



# DEMOGRAPHIC RESEARCH

*A peer-reviewed, open-access journal of population sciences*

---

## ***DEMOGRAPHIC RESEARCH***

**VOLUME 28, ARTICLE 38, PAGES 1093-1144**

**PUBLISHED 29 MAY 2013**

<http://www.demographic-research.org/Volumes/Vol28/38/>

DOI: 10.4054/DemRes.2013.28.38

*Research Material*

**A Stata module for computing  
fertility rates and TFRs from  
birth histories: `tfrr2`**

**Bruno Schoumaker**

© 2013 Bruno Schoumaker.

*This open-access work is published under the terms of the Creative Commons Attribution NonCommercial License 2.0 Germany, which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit.*

*See <http://creativecommons.org/licenses/by-nc/2.0/de/>*

## Table of Contents

1	Introduction	1094
2	tfr2 in brief	1094
2.1	Why a Stata module for fertility rates?	1095
3	Birth histories in DHS surveys and recode data files	1096
4	Computing events and exposure from birth histories	1097
4.1	Example 1: Rates for the three years preceding the survey	1097
4.2	Example 2: Rates by calendar year (2005–2007)	1100
5	The tabexp command: Computing events and exposure	1102
5.1	Examples of tabexp	1105
5.1.1	Preparing a table of births and exposure for the three years preceding the survey	1105
5.1.2	Preparing a table of births and exposure for three calendar years	1107
5.1.3	Preparing a table using all women factors	1109
6	Poisson regression to compute fertility rates	1110
6.1	Age-specific fertility rates and TFR	1111
6.2	Reconstructing fertility trends	1112
6.3	Multivariate analyses of fertility	1113
7	Birth histories analyzed by tfr2	1114
7.1	Fertility rates and TFR computed by tfr2	1115
7.1.1	Age-specific fertility rates and TFR for the last three years	1115
7.1.2	Fertility rates by single year of age for the last five calendar years	1118
7.1.3	Fertility rates for sub-populations	1118
7.1.4	Using tfr2 with WFS and MICS surveys	1120
7.2	Reconstructing fertility trends using tfr2	1121
7.2.1	Reconstructing the TFR (15–49) over 15 years	1122
7.2.2	Reconstructing adolescent fertility over 30 years	1123
7.2.3	Comparing reconstructed fertility trends from successive surveys	1125
7.3	Rate ratios computed using tfr2	1127
7.3.1	Fertility differentials by education	1128
7.3.2	Multivariate model of recent fertility	1131
8	Conclusion	1133
	References	1134
	Appendices	1135

## **A Stata module for computing fertility rates and TFRs from birth histories: `tfr2`**

**Bruno Schoumaker<sup>1</sup>**

### **Abstract**

#### **BACKGROUND**

Since the 1970s, birth history data have become widely available, thanks to the World Fertility survey and the Demographic and Health Surveys programs. Despite their wide availability, these data remain under-exploited. Computation, even of simple indicators (fertility rates, total fertility rates, mean age at childbearing) and their standard errors, is not direct with such data, and other types of analysis (fertility differentials, reconstruction of fertility trends et cetera) may also involve reorganization of data sets and statistical modeling that present a barrier to the use of birth history data.

#### **OBJECTIVE**

This paper presents a Stata software module (`tfr2`) that was prepared to analyze birth history data in a user-friendly and flexible way. It is designed to be used primarily with DHS data, but can also be used easily with birth histories from other sources. Three types of analysis are performed by `tfr2`: (1) the computation of age-specific fertility rates and TFRs, as well as their standard errors, (2) the reconstruction of fertility trends, and (3) the estimation of fertility differentials (rate ratios).

#### **METHODS**

The `tfr2` module is composed of two parts: (1) a Stata command to transform birth history data into a table of births and exposure (`tabexp`), and (2) a Poisson regression model to compute fertility rates, fertility trends and fertility differentials from a table of births and exposure (produced by `tabexp`).

#### **COMMENTS**

One can obtain `tfr2` free of charge. It will work with Stata 10 and more recent versions of Stata.

---

<sup>1</sup> Université catholique de Louvain, Belgium. E-mail: [bruno.schoumaker@uclouvain.be](mailto:bruno.schoumaker@uclouvain.be).

