European Population Conference 2006 - Session 901 – Data and Methods

The reconstruction of fertility trends with DHS birth histories
Application of Poisson regression to person-period data

Bruno SCHOUMAKER
Institut de démographie, Université catholique de Louvain, Belgium
schoumaker@demo.ucl.ac.be

Context and objective

The reconstitution of birth histories—or maternity histories—is a widely used approach for collecting data on fertility in developing countries. Since the 1970s with the World Fertility Survey (WFS), and even more so since the mid-1980s with the program of Demographic and Health Surveys (DHS), birth histories have become an indispensable source of data for studying fertility levels, trends, and determinants. The principle is well-known: a sample of women are asked about their reproductive history, and the birth dates of each of their children, from the first birth until the time of the survey, are recorded.

These birth histories are mainly used to calculate the classic indicators of fertility, in particular fertility rates and total fertility rates, and to reconstitute fertility trends over the past ten to fifteen years. When combined with socio-economic data collected by fertility surveys, they can also be used for explicative analyses of fertility behaviour. Finally, although these data are most often used to study recent fertility, occasionally researchers carry out explicative analyses that exploit their longitudinal nature.

In this paper, we present a simple method that can be used for a variety of analyses of fertility levels, trends and differentials. We then apply the method to the reconstruction of fertility trends in a variety of countries at different stages of fertility transition using DHS birth histories. This paper is an extension of the applications on fertility trends of the attached paper “A person period approach to analyzing birth histories” (Schoumaker, 2004).

Calculating fertility rates using person-period data and Poisson regression

The method relies on the organisation of the birth histories as a person period data file, which can be analyzed with Poisson regression (log rates models). This approach proves to be very flexible in various ways: not only does it simplify the programs for calculating classical fertility rates and TFRs as they are published in DHS reports, but it also makes it possible to include time-varying explanatory variables in regression models. The principle is the following: for each woman, the period over which rates are to be calculated (for example five years) is divided into several sub-periods (or segments) over the course of which the explanatory variables (the age groups for example) are constant. For example, instead of having a single observation for a woman aged 22 at the time of the survey, two sub-periods are distinguished over the course of which the age group is constant, and observations corresponding to each period are created in the data file. One observation covers the period from exact age 17 to exact age 20 (age group 15-19), and the second observation applies to the period from exact age 20 to exact age 22 (age group 20-24). The table below illustrates the organisation of the data into person-periods for two cases. The last column presents an individual fertility "rate" for the sub-period (r_{ij}), calculated as the ratio of the numberof births (n_{ij}) to the length of the sub-period (t_{ij}).

EXAMPLE OF DATA ORGANISED INTO PERSON-PERIODS

Woman number (j)	Sub-period number (i)	Exact age at time of survey	Age group belonged to during the sub-period	Births during the sub-period (y_{ij})	Duration of exposure in years (t_{ij})	Individual fertility rate for the sub-period (r_{ij})
1	1	22.0	15-19	1	3.0	0.33
1	2	22.0	20-24	1	2.0	0.50
2	1	24.5	15-19	0	0.5	0.00
2	2	24.5	20-24	2	4.5	0.44

Such data can be analyzed with Poisson regression to calculate fertility rates. The dependent variable is the number of births over the course of each sub-period, the independent variables are the five-year age groups (as dummy variables), and the length of each sub-period is controlled by the offset term. The exponentials of the explanatory variables (age groups in this example) are simply equal to age-specific fertility rates, and can be summed to calculate the total fertility rate. This approach leads to strictly identical results as those published in DHS reports.

Reconstructing fertility trends

The main advantage of this method is that it is possible to include time-varying explanatory variable in a very simple way. Again, the method consists of expanding each observation (line in the file) into new observations at each change in the value of the explanatory variables. A simple example of a time-changing variable is the period (year, five-year period, etc.), and one possible application of Poisson regression to person-period data is the reconstitution of fertility trends based on birth histories. In this case, the data are organised by calendar year, and sub-divided when a woman passes from one age group to another during the course of the year. This structure makes it possible to include both age groups and years as explanatory variables in the Poisson regression and to estimate annual variation in fertility levels.

In the following example, we hypothesise here that the fertility schedule is constant, that is, that the proportional distribution of age-specific rates is constant. As a result, the exponentiated regression coefficient for a given year measures the relationship between fertility (TFR) for this year and fertility in the reference year. The TFR for the reference year is calculated from the regression coefficients for the five year age groups, and then multiplied by the exponentiated regression coef ficient for the specified year to obtain an estimate of the TFR for that year.

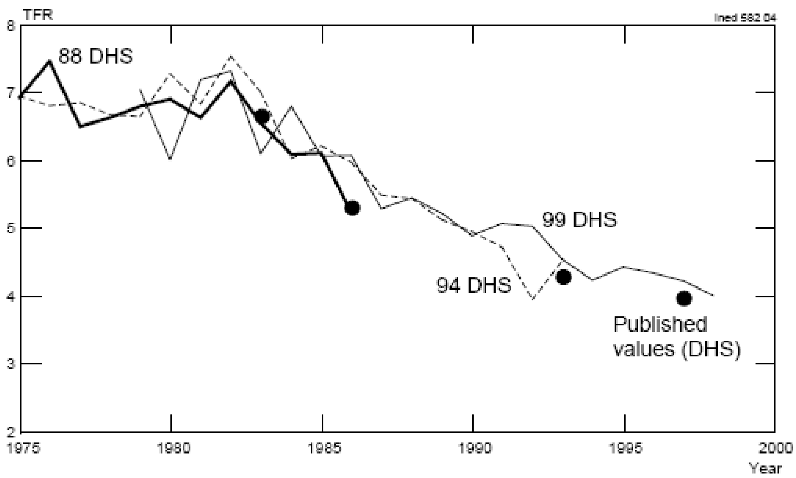
The Table below illustrates the application of this method to fertility data for the twenty years preceding the 1999 Zimbabwe DHS. The reference year is 1998, and the TFR for this year (4.00 children per woman) is obtained from the regression coefficients for the five-year age groups. The TFR for 1997 (4.22) is equal to the fertility of the reference year (4.00) multiplied by the exponentiated coefficient of the year 1997 (1.055), and similarly for other years. Change over time in fertility is depicted in the Figure below. The values estimated by applying the same method to birth histories from the 1988 and 1994 DHS are also included in this figure, in addition to the published values from the DHS reports for the different dates. Note that the levels and trends estimated from the three surveys match very well overall, and the retrospective estimates are also very close to the published values for different dates.

