.95
	Mean age at marriage	22.8	22.5	22.0	23.6	23.1	22.9
	Median age at marriage	21.8	21.8	21.0	22.7	22.5	22.1
Education							
Low	PPR (p_B)	0.93	0.91	0.97	0.92	0.89	0.96
	Mean age at marriage	21.9	21.4	20.3	22.3	21.8	20.6
	Median age at marriage	20.3	20.1	19.2	20.6	20.4	19.5
Medium	PPR (p_B)	0.87	0.89	0.94	0.87	0.88	0.94
	Mean age at marriage	23.1	23.1	22.4	23.2	23.3	22.4
	Median age at marriage	22.1	22.2	21.5	22.2	22.4	21.6
High	PPR (p_B)	0.80	0.88	0.88	0.80	0.88	0.88
	Mean age at marriage	25.9	25.5	25.2	25.8	25.6	25.1
	Median age at marriage	25.4	25.0	24.6	25.4	25.1	24.6
Total	PPR (p_B)	0.86	0.89	0.92			
	Mean age at marriage	23.9	23.8	23.2			
	Median age at marriage	23.0	23.0	22.4			

Notes: Estimates are based on separate CLL models for each of the three surveys. In the calculation of adjusted estimates for a particular survey, the control variable is held constant at its interval-specific mean values pertaining to that survey; in other words, a different set of means is used for each survey. Estimates for "Total" are based on CLL models in which the dummy variables presenting life table time intervals are the only predictor variables. Estimates pertain to the 5-year period before each survey. Calendar years are defined as years before the survey, not January 1 - December 31. In the case of the 1993 survey, for example, the year before the survey occurs partly in 1993 and partly in 1992. But it occurs mostly in 1992 and is therefore labeled 1992. The medians shown in the table are true medians, based on all failures (in this case all first marriages) 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.

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

		Unadjusted			Adjusted		
		1988-92	1993-97	1998-02	1988-92	1993-97	1998-02
Residence							
Urban	PPR (p_3)	0.71	0.61	0.59	0.71	0.63	0.61
	Mean closed interval	3.32	3.29	3.51	3.29	3.26	3.50
	Median closed interval	2.67	2.63	2.76	2.66	2.62	2.76
Rural	PPR (p_3)	0.78	0.73	0.70	0.76	0.70	0.67
	Mean closed interval	3.10	3.13	3.58	3.13	3.16	3.61
	Median closed interval	2.59	2.62	2.84	2.60	2.63	2.83
Education							
Low	PPR (p_3)	0.82	0.76	0.74	0.81	0.74	0.73
	Mean closed interval	3.06	3.13	3.48	3.08	3.15	3.46
	Median closed interval	2.60	2.62	2.83	2.61	2.62	2.81
Medium	PPR (p_3)	0.72	0.67	0.64	0.72	0.67	0.64
	Mean closed interval	3.28	3.01	3.48	3.27	3.01	3.47
	Median closed interval	2.62	2.52	2.77	2.62	2.51	2.76
High	PPR (p_3)	0.65	0.55	0.53	0.65	0.56	0.54
	Mean closed interval	3.42	3.59	3.74	3.39	3.56	3.76
	Median closed interval	2.73	2.80	2.85	2.72	2.79	2.86
Total	PPR (p_3)	0.74	0.67	0.64			
	Mean closed interval	3.2	3.2	3.5			
	Median closed interval	2.6	2.6	2.8			

Notes: Estimates are based on separate CLL models for each of the three surveys. In the calculation of adjusted estimates for a particular survey, the control variable is held constant at its interval-specific mean values pertaining to that survey; in other words, a different set of means is used for each survey. Estimates for "Total" are based on CLL models in which the dummy variables presenting life table time intervals are the only predictor variables. Estimates pertain to the 5-year period before each survey. Calendar years are defined as years before the survey, not January 1 - December 31. In the case of the 1993 survey, for example, the year before the survey occurs partly in 1993 and partly in 1992. But it occurs mostly in 1992 and is therefore labeled 1992. The medians shown in the table are true medians, based on all failures (in this case all first marriages) 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.