## 1. Introduction

Since the 1970s, birth histories have become a major source of data on fertility in developing countries. Thanks to the World Fertility Surveys (WFS) and Demographic and Health Surveys (DHS), birth histories have been collected in a large number of countries and are publicly available and well documented.<sup>2</sup> Despite their wide availability, these data remain under-exploited. A possible reason for this is the fact that using birth history data is not a straightforward process; it usually involves data transformation, and even the computation of simple indicators (fertility rates, total fertility rates, mean age at childbearing) and their standard errors is not direct with such data.<sup>3</sup> Other types of analysis (such as fertility differentials and the reconstruction of fertility trends) also involve the reorganization of data and statistical modeling that may present a barrier to the use of birth history data.

The Stata module `tfr2`<sup>4</sup> was created to analyze birth history data in a user-friendly and flexible way. It is designed to be used primarily with DHS data, but can also easily be used with birth histories from other sources. In this paper, I present the way `tfr2` and its companion `tabexp` work, and I illustrate their use with birth histories from DHS, WFS and MICS. I discuss a few examples of analyses that can be done with `tfr2`, such as computing rates and TFRs on various types of periods, reconstructing fertility trends, and estimating multivariate models of recent fertility.

## 2. `tfr2` in brief

Stata command `tfr2` (.ado file) analyzes birth history data. Three types of analysis are performed by `tfr2`: (1) the computation of age-specific fertility rates and TFRs, as

---

<sup>2</sup> Birth histories have also been collected through other types of survey, such as the Multiple Indicator Cluster Surveys (MICS) and the World Health Surveys (WHS), which are also available free of charge.

<sup>3</sup> The MEASURE DHS project provides SPSS and SAS programs to compute fertility rates from birth histories, but the use of these programs is not straightforward, if one wants to compute fertility rates over different time periods or use them with other types of survey. The computation of correct standard errors of rates and TFRs, taking account of the clustering of observations, is not implemented in these syntaxes, either. Other researchers (Rodríguez 2006; Moultrie 2012; Pullum 2012) have also produced Stata programs to compute age-specific fertility rates. `tfr2` is designed to be more general (not limited to fertility rates) and user-friendly (a Stata command, rather than a program).

<sup>4</sup> Stata (StataCorp, 2011) is a software package widely used by demographers that offers powerful data management and statistical tools. `tfr2` uses Stata's capability to integrate users' commands that can be run in the same way as official Stata commands.

well as their standard errors, (2) the reconstruction of fertility trends, and (3) the estimation of fertility differentials (rate ratios). There are two parts composing `tfr2`:

- 1) A tool to transform birth history data into a table of births and exposure. A Stata command, `tabexp`, was created for this. It is used automatically by `tfr2` to transform data, but it can also be used separately.
- 2) A Poisson regression model to compute fertility rates, fertility trends and fertility differentials from a table of births and exposure (produced by `tabexp`). Standard errors are also computed by `tfr2`.

## 2.1 Why a Stata module for fertility rates?

The idea of this Stata module stems from several needs:

- 1) *Flexibility* – In various situations, it may be necessary to compute fertility rates that do not correspond to those published in the survey reports (or on MEASURE DHS' STAT compiler). For instance, fertility rates are usually published for the three years before the survey, but a longer period (e.g. five years) may be preferable in some cases, particularly when working on smaller populations, or on populations disaggregated by covariates. In some instances, rates need to be computed on calendar years instead of on years preceding the survey.
- 2) *User-friendliness* – The computation of fertility rates as published in DHS reports is not straightforward (Rutstein and Rojas 2006). Some programming is needed to compute the number of births and exposure between exact ages. The existing syntaxes provided by DHS can be adapted to other situations, but this is time-consuming and not necessarily easily done. A user-friendly tool that organizes the datasets in a flexible way facilitates the computation of fertility rates.
- 3) *Versatility* – The combination of a properly organized dataset and Poisson regression makes possible computing classical indicators of fertility (rates, TFR), reconstructing fertility trends, and conducting multivariate analyses within the same framework. Using the same framework makes the link between descriptive and multivariate analyses more explicit.

This Stata module is expected to:

- 1) Facilitate the computation of fertility rates and correct standard errors with birth history data from various types of surveys (eg. DHS, WFS, MICS).
- 2) Improve the evaluation of data quality, for instance by the computation of rates by single year of age and the reconstruction of trends of fertility rates by year.
- 3) Stimulate exploratory analysis of fertility trends and differentials.