AGE-SPECIFIC FERTILITY IN 1998 AND RECONSTRUCTION OF FERTILITY TRENDS OVER THE TWENTY YEARS PRECEDING THE SURVEY IN ZIMBABWE. RESULTS FROM A POISSON REGRESSION ON PERSON-PERIOD DATA

Age groups	Regression coefficients (β)	exp (β)	Year	Regression coefficients (β)	exp (β)	Estimated TFR
15-19	– 2.399	0.091	1998 (*Ref.*)	–	–	4.00
20-24	– 1.704	0.182	1997	0.0524	1.055	4.22
25-29	– 1.759	0.172	1996	0.0796	1.082	4.33
30-34	– 1.868	0.154	…	…		…
35-39	– 2.098	0.123	1988	0.3075	1.360	5.44
40-44	– 2.777	0.062	…	…		…
45-49	– 4.121	0.016	1980	0.4053	1.500	6.03
1998 TFR (reference year)		4.00	1979	0.5637	1.757	7.03
Source: Zimbabwe DHS, 1999.						

-Reconstruction of the TFR during the period 1975-1998 in Zimbabwe. Results from a Poisson regression on person-period data

Source: 1988, 1994, and 1999 DHS.



This method has several advantages relative to the classic approach of calculating TFRs separately for different periods. First, because only a single regression model is required to reconstitute trends over fifteen or twenty years, the method is easier to implement. Second, the results are interpretable in terms of total fertility rates between the ages of 15 and 49, while with the classic approach TFRs are estimated only to age 35 or 40. A third benefit is that fertility trends can be integrated into the regression model itself. Rather than treating years as dummy variables in the model, it is possible to include a function of time (linear, quadratic, spline, etc.) in the regression. Finally, this method allows the user to include explanatory variables in the model and estimate the effect of these variables on annual fertility levels. For example, the effect of changes in the socio-economic characteristics of the population on fertility could be evaluated by including individual variables that are fixed in time; similarly, time-varying variables at the individual, contextual (such as the presence of family planning services in the village) or global (per capita GDP, etc.) level can be incorporated to explain changes over time in fertility.

Applications and extensions of the method

In this paper we propose to use and extend the method described above in three directions:

- The method will be applied to reconstructing fertility trends, using DHS birth histories, in about 20 countries at different stages of fertility transition and from different regions of the world (Burkina Faso, Bangladesh, Cambodia, Egypt, Kenya, Madagascar, Morocco, Peru…). When several surveys are available in a country, reconstructed fertility trends from different surveys in the same country will be compared to evaluate their consistency and the possible sources of inconsistencies ;
- One possible source of bias in fertility levels and trends estimated by this method is the assumption that the proportional distribution of age-specific rates (fertility schedule) is constant over time[1]. This assumption will be tested and relaxed in the applications if this proves necessary. This will be done through modelling the relationship between age and fertility by a parsimonious function of age, and letting the parameters of the function to vary over time ;
- In the example described above, the trend in the total fertility rate is estimated in a non-parametric way, i.e. by including dummy variables for each year. We will test more parsimonious ways to model fertility trends by using spline functions.

[1] If the fertility schedule is not constant, and fertility decreases proportionally more at certain ages than at others—which is currently the case—the estimation of fertility levels and trends could be affected. In the case of Zimbabwe, our method slightly overestimates fertility over the recent period compared to the published TFR, which could be explained by the fact that fertility has decreased more at older ages. However, in the cas of Zimbabwe, this distortion has only a slight impact on the estimate of the general trend.

A Person-Period Approach to Analysing Birth Histories

Bruno SCHOUMAKER*

The reconstitution of birth histories — or maternity histories — is a widely used approach for collecting data on fertility in developing countries. Since the 1970s with the World Fertility Survey (WFS), and even more so since the mid-1980s with the program of Demographic and Health Surveys (DHS), birth histories have become an indispensable source of data for studying fertility levels, trends, and determinants. The principle is well-known: a sample of women are asked about their reproductive history, and the birth dates of each of their children, from the first birth until the time of the survey, are recorded. These birth histories are mainly used to calculate the classic indicators of fertility, in particular fertility rates and total fertility rates, and to reconstitute fertility trends over the past ten to fifteen years (Garenne and Joseph, 2002; Potter, 1977). When combined with socio-economic data collected by fertility surveys, they can also be used for explicative analyses of fertility behaviour (Cleland and Rodriguez, 1988; White et al., 2002). Finally, although these data are most often used to study recent fertility, occasionally researchers carry out explicative analyses that exploit their longitudinal nature (Angeles et al., 1998; Raftery et al., 1995).

Birth histories constitute the primary material for the majority of studies on fertility in developing countries. Methods of analysis differ substantially depending on the type of study (descriptive or explanatory). The age-specific rates and TFRs presented in survey reports and estimates of fertility trends are calculated using classic methods of demographic analysis. Explanatory studies use regression methods: logistic regression (Angeles et al., 1998), Poisson regression (Mencarini, 1999), or event history methods (Raftery et al., 1995). Although the principle is rarely described in manuals of demographic analysis, regression methods, particularly Poisson regression, can also be used to calculate classic demographic measures such as fertility rates (Powers and Xie, 2000) and total fertility rates. When using regression, it becomes possible to include variables that explain fertility as well — in fact, this is the usual reason for using regression models.