Table 5: Unadjusted and adjusted period estimates 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

		TFR			TMFR		
		1988-92	1993-97	1998-02	1988-92	1993-97	1998-02
Residence							
Urban	Unadjusted	3.12	2.61	2.58	3.74	3.14	2.87
	Adjusted	3.13	2.62	2.59	3.80	3.19	2.90
Rural	Unadjusted	4.22	4.06	3.59	4.66	4.25	3.73
	Adjusted	3.84	3.81	3.31	4.34	4.00	3.48
Education							
Low	Unadjusted	4.64	4.00	3.89	5.02	4.38	4.01
	Adjusted	4.49	3.68	3.69	4.90	4.16	3.84
Medium	Unadjusted	3.59	3.17	3.07	4.11	3.57	3.25
	Adjusted	3.59	3.13	3.04	4.13	3.56	3.23
High	Unadjusted	2.42	2.55	2.28	3.02	2.91	2.59
	Adjusted	2.45	2.60	2.32	3.05	2.96	2.63
Total		3.60	3.23	2.99	4.17	3.64	3.24

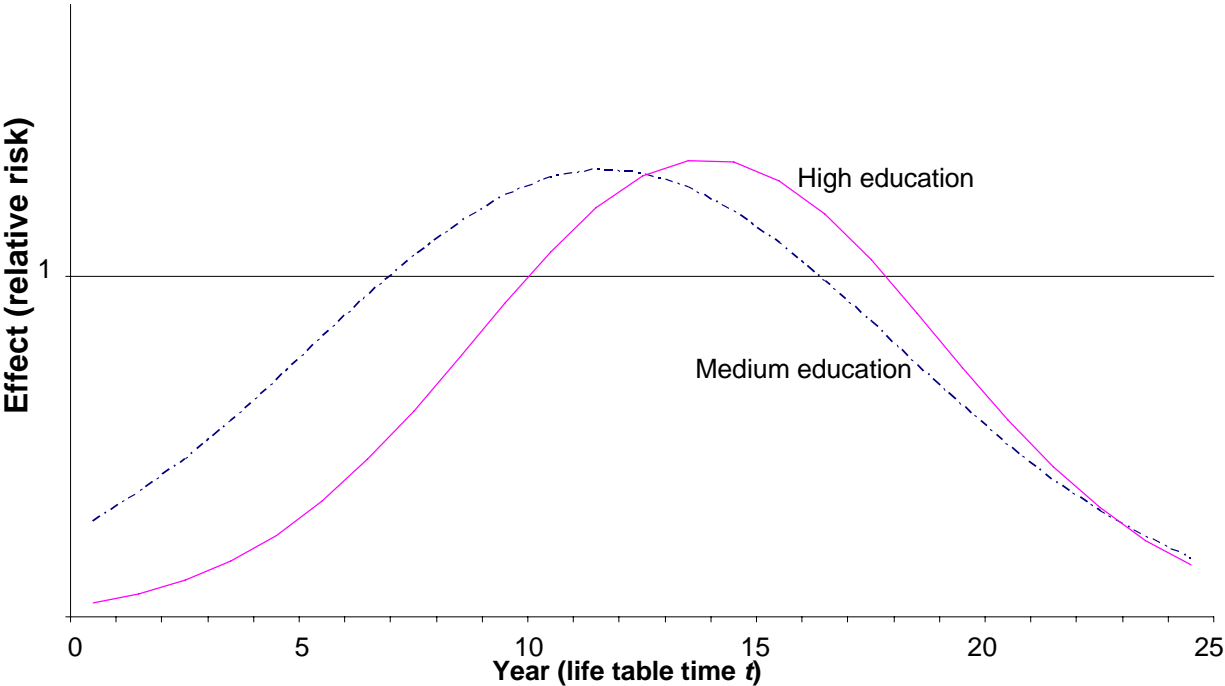
Table 6: Unadjusted and adjusted trends in the period TFR and its components
(pooled data analysis)