### 3. Birth histories in DHS surveys and recode data files

Given the wide availability of DHS data, the examples presented in this paper have been mainly based on them. For this reason, I briefly explain the way birth histories are collected and organized in DHS data files. This explanation is based on the use of standard recode data files in Stata format (individual recode Stata system file, for instance, `boir51dt.dta` for the 2008 Bolivia survey).<sup>5</sup>

A birth history collects the dates of birth of all the children a woman has had in her life, starting from her first child until the time of the survey. In DHS, both the year and month of birth are recorded, and information on child survival is also collected. Birth histories are usually collected from a sample of women aged 15–49 at the time of the survey. In most DHS, birth histories are collected among all women, but in some countries, only women who have ever been married are eligible for fertility data.

Three types of information are necessary to compute fertility rates from birth histories: (1) the dates of births of the children, (2) the date of birth of each woman (whether or not she has ever given birth), and (3) the date of the survey.<sup>6</sup> These dates allow for locating events and computing exposure by age, period and cohort. Two other variables are, in some cases, also necessary: (1) a sampling weight variable to correct for the over- or under-sampling of some women because of sample design or differential response rate (Rutstein and Rojas 2006), and (2) an *all women factor*, which is used to compute age-specific fertility rates for all women when the sample is limited to women who have ever been married (Rutstein and Rojas 2006).

Table 1 illustrates typical birth history data with a few cases from the 2008 Bolivia DHS data file (`boir51dt.dta`). This file includes 16,939 observations (women), of which the first 10 are shown. The first variable (`caseid`) is the woman identifier.<sup>7</sup> The `v005` variable is the sampling weight variable.<sup>8</sup> The date of survey is recorded in `v008`, and is expressed in the Century Month Code (CMC).<sup>9</sup> The `v011` variable is the date of birth of the woman, also recorded in the CMC. Finally, all the births of the birth history are recorded (in the CMC) in variables `b3_01` to `b3_20` (only `b3_01` to `b3_10` are shown in Table 1), with `b3_01` corresponding to the most recent birth. Some women have never

---

<sup>5</sup> These files can be downloaded from the MEASURE DHS website ([www.measuredhs.com](http://www.measuredhs.com)).

<sup>6</sup> This is necessary only if rates are computed for a period defined by reference to the date of the survey (e.g. three years before the survey).

<sup>7</sup> For this example, I recoded this variable from 1 to n.

<sup>8</sup> This variable is “calculated to six decimals but [is] presented in the standard recode files without the decimal” (Rutstein and Rojas 2006, p. 14).

<sup>9</sup> The CMC code corresponds to the number of months since January 1900. It is “calculated by multiplying by 12 the difference between the year of an event and 1900. [...] The month of the event is added to the previous result” (Rutstein and Rojas 2006, p. 14).

given birth (e.g. the third woman). In such cases, dates of birth contain only missing values.

**Table 1: Illustration of birth history data in a DHS survey**

caseid	v005	v008	v011	b3_01	b3_02	b3_03	b3_04	b3_05	b3_06	b3_07	b3_08	b3_09	b3_10
1	773970	1299	972	1230	1217								
2	773970	1299	780	1294	1268	1227	1205	1178	1153	1122	1097	1079	1052
3	773970	1299	1079										
4	773970	1299	1097										
5	773970	1299	931	1283	1214	1189	1165						
6	773970	1299	1093										
7	773970	1299	961	1244	1220	1197	1175	1164					
8	773970	1299	1035										
9	773970	1299	800	1214	1130	1078							
10	773970	1302	1036	1270	1250								

## 4. Computing events and exposure from birth histories

The computation of period age-specific fertility rates requires counting events and measuring exposure in age groups for a defined period. In DHS reports, the rates are usually computed by five-year age groups (between exact ages) for the three years preceding the survey (Rutstein and Rojas 2006). The following section presents the way the computation of events and exposure is implemented in tfr2. We first use the example of rates published in DHS reports (five-year age groups, last three years), and then we discuss the computation of rates by calendar year over a three-year period.