* Catholic University of Louvain, Belgium.
Translated by Sarah R. Hayford.

In our opinion, using the same method for descriptive and explanatory analyses not only simplifies the analysis from a technical point of view but also facilitates the interpretation of regression coefficients. The *first objective* of this note will be to summarize how the same method — Poisson regression — can be used both to calculate classic fertility measures and to incorporate explanatory variables whose effects are easily interpretable from a demographic point of view. This first objective is primarily didactic, since the use of Poisson regression for the analysis of fertility data has already been described by various authors (Rodriguez and Cleland, 1988; Trussell and Rodriguez, 1990; Winkelmann and Zimmermann, 1994). The *second objective* of this note, and its original contribution, is to show how reorganising birth history data into person-periods and using Poisson regression to analyse them offers more flexibility in calculating fertility measures and measuring fertility trends and determinants. This method allows researchers to calculate TFRs such as the ones reported in the DHS survey reports in a simple way, to reconstitute past trends in fertility, and to carry out explanatory analyses including time-changing variables. In sum, this note has a practical rather than theoretical bent; its goal is to propose a method that, we believe, simplifies the analysis of birth history data.

We begin with a brief review of the two approaches generally used to calculate birth rates and total fertility rates from retrospective survey data. We then proceed to introduce two ways of using Poisson regression to analyse fertility data. The standard approach relies on individual observations; we apply this method to data from the 1998-99 Burkina Faso DHS to estimate fertility rates and TFRs via Poisson regression. We next present the second approach, which is based on the reorganisation of data into person-periods. We first apply this method to data from two DHS surveys (Zimbabwe 1999 and Burkina Faso 1998-99) to calculate fertility rates and TFRs in these two countries. We next use it to estimate differences in fertility according to socio-economic status in Burkina Faso. Finally, we apply this method to fertility trends in Zimbabwe in order to illustrate how organising the data into person-periods allows for the incorporation of time-changing variables.

I. Calculating fertility rates and TFR

In general, age-specific fertility rates are calculated by dividing the number of births to women in a given age group by the number of years lived by these women in the period under study (Vandeschrick, 1995). These rates can be calculated in two ways using retrospective survey data. The simplest approach (here called method 1) calculates rates by dividing the total number of births over the past five years reported by women in a five-year age group at the time of the survey by the number of years lived by these women over the course of the preceding five years (number of women multiplied by five). In practice, this method is equivalent to calculating the average of an individual variable (the number of births over the past five years divided by the number of years) for each five-year age group. These rates are added and multiplied by five to obtain the total fertility rate.

A second approach (here called method 2) consists of calculating age-specific fertility rates by dividing the total number of births between two exact ages by the total number of years lived by women over the course of the period. This is the method used in the DHS reports. In contrast to those produced by method 1, rates calculated in this manner describe five-year age groups based on women's age at the time of the birth of their child, not at the time of the survey. Figure 1 illustrates these two approaches on a Lexis diagram. One advantage of the second method is that the period for which fertility rates and TFRs are calculated does not have to be a multiple of five(1). In fact, the reference period is often three or four years in the DHS reports, which limits the bias in some of these surveys(2). However, the calculations are less straightforward than with the first method, since they require that each birth be placed in the "right square" of the Lexis diagram and that each woman's time be partitioned into age groups. In practice — for instance in the programmes provided by Macro International(3) — the analyst usually produces two tables, one establishing the number of births in each square of the Lexis diagram and the second totalling the number of years lived by the women in each age group. Rates are

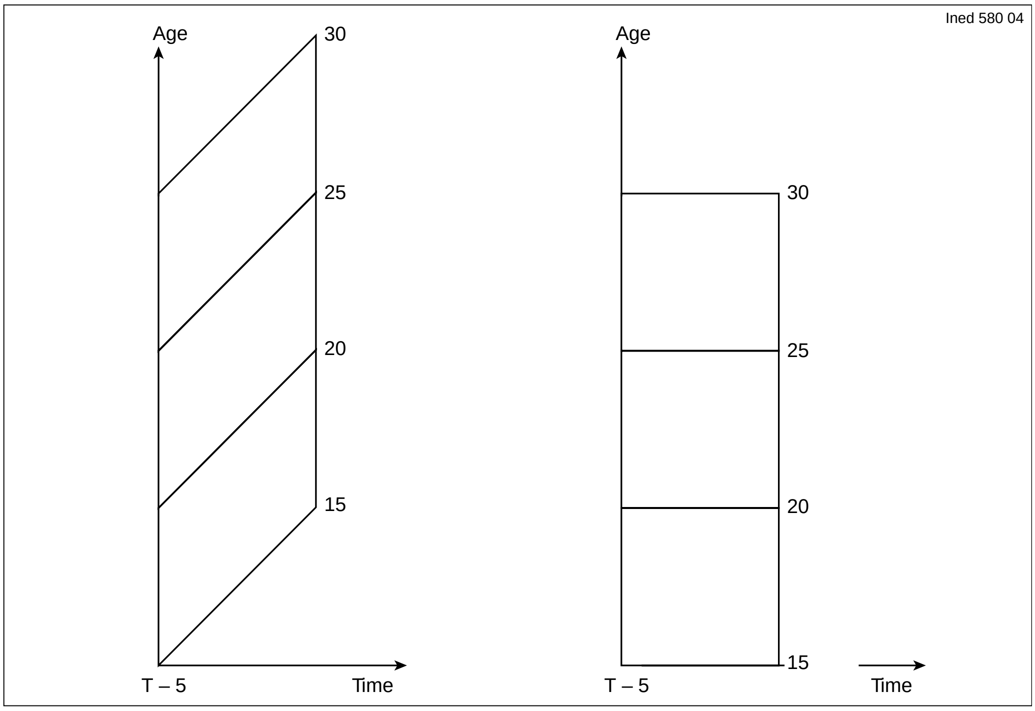


Figure 1.–Lexis diagrams comparing two approaches to calculating fertility rates for five-year age groups

(1) With the first approach, it is not strictly necessary to calculate rates for a five-year period, but using other lengths of time requires basing the rates on non-standard ages.