Transition	Measure	1988-92	1993-97	1998-02		
B-M	ρ_B	Unadjusted	0.87	0.88	0.92	
		Adjusted	0.84	0.87	0.92	
	Mean age at first marr.	Unadjusted	23.9	23.7	23.2	
		Adjusted	24.5	24.2	23.6	
	Median age at first marr.	Unadjusted	23.1	22.9	22.4	
		Adjusted	23.8	23.5	22.9	
	M-1	ρ_M	Unadjusted	0.96	0.96	0.94
			Adjusted	0.96	0.96	0.94
		Mean CBI	Unadjusted	1.31	1.33	1.43
Adjusted			1.30	1.33	1.43	
Median CBI		Unadjusted	0.97	0.98	1.07	
		Adjusted	0.96	0.98	1.07	
1-2		ρ_1	Unadjusted	0.89	0.86	0.83
			Adjusted	0.89	0.86	0.83
		Mean CBI	Unadjusted	2.88	3.00	3.11
	Adjusted		2.90	3.00	3.10	
	Median CBI	Unadjusted	2.32	2.42	2.52	
		Adjusted	2.34	2.42	2.51	
	2-3	ρ_2	Unadjusted	0.82	0.79	0.72
			Adjusted	0.80	0.78	0.71
		Mean CBI	Unadjusted	3.19	3.26	3.42
Adjusted			3.21	3.26	3.41	
Median CBI		Unadjusted	2.62	2.67	2.79	
		Adjusted	2.63	2.67	2.77	

3-4	p_3	Unadjusted	0.74	0.68	0.63	
		Adjusted	0.73	0.67	0.63	
	Mean CBI	Unadjusted	3.20	3.32	3.40	
		Adjusted	3.22	3.32	3.39	
	Median CBI	Unadjusted	2.61	2.69	2.74	
		Adjusted	2.62	2.68	2.73	
	4-5	p_4	Unadjusted	0.69	0.66	0.58
			Adjusted	0.67	0.64	0.58
Mean CBI		Unadjusted	3.16	3.20	3.30	
		Adjusted	3.16	3.21	3.29	
Median CBI		Unadjusted	2.70	2.74	2.81	
		Adjusted	2.70	2.73	2.80	
5-6		p_5	Unadjusted	0.72	0.65	0.60
			Adjusted	0.69	0.64	0.59
	Mean CBI	Unadjusted	3.12	3.21	3.28	
		Adjusted	3.11	3.18	3.24	
	Median CBI	Unadjusted	2.68	2.75	2.79	
		Adjusted	2.67	2.72	2.76	
	6+ - 7+	p_{6+}	Unadjusted	0.68	0.63	0.63
			Adjusted	0.67	0.63	0.62
Mean CBI		Unadjusted	3.02	3.08	3.09	
		Adjusted	3.00	3.06	3.06	
Median CBI		Unadjusted	2.57	2.62	2.62	
		Adjusted	2.56	2.60	2.61	
TFR		Unadjusted	3.64	3.26	2.94	
		Adjusted	3.39	3.14	2.88	
TMFR	Unadjusted	4.21	3.69	3.17		
	Adjusted	4.03	3.60	3.13		

Note: CBI denotes closed birth interval. Unadjusted estimates are based on CLL models in which the dummy variables representing life table time intervals and the dummy variables representing the three time periods are the only predictor variables. Adjusted estimates are based on CLL models in which the set of predictor variables additionally includes the dummy variables representing residence and education. In the calculation of adjusted estimates, the control variable is held constant at its interval-specific mean values calculated from the pooled data set; in other words, the same set of means is used throughout, regardless of time period.

Figure 1: The effect of medium education, $\exp(f+gt+ht^2)$, and the effect of high education, $\exp(j+kt+mt^2)$, relative to low education, on progression from birth to first marriage: period estimates based on the 5-year period before the 1993 Philippines DHS survey



Note: Effects of medium and high education are relative to low education. See equation (1) in the text.