### 4.1 Example 1: Rates for the three years preceding the survey

The Lexis diagram (Figure 1) shows the birth histories of the first five women in Table 1. Each woman's life is represented by a thin diagonal line; births are indicated by dots. The thick diagonal line represents the "life" of the (hypothetical) oldest woman in the data set; rates can only be computed for ages and periods below that diagonal. In this example, I consider that the survey was conducted in March 2008 (CMC 1299) for all the women.<sup>10</sup> Because the last month is incomplete, it is dropped for the computation of rates.<sup>11</sup> The period covered by rates for the last three years thus starts in March 2005

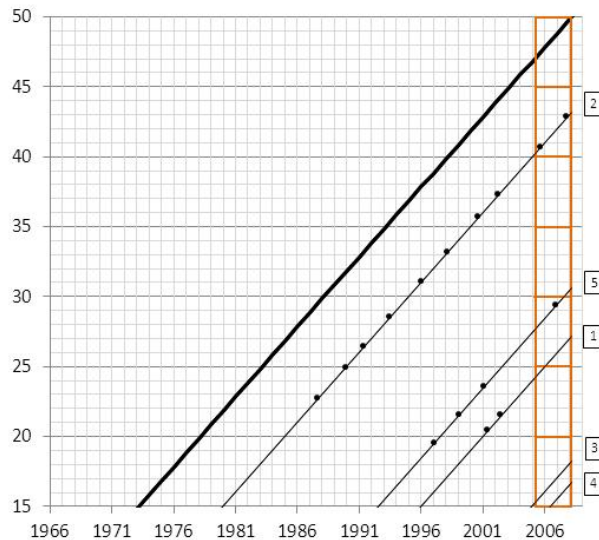
<sup>10</sup> This is correct for the five women represented on the Lexis diagram (Figure 1), but data collection usually extends over several months, so that in practice the date of survey varies across women.

<sup>11</sup> Any birth occurring in month 1299 will be dropped, and exposure in that month will not be included.

(beginning of month 1263) and ends in February 2008 (end of month 1298). Fertility rates are computed by dividing the number of births by the total exposure in each (orange) rectangle of the Lexis diagram. For the last age group (45–49), the rates will be slightly biased (upward) because of truncation.

A flexible approach to computing the number of births and exposure in the rectangles of the Lexis diagram consists of transforming the birth history into a person-period data file (Schoumaker 2004), and then aggregating the data by age groups into a table of births and exposure.

**Figure 1:** Illustration of birth history data on a Lexis diagram (births and exposure in five-year age groups for the three years preceding the survey)



(1) Transforming the birth history into a person-period data file consists of splitting each observation (woman) in the original data file into one or several lines, each line corresponding to a period in which the age group is constant. The number of births and exposure is computed in each period for each woman. We illustrate this in Table 2 using the five women in Figure 1. In the three years preceding the survey, the first woman spent nine months in the 20–24 age group, and 27 months in the 25–29 age



group.<sup>12</sup> She did not give birth during that period. The first line in the data file represents the period she spent in the 20–24 age group, and the second line the period she spent in the 25–29 age group. The number of births and exposure are measured for each period. The second woman spent 36 months in the 40–44 age group, and had two births during that period. There is only one line for her in the data file, as her age group did not change in the three years preceding the survey. The third woman spent 36 months in age group 15–19 and had no birth. The fourth woman turned 15 in month 1277; she spent 22 months in the 15–19 age group and had no birth during that period. Finally, the fifth woman spent 28 months in age group 25–29, and eight months in age group 30–34; she gave birth to one child in the three years preceding the survey, when she was aged 25–29. The information for all five women is represented over seven lines in the data file.

**Table 2: Illustration of the transformation of birth history data into a person-period data file (births and exposure in five-year age groups for the three years preceding the survey)**

caseid	v005	age_g	births	expos_m	expos_y
1	773970	20-24	0	9	0.75
1	773970	25-29	0	27	2.25
2	773970	40-44	2	36	3.00
3	773970	15-19	0	36	3.00
4	773970	15-19	0	22	1.83
5	773970	25-29	1	28	2.33
5	773970	30-34	0	8	0.67

Note: expos\_m: exposure in months; expos\_y: exposure in years.