(2) In effect, the ages of children who are at the limit of the threshold of eligibility for the health modules are artificially inflated (Marckwardt and Rutstein, 1996). If these modules apply to children under five years old, TFRs calculated over five years will be underestimated.

(3) See http://www.measuredhs.com/zip/frsas.zip for SAS programs provided by Macro International.

then calculated by dividing the number of births by the number of person-years lived in each age group.

II. Calculating fertility rates using Poisson regression: Method 1

The first way of calculating fertility rates has a practical advantage, since rates are obtained in a single step by averaging an individual variable (the number of births divided by the length of the period) over each age group. These same rates can also be estimated via Poisson regression (see inset). The advantage of Poisson regression is that it allows the incorporation of explanatory variables, whose effects are expressed in the form of ratios of rates (Powers and Xie, 2000) and whose significance can be tested(4).

We use data from the 1998-99 Burkina Faso DHS to illustrate the estimation of fertility rates and TFRs using individual data and Poisson regression (Table 1). The dependent variable is the number of births over the five years preceding the survey (variable predefined in the DHS files); we include five-year age groups (as dummy variables) on the right hand side of the model and control for the length of exposure (five years for each woman) using a term called the offset. Fertility rates are obtained by exponentiating the regression coefficients for each of the seven age groups(5), and the total fertility rate is equal to the sum of the rates multiplied by five. Table 1 presents the regression coefficients, fertility rates, and TFR (6.73 children per woman) (column 1). It also shows results from two separate regressions on women living in urban areas (column 2) and rural areas (column 3). The TFR is 4.04 children per woman in urban areas and 7.23 children per woman in rural areas, for a TFR that is 1.79 times higher in rural areas.

Using Poisson regression, it is also possible to estimate the effects of explanatory variables in the form of rate ratios (for recent applications, see, e.g., Gregson et al., 1997; White et al., 2002). Rather than estimating models separately by area of residence, we could estimate the effect of area of residence by introducing this variable into the regression. The results are displayed in column 4. Urban residence is taken as the reference category, and the TFR of women living in urban areas (4.07) is obtained based on the regression coefficients for each age group. The exponentiated regression coefficient for area of residence represents the ratio of the fertility rate of rural women to the rate of urban women (1.78). By hypothesis, the multiplicative effect is the same at every age (identical age schedule), so the TFR of rural women (7.23) is obtained by multiplying the TFR of women living in urban areas by this value. The TFRs estimated separately and by including area of residence in the regression are very close, and the multiplicative effect of area of residence obtained using Poisson regression (1.78) is practically identical to the ratio of the rural to urban TFRs estimated separately (1.79). In summary, introducing an explanatory variable into the regression gives very similar results to those obtained by estimating separate models.

(4) Poisson regressions can be estimated with the SAS and Stata software.
(5) We treat the constant as omitted from the model.

TABLE 1.– ESTIMATION OF FERTILITY RATES AND TFR OVER THE FIVE YEARS PRECEDING THE SURVEY IN BURKINA FASO.
RESULTS OF POISSON REGRESSIONS ON INDIVIDUAL DATA

Explanatory variables	Model 1		Model 2		Model 3		Model 4	
	Total		Urban areas		Rural areas		Total with area of residence as explanatory variable	
	Coefficients (β)	exp (β)	Coefficients (β)	exp (β)	Coefficients (β)	exp (β)	Coefficients (β)	exp (β)
Age group								
15-19	– 3.066	0.047	– 3.758	0.023	– 2.938	0.053	– 3.543	0.029
20-24	– 1.386	0.250	– 1.917	0.147	– 1.288	0.276	– 1.870	0.154
25-29	– 1.214	0.297	– 1.596	0.203	– 1.158	0.314	– 1.719	0.179
30-34	– 1.290	0.275	– 1.724	0.178	– 1.228	0.293	– 1.796	0.166
35-39	– 1.383	0.251	– 1.874	0.153	– 1.325	0.266	– 1.898	0.150
40-44	– 1.825	0.161	– 2.573	0.076	– 1.740	0.176	– 2.335	0.097
45-49	– 2.741	0.064	– 3.638	0.026	– 2.671	0.069	– 3.267	0.038
Area of residence								
Urban	–	–	–	–	–	–	0.000	1.000
Rural	–	–	–	–	–	–	0.575	1.777
Urban TFR				4.04				4.07
Rural TFR						7.23		7.23
Total TFR		6.73						

Source: Burkina Faso DHS, 1998-1999.

Poisson regression

Poisson regression is used to analyse non-negative whole number variables (count data) such as the number of births occurring to women over the course of a given period. It is a particular case of the generalised linear model, in which the conditional distribution of the dependent variable follows a Poisson law and the link function is logarithmic (Winkelmann et al., 1994; Trussell and Rodriguez, 1990; Cameron et al., 1998). It presents several advantages for the statistical analysis of fertility. Notably, it makes it possible to control the length of exposure in the models via a term called the offset. The offset is an independent variable whose coefficient is fixed at one (Trussell and Rodriguez, 1990); including it is equivalent to assuming that the risk is proportional to the duration. Poisson regression estimates the effects of explanatory variables on rates; the logarithmic form of the model is such that the exponents of the regression coefficients represent the relationships between the fertility rates of different groups of women.

In the case of fertility, the dependent variable is the number of births (y_i) occurring to women (i) over the course of a given period, and the probability that the random variable Y_i is equal to y_i (number of births observed) is assumed to follow a Poisson distribution with average μ_i:

$$P(Y_i = y_i | \mu_i) = \frac{e^{\mu_i}\mu_i^{y_i}}{y_i!} \quad [1]$$

The average μ_i, the average number of births per period, can be decomposed into the product of a fertility rate (λ_i) and a length of exposure (t_i):

$$\mu_i = t_i\lambda_i \quad [2]$$

The logarithm of the average (μ_i) is thus equal to the sum of the logarithms of the length of exposure (t_i) and the fertility rate (λ_i):

$$\ln\mu_i = \ln t_i + \ln\lambda_i \quad [3]$$

The logarithm of the length of exposure is the offset, and the logarithm of the fertility rates (λ_i) is modelled as a linear function of k explanatory variables:

$$\ln\lambda_i = \sum_{k=1}^{K} \beta_k x_{ki} \quad [4]$$

From which:

$$\ln\mu_i = \ln t_i + \sum_{k=1}^{K} \beta_k x_{ki} \quad [5]$$

By exponentiating the above equation, we see that the explanatory variables have multiplicative effects on the rate (λ_i), since:

$$\lambda_i = \exp\sum_{k=1}^{K} \beta_k x_{ki} = \prod_{k=1}^{K} \exp(\beta_k x_{ki}) \quad [6]$$

The exponent of the regression coefficient (β_k) for an explanatory variable (x_k) thus expresses the relationship between the fertility rate of women for which the explanatory variable has a given value and the fertility rate of women for which the variable has that value minus one, all other things being equal. For example, for a dichotomous variable, the exponent of the coefficient of this variable is equal to the ratio of the fertility rate of women in a category to the fertility rate of women in the reference category. Several of the examples covered in this note illustrate the interpretation of the regression coefficients.

III. Calculating fertility rates using Poisson regression: Method 2

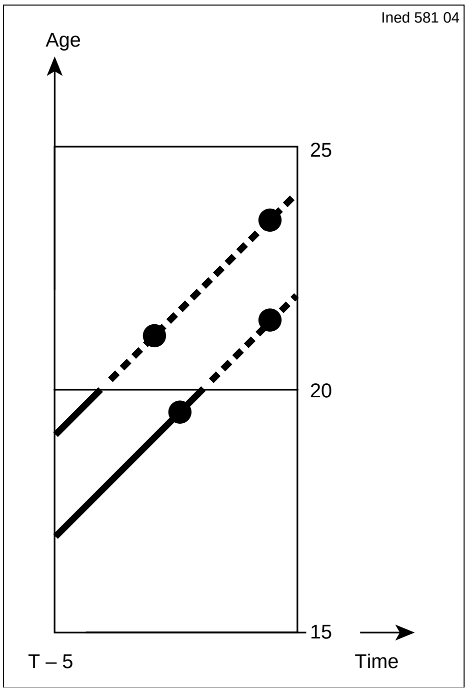


Figure 2.–Lexis diagram illustrating how data are divided into person-periods to calculate fertility rates for five-year age groups

Recall that with method 1, age groups are defined as a function of the age of women *at the time of the survey.* With method 2, on the other hand, rates and TFR are calculated as a function of the age of the mothers *at the time of the birth of the children*, and a woman can pass through two age groups over the course of a period. For example, consider a woman who is exact age 22 at the time of the survey and who has two children over the course of the five years preceding the survey, one at age 19 and one at age 21 (Figure 2, lower line). This woman spent three years in the 15-19 age group (continuous line) and two years in the 20-24 age group (dotted line), and it is thus not possible to assign her a single value for the age group variable. Using method 2 based on individual observations, it is not possible to calculate fertility rates or TFR either as simple averages or by including age groups in the regression.

The solution proposed here consists of moving from a file of individual observations to a person-period file. This approach proves to be very flexible in various ways: not only does it simplify the programs for calculating fertility rates and TFRs by method 2, but it also makes it possible to include time-varying variables in the explanatory models. The principle is the following: for each woman, the period over which rates are to be calculated (for example five years) is divided into several sub-periods (or segments) over the course of which the explanatory variables (the age groups in this example) are constant. Thus, instead of having a single observation for the woman aged 22 at the time of the survey (Figure 2), two sub-periods are distinguished over the course of which the age group is constant, and observations corresponding to each period are created in the data file. One observation covers the period from exact age 17 to exact age 20 (Figure 2, lower continuous line), and the second observation applies to the period from exact age 20 to exact age 22 (dotted line). The dependent variable is the number of births, here one in each segment. The length of each segment is also included in the file, here three years between ages 17 and 20 and two years between ages 20 and 22. Table 2 illustrates the organisation of the data into person-periods for the two cases represented in Figure 2. The last column presents an individual fertility “rate” for the sub-period (r_{ij}), calculated as the ratio of the number of births (n_{ij}) to the length of the sub-period (t_{ij}).

TABLE 2.– EXAMPLE OF DATA ORGANISED INTO PERSON-PERIODS
(DATA ILLUSTRATED IN FIGURE 2)

Woman number (j)	Sub-period number (i)	Exact age at time of survey	Age group belonged to during the sub-period	Births during the sub-period (y_{ij})	Duration of exposure in years (t_{ij})	Individual fertility rate for the sub-period (r_{ij})
1	1	22.0	15-19	1	3.0	0.33
1	2	22.0	20-24	1	2.0	0.50
2	1	24.5	15-19	0	0.5	0.00
2	2	24.5	20-24	2	4.5	0.44

There are several ways of calculating age-specific fertility rates from person-period data (method 2). One possibility is to calculate for each age group the average of the individual fertility rates by sub-period (r_{ij}) weighted by the lengths of the sub-periods (t_{ij}). When the data are reorganised into person-periods, it is possible to obtain fertility rates and TFRs identical to those in the DHS reports from simple weighted averages rather than producing separate tables. The second possibility consists of applying a Poisson regression to the person-period data. The dependent variable is the number of births over the course of each sub-period, the independent variables are the five-year age groups (as dummy variables), and the length of each sub-period is controlled by the offset term. The primary difference relative to the application of Poisson regression to individual data is that the independent variables (the age groups) are time-varying and the time spent in each age group is controlled by the offset term.