(2) *Aggregating the person-period data file into a table of births and exposure* is done by summing the number of births and the exposure by age group (Table 3).<sup>13</sup> This preserves all the information needed to compute age-specific fertility rates and their standard errors (the total number of births and the total exposure in each age group). Both Table 2 and Table 3 can be analyzed by Poisson regression, which will lead to identical results. Poisson regression indeed provides equivalent results whether one works with individual data, person-period data or grouped data such as the tables of births and exposure (Powers and Xie 2000; Rodríguez 2007). Because the table of births and exposure is much smaller than the person-period data file, data storage needs

<sup>12</sup> She turns 25 in month  $972+300=1272$ . The three-year period starts in month 1263 (March 2005). The woman thus spends nine months (1272–1263) in the 20–24 age group, and 27 months (36–9) in the 25–29 age group.

<sup>13</sup> Sampling weights and all women factors are not used in this example. Their use is discussed later in the paper.

and computation time are greatly reduced. This is why `tfrr2` transforms birth history data into a table of events and exposure (using `tabexp`).

**Table 3: Illustration of the transformation of birth history data into a table of births and exposure in a DHS survey (births and exposure in five-year age groups for the three years preceding the survey)**

<code>age_g</code>	<code>births</code>	<code>expos_m</code>	<code>expos_y</code>
15-19	0	58	4.83
20-24	0	9	0.75
25-29	1	55	4.58
30-34	0	8	0.67
40-44	2	36	3.00

Note: `expos_m`: exposure in months; `expos_y`: exposure in years.

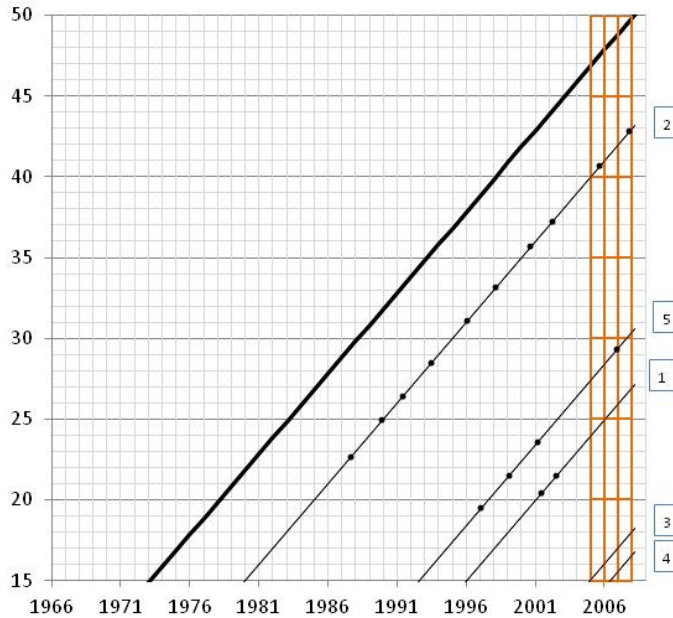
## 4.2 Example 2: Rates by calendar year (2005–2007)

The same principle can be used for any type of rate. Let us suppose that we want to compute fertility rates in five-year age groups by calendar year for 2005, 2006 and 2007 (rectangles in Figure 2).

The person-period data file will be constructed by splitting observations for each change of age group and year.

In this example, the person-period data file will have 15 lines (Table 4). The first woman was born in January 1981, and turns 25 in month 1272 (January 2006). She spends 12 months in the age group 20–24 in 2005; her change of age group coincides with the beginning of year 2006. She spends 12 months in the age group 25–29 in 2006, and 12 months in 2007. She is thus represented by three lines in the data file. She did not give birth in any of the periods. The second woman was born in January 1965; three lines will also be created in the data file, lasting 12 months each. She gave birth in 2005 and 2007. The third woman was born in December 1989; she turned 15 in December 2004, and she is also represented by three lines, each lasting 12 months (in the 15–19 age group). The fourth woman was born in May 1991 (month 1097). She turns 15 in May 2006. As a result, her exposure in the 15–19 age group in year 2006 is eight months, and she spends 12 months in that age group in 2007. The fifth woman was born in month 931 (July 1977). She spends 12 months in the age group 25–29 in 2005, 12 months in 2006, and six months in 2007. In July 2007, she turned 30, and spent six months in 2007 in the age group 30–34. She had a birth in 2006, when she was in the 25–29 age group.