Before passing to some applied examples, we emphasise that the application of Poisson regression to person-year data is not novel in itself. The same principle is used in duration models (Blossfeld and Rohwer, 2002; Courgeau and Lelièvre, 1992) and models used to analyse repeated events, in epidemiology for instance (Clayton, 1994). The originality of the approach proposed here is primarily in the application of this principle to the calculation of classic measures of fertility (rates, TFR) and the measure of fertility trends and determinants based on birth histories. To our knowledge, this application has not been presented before, except by this author in a different form (Schoumaker, 2001).

We illustrate this approach using several examples. Table 3 compares rates and TFRs obtained using Poisson regression on person-period data and results published in the DHS reports for the 1998-99 Burkina Faso survey (rates calculated over the past five years) and the 1999 Zimbabwe survey (past three years). For each country, the first column contains the regression coefficients estimated for the seven age groups and the second column presents the exponentiated coefficients, that is, the fertility rates. The fertility rates published in the DHS reports, printed in the third column, are exactly identical to the rates estimated using regression

As with the first method, explanatory variables can be incorporated in the model, and the statistical significance of their coefficients can be tested. Table 4 presents the results of a Poisson regression using data from the 1998-99 Burkina Faso DHS and including a variable measuring standard of living(6) in addition to the age groups. For each of the five categories of standard of living, the TFR is obtained by multiplying the TFR in the reference category (the poorest women) by the exponentiated regression coefficient. The asterisks indicate the extent to which the TFRs are significantly different from that of the reference category. We note here that the fertility of very poor women is slightly higher (significant difference) than that of the poorest women, and that the fertility of the most well-off women is significantly lower. The other differences are not significant. This approach provides a simple way of testing the significance of differences in fertility for different groups of women.

TABLE 3.– AGE-SPECIFIC FERTILITY AND TFR IN ZIMBABWE AND IN BURKINA FASO. COMPARISON OF RESULTS FROM POISSON REGRESSION ON PERSON-PERIOD DATA AND PUBLISHED RESULTS FROM THE DHS REPORTS

Age group	Zimbabwe 1999 (past 3 years)			Burkina Faso 1998-1999 (past 5 years)		
	Coefficients (β)	exp (β)	Published results	Coefficients (β)	exp (β)	Published results
15-19	– 2.193	0.112	0.112	– 1.940	0.144	0.144
20-24	– 1.613	0.199	0.199	– 1.186	0.305	0.305
25-29	– 1.717	0.180	0.180	– 1.226	0.293	0.293
30-34	– 2.005	0.135	0.135	– 1.332	0.264	0.264
35-39	– 2.228	0.108	0.108	– 1.543	0.214	0.214
40-44	– 3.088	0.046	0.046	– 2.190	0.112	0.112
45-49	– 4.233	0.015	0.015	– 3.563	0.028	0.028
TFR	–	3.97	3.97		6.80	6.80

Sources for published data: Central Statistical Office (2000) for Zimbabwe; INSD (2000) for Burkina Faso.

TABLEAU 4.– ESTIMATION OF THE RELATIONSHIP BETWEEN STANDARD OF LIVING AND FERTILITY IN BURKINA FASO. RESULTS OF A POISSON REGRESSION USING PERSON-PERIOD DATA FOR THE FIVE YEARS BEFORE THE SURVEY

Standard of living	TFR
Lowest (*Ref.*)	6.79
Very low	7.50**
Low	7.16
Average	6.73
Well-off	4.36***

*** $p < 0.01$; ** $p < 0.05$; ** $p < 0.10$.
Source: Burkina Faso DHS, 1998-1999.

(6) The standard of living variable is based on household possessions.

We emphasize that working with person-period data does not artificially inflate the size of the sample when Poisson regression is used to analyse the data. Just as Poisson regression can be used on grouped or individual data with equivalent results (Rodriguez, 2001), dividing individual observations into several data points does not change results (regression coefficients or standard errors) when the data are analysed using this method[7].

IV. Reconstructing fertility trends

As noted above, the reorganisation of data into person-periods can be extended to incorporate time-varying explanatory variables. The method consists of expanding each observation (line in the file) into new observations at each change in the value of the explanatory variables. A simple example of a time-changing variable is the period (year, five-year period, etc.), and one possible application of Poisson regression to person-period data is the reconstitution of fertility trends based on birth histories. In this case, the data are organised by calendar year, and sub-divided when a woman passes from one age group to another during the course of the year. This structure makes it possible to include both age groups and years as explanatory variables in the Poisson regression and to estimate annual variation in fertility levels. As in the case of fixed explanatory variables, we hypothesise here that the fertility schedule is constant, that is, that the proportional distribution of age-specific rates is constant. The exponentiated regression coefficient for a given year measures the relationship between fertility (TFR) for this year and fertility in the reference year. The TFR for the reference year is calculated from the regression coefficients for the five-year age groups, and then multiplied by the exponentiated regression coefficient for the specified year to obtain an estimate of the TFR for that year.