**Figure 2: Illustration of birth history data on a Lexis diagram (births and exposure in five-year age groups for 2005, 2006, and 2007)**



**Table 4: Illustration of the transformation of birth history data into a person-period data file (births and exposure in five-year age groups for 2005, 2006, and 2007)**

caseid	v005	age_g	year	births	expos_m	expos_y
1	773970	20-24	2005	0	12	1.00
1	773970	25-29	2006	0	12	1.00
1	773970	25-29	2007	0	12	1.00
2	773970	40-44	2005	1	12	1.00
2	773970	40-44	2006	0	12	1.00
2	773970	40-44	2007	1	12	1.00
3	773970	15-19	2005	0	12	1.00
3	773970	15-19	2006	0	12	1.00
3	773970	15-19	2007	0	12	1.00
4	773970	15-19	2006	0	8	0.67
4	773970	15-19	2007	0	12	1.00
5	773970	25-29	2005	0	12	1.00
5	773970	25-29	2006	1	12	1.00
5	773970	25-29	2007	0	6	0.50
5	773970	30-34	2007	0	6	0.50

Note: expos\_m: exposure in months; expos\_y: exposure in years.

This person-period data file can be aggregated by age group and year (Table 5), and the table of births and exposure can be used to compute age-specific fertility rates by year.

**Table 5: Illustration of the transformation of birth history data into a table of births and exposure (births and exposure in five-year age groups for 2005, 2006, and 2007)**

<code>age_g</code>	<code>year</code>	<code>births</code>	<code>expos_m</code>	<code>expos_y</code>
20-24	2005	0	12	1.00
25-29	2005	0	12	1.00
25-29	2006	1	24	2.00
25-29	2007	0	18	1.50
30-34	2007	0	6	0.50
40-44	2005	1	12	1.00
40-44	2006	0	12	1.00
40-44	2007	1	12	1.00

Note: `expos_m`: exposure in months; `expos_y`: exposure in years

Sampling weights were not used in these examples, but their use is straightforward. Weights are normalized so that their sum is equal to the sample size of women. The weights are then used for the construction of the table of births and exposure, by computing weighted sums of births and exposure.<sup>14</sup> All women factors were not used in this example, either. When all women factors need to be used (in surveys in which only women who had ever been married were interviewed), individual exposure is first multiplied by the all women factor, and the table of births and exposure is then computed as described above (using sampling weights if necessary).

## 5. The `tabexp` command: Computing events and exposure

Although created to be part of `tfr2`, `tabexp` can be used as a stand-alone command to produce tables of events and exposure, as explained in section 4.<sup>15</sup> It is also used by `tfr2` to transform the birth history data into a proper table for analysis with Poisson regression. Even though it is not necessary to use `tabexp` separately to use `tfr2` (it

<sup>14</sup> The use of weights means that the number of births is not necessarily an integer. Although the Poisson model is supposed to be used for count data, it can also be estimated when the number of births is not an integer.

<sup>15</sup> `tabexp` uses important commands available in Stata, as the `stset` and `stsplit` commands to create the person-period data file, and the `collapse` command to produce a table of events and exposure.

is used automatically by `tfr2`), it is worth illustrating the way `tabexp` works, as it facilitates an understanding of `tfr2`.

The number of births and exposure between exact ages (in rectangles in the Lexis diagram) is computed by `tabexp` for time periods defined in various ways. The general syntax for the `tabexp` command is as follows:

```
tabexp [varlist] [if exp] [pweight = exp] [, options]
```

`varlist` is used to include covariates in the table of events and exposure. Using covariates will produce tables of births and exposure for all the values of these variables.

`pweight` allows the use of sampling weights. It automatically ensures that their sum is equal to the sample size (normalized weights). By default, `v005` (weight variable in DHS) is used as the sampling weight variable.