Table 5 illustrates the application of this method to fertility data for the twenty years preceding the 1999 Zimbabwe DHS. The reference year is 1998, and the TFR for this year (4.00 children per woman) is obtained from the regression coefficients for the five-year age groups. The TFR for 1997 (4.22) is equal to the fertility of the reference year (4.00) multiplied by the exponentiated coefficient of the year 1997 (1.055), and similarly for other years. Change over time in fertility is depicted in Figure 3. The values estimated by applying the same method to birth histories from the 1988 and 1994 DHS are also included in this figure, in addition to the published values from the DHS reports for the different dates. Note that the levels and

---

[7] As in any other analysis, we could correct the standard errors, notably to take into account the effects of clustering linked to the complex sampling frame (Lee et al., 1989). It is also possible to correct the biased standard errors that result when the hypothesis of equidispersion required for Poisson regression (variance of the dependent variable equal to its mean) does not hold. Over-dispersion (variance higher than the mean) leads to underestimation of the standard errors of the regression coefficients, while under-dispersion (the opposite case) leads to overestimation of standard errors (Winkelmann and Zimmermann, 1994). The negative binomial model is a common approach to account for over-dispersion. However, fertility data are more likely to be under- than over-dispersed, (Covas and Santos Silva, 2000; Winkelmann and Zimmermann, 1994) and the negative binomial model cannot be applied in the case of under-dispersion (Winkelmann and Zimmermann, 1994). Under-dispersed data can be analysed using generalized event count models (Winkelmann and Zimmermann, 1994; King, 1989). There are also simple ways to correct coefficient standard errors in the case of under- or over-dispersion (Allison, 1999).

trends estimated from the three surveys match very well overall, and the retrospective estimates are also very close to the published values for different dates[8].

TABLE 5.– AGE-SPECIFIC FERTILITY IN 1998 AND RECONSTRUCTION OF FERTILITY TRENDS OVER THE TWENTY YEARS PRECEDING THE SURVEY IN ZIMBABWE. RESULTS FROM A POISSON REGRESSION ON PERSON-PERIOD DATA

Age groups	Regression coefficients (β)	exp (β)	Year	Regression coefficients (β)	exp (β)	Estimated TFR
15-19	– 2.399	0.091	1998 (*Ref.*)	–	–	4.00
20-24	– 1.704	0.182	1997	0.0524	1.055	4.22
25-29	– 1.759	0.172	1996	0.0796	1.082	4.33
30-34	– 1.868	0.154	…	…		…
35-39	– 2.098	0.123	1988	0.3075	1.360	5.44
40-44	– 2.777	0.062	…	…		…
45-49	– 4.121	0.016	1980	0.4053	1.500	6.03
1998 TFR (reference year)		4.00	1979	0.5637	1.757	7.03

Source: Zimbabwe DHS, 1999.

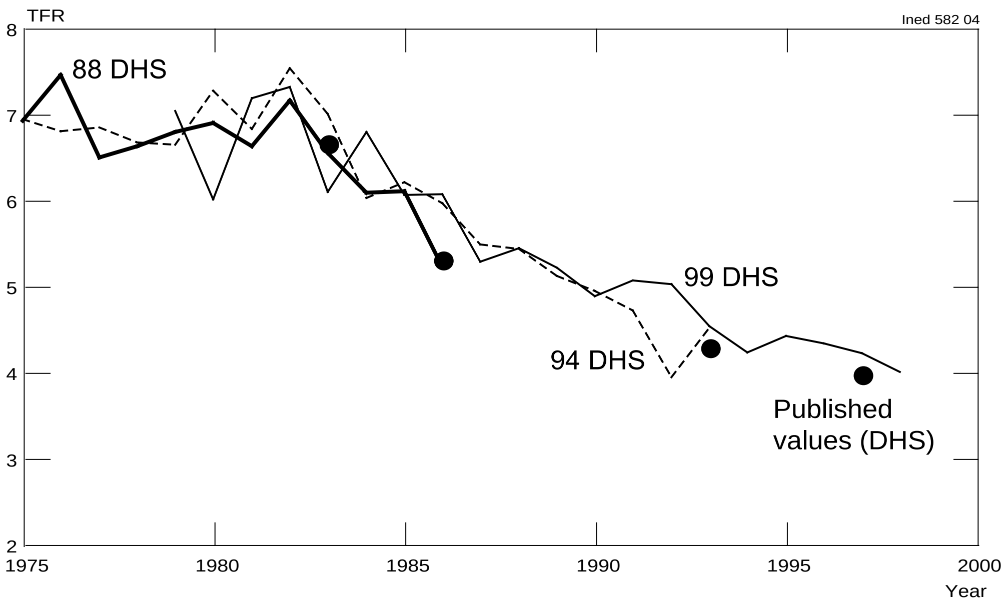


Figure 3.–Reconstruction of the TFR during the period 1975-1998 in Zimbabwe. Results from a Poisson regression on person-period data

Source: 1988, 1994, and 1999 DHS.

[8] If the fertility schedule is not constant, and fertility decreases proportionally more at certain ages than at others—which is currently the case—the estimation of fertility levels and trends could be affected. In the case of Zimbabwe, our method slightly overestimates fertility over the recent period compared to the published TFR, which could be explained by the fact that fertility has decreased more at older ages. However, this distortion has only a slight impact on the estimate of the general trend.

This method has several advantages relative to the classic approach of calculating TFRs separately for different periods (see for example Garenne and Joseph, 2002). First, because only a single regression model is required to reconstitute trends over fifteen or twenty years, the method is easier to implement. Second, the results are interpretable in terms of total fertility rates between the ages of 15 and 49, while with the classic approach TFRs are estimated only to age 35 or 40(9). A third benefit is that fertility trends can be integrated into the regression model itself. Rather than treating years as dummy variables in the model, it is possible to include a function of time (linear, quadratic, spline, etc.) in the regression. Finally, this method allows the user to include explanatory variables in the model and estimate the effect of these variables on annual fertility levels. For example, the effect of changes in the socio-economic characteristics of the population on fertility could be evaluated by including individual variables that are fixed in time; similarly, time-varying variables at the individual, contextual (such as the presence of family planning services in the village) or global (per capita GDP, etc.) level can be incorporated to explain changes over time in fertility.

Conclusion

We have demonstrated how organising birth history data into person-periods and applying Poisson regression to these data can provide a flexible approach for the analysis of levels, trends, and determinants of fertility. It is a (more) simple way of calculating age-specific fertility rates and total fertility rates, but also makes it possible to estimate explanatory models and reconstruct fertility trends. In sum, it allows the researcher to carry out descriptive and explanatory analyses of fertility using a common approach, with the same method and the same data file.

The examples addressed in this note are simple and could be complicated in several ways. It is of course possible to include additional explanatory variables in the models, in particular time-changing variables. Other extensions can also be envisioned. For example, the model of legitimate fertility proposed by Rodriguez and Cleland (1988) could easily be estimated using person-period data. Multi-level analyses on such data are also possible, since multi-level Poisson models can be estimated with various software. Finally, the combination of temporal and spatial dimensions in these models could also be useful in a study of the processes of diffusion of changes in fertility.

(9) This stems from the fact that the birth histories collected in these surveys apply to women aged 15 to 49 at the time of the survey, and therefore do not apply to certain age groups for older periods. It is therefore impossible to calculate a TFR between ages 15 and 49 using the classic approaches when we go back in time, except by estimating the rates at older ages separately for the earlier periods. With Poisson regression, these rates are estimated directly in the model, under the hypothesis that the fertility schedule is constant over time.

REFERENCES

ALLISON P., 1999, *Logistic Regression Using the SAS System*, Cary, SAS Institute Inc.

ANGELES G., D. GUILKEY, T. MROZ, 1998, "Purposive program placement and the estimation of family planning program effects in Tanzania", *Journal of the American Statistical Association*, 93, pp. 884-899.

BLOSSFELD H.-P., G. ROHWER, 2002, *Techniques of Event-History Modeling. New Approaches to Causal Analysis*, Mahwah (NJ), Lawrence Erlbaum Associates.

CAMERON A., P. TRIVEDI, 1998, *Regression Analysis of Count Data*, Cambridge, Cambridge University Press.

CENTRAL STATISTICAL OFFICE, MACRO INTERNATIONAL, 2000, *Zimbabwe Demographic and Health Survey 1999*, Harare/Calverton, Central Statistical Office/Macro International.

CLAYTON D., 1994, "Some approaches to the analyses of recurrent event data", *Statistical Methods in Medical Research*, 3, pp. 244-262.

CLELAND J., G. RODRIGUEZ, 1988, "The effect of parental education on marital fertility in developing countries", *Population Studies*, 42(3), pp. 419-442.

COURGEAU D., É. LELIÈVRE, 1992, *Event History Analysis in Demography*, Oxford, Clarendon Press.

COVAS F., J. SANTOS SILVA, 2000, "A modified hurdle model for completed fertility", *Journal of Population Economics*, 13(2), pp. 173-188.

GARENNE M., V. JOSEPH, 2002, "The timing of the fertility transition in sub-Saharan Africa", *World Development*, 30(10), pp. 1835-1843.

GREGSON S., T. ZHUWAU, R. ANDERSON, S. CHANDIWANA, 1997,"HIV and fertility change in rural Zimbabwe", *Health Transition Review*, 7(suppl. 2), pp. 89-112.

INSD, MACRO INTERNATIONAL, 2000, *Enquête démographique et de santé, Burkina Faso 1998-1999*, Calverton, Macro International.

KING G., 1989, "Variance specification in event count models: from restrictive assumptions to a generalized estimator", *American Journal of Political Science*, 33(3), pp. 762-784.

LEE E. S., R. FORTHOFER, R. LORIMER, 1989, *Analyzing Complex Survey Data*, Newbury Park, Sage Publications.

MARCKWARDT A., O. RUTSTEIN, 1996, "Accuracy of DHS-II demographic data: gains and losses in comparison with earlier surveys", *DHS Working Paper 19*, Calverton, Macro International.

MENCARINI L., 1999, "An analysis of fertility and infant mortality in South Africa based on 1993 LSDS data", *in* UEPA (ed.),Durban, UEPA, pp. 109-128.

POTTER J., 1977, "Problems in using birth-history analysis to estimate trends in fertility", *Population Studies*, 31(2), pp. 335-364.

POWERS D., Y. XIE, 2000, *Statistical Methods for Categorical Data Analysis*, San Diego, Academic Press.

RAFTERY A., S. LEWIS, A. AGHAJANIAN, 1995, "Demand or ideation? Evidence from the Iranian marital fertility decline", *Demography*, 32(2), pp. 159-182.

RODRIGUEZ G., 2001, *Generalized Linear Models. Lecture Notes*, Princeton, Princeton University.

RODRIGUEZ G., J. CLELAND, 1988,"Modelling marital fertility by age and duration : an empirical appraisal of the Page model", *Population Studies*, 42(2), pp. 241-257.

SCHOUMAKER B., 2001, "Une analyse multi-niveaux dynamique de la fécondité légitime au Maroc rural", *IUSSP General Population Conference*, Salvador de Bahia (Brasil).

TRUSSELL J., G. RODRIGUEZ, 1990, "Heterogeneity in demographic research", in J. Adams, D. Lam, A. Hermalin, P. Smouse (eds.), *Convergent Issues in Genetics and Demography*, New York, Oxford University Press, pp. 111-132.

VANDESCHRICK C., 1995, *Analyse démographique*, Louvain-la-Neuve, Academia/L'Harmattan.

WHITE M., E. TAGOE, C. STIFF, K. ADAZU, D. SMITH, 2002, *Urbanization and the Fertility Transition in Ghana*, Providence, Brown University.

WINKELMANN R., K. ZIMMERMANN, 1994, "Count data models for demographic data", *Mathematical Population Studies*, 4(3), pp. 205-221.

Bruno SCHOUMAKER, Institut de démographie, Université catholique de Louvain,
1/17 Place Montesquieu, B-1348, Louvain-la-Neuve, Belgique, Tel: 32 10 474136,
e-mail: schoumaker@demo.ucl.ac.be