The other main other options indicate:

- The variables containing the relevant dates:<sup>16</sup>
  - `dates(varname)` indicates the date of the survey (`v008` in DHS)
  - `wbirth(varname)` indicates the dates of birth of the women (`v011` in DHS)
  - the variables containing the dates of births of children (`b3_01` to `b3_20` in DHS) are included in `bvar(varlist)`
- The size of the age groups (one or five years), and the minimum and maximum ages:
  - `ageg(#)` indicates the size of age groups: `ageg(5)` for five-year age groups, and `ageg(1)` for single ages.
  - `minage(#)` defines the lower age, and `maxage(#)` defines the upper age (by default, equal respectively to 15 and 49).
- The definition of the time period(s) for the preparation of the table:
  - `length(#)` defines the length of the period: `length(3)` means that births and exposure are computed for a three-year period.
  - The option `trend(#)` is used to set the width of the sub-periods for the computation of births and exposure. For instance, to produce a table by three-year periods over the last 15 years, the options `length(15) trend(3)` are used (`length` must be a multiple of `trend`).

---

<sup>16</sup> Currently, data on birth dates are expected to contain no missing or imprecise values. Missing values on birth dates should be imputed by the user before using `tfr2`. Missing values for characteristics of women (e.g. age, education) should also be treated separately by the user.

- `cy` indicates that events and exposure should be computed by calendar year (in contrast to years before the survey).
- `endy` can be used to indicate the last year of the period for the table. For instance, using options `length(5) endy(2007)` will prepare a table for years 2003 to 2007.
- The date of entry into the risk set:
  - `entry(varname)` can be used to indicate the date of entry of the individual into the risk set. For instance, to compute marital fertility rates, the date of entry would be the date of marriage. If analyses are restricted to periods after a migration (to a city, for example) the date of migration of each individual can be included. Only births and exposure after that date are taken into account. Dates should be indicated in CMC.

Additional options:

- `awf(varname)` is used to indicate the variable containing the all women factor (in DHS). By default `awf(awfactt)` is used, and `tabexp` automatically detects if the `awfactt` variable exists.
- `force` replaces the data file in memory by the table of events and exposure after it has been created.
- `frm` allows for fractional months. In most surveys, dates are collected only using months and year. This option randomly adds a fraction of a month to the dates of birth of children and of women (see example in Appendix 4).
- The option `rates` is used to display fertility rates (births divided by exposure) and their standard errors. This is a quick way to calculate rates, without all the options available in `tfr2` (trends, rate ratios).
- `savetab(filename)` saves the table of events and exposure as a Stata file.
- `nodis` disables the display of the results. It is used when the table of events and exposure needs to be prepared, but should not be displayed (for instance, when used by `tfr2`).
- `cluster(varname)` indicates that a cluster (primary sampling unit) variable should be taken into account. The computation of events and exposure will be done in each cluster separately (in a similar way to when using covariates).

## 5.1 Examples of `tabexp`

A few examples are used below to illustrate how `tabexp` works in typical situations. Appendix 1 also shows a series of Lexis diagrams and how `tabexp` can be used to produce the tables of events and exposure corresponding to the rectangles in the Lexis diagrams. Additional Stata do-files are available in appendix 4, and illustrate possible uses of `tabexp`.

### 5.1.1 Preparing a table of births and exposure for the three years preceding the survey

The first example produces a table of births and exposure for five-year age groups and for the three years preceding the survey. This corresponds to the table needed to compute fertility rates published in DHS reports (Figure 1). The following command will be used:

```
tabexp [pweight=v005], length(3) ageg(5) bvar(b3_*) dates(v008)
wbirth(v011)
```

Inputting only

```
tabexp
```

will produce the same result, because the default values/variables of `tabexp` were set in order to reproduce tables and rates in the DHS reports.

**Table 6: Computing births and exposure by five-year age groups for the three years preceding the survey, 2008 Bolivia DHS (computation using `tabexp`)**

<pre> . use BOIR51FL.DTA, clear . tabexp weight variable is v005 Preparing table of events and exposure for 3 year(s) preceding the survey Period covered: 3/2005 to 2/2008 Central date is 2006.7433 Number of cases (women): 16912 Number of person-years (weighted): 47524.059 Number of events (weighted): 5345.8384                 </pre>				
period	ageg	events	exposure	centry
0	15	882.43	10075.73	2006.743
0	20	1391.53	8020.29	2006.743
0	25	1379.8	7967.786	2006.743
0	30	852.623	6673.882	2006.743
0	35	586.892	6184.594	2006.743
0	40	226.836	5303.094	2006.743
0	45	25.7228	3298.686	2006.743

Table 6 shows the output of `tabexp` for the Bolivia 2008 DHS. The output contains 5 variables. The first variable indicates the period; here it is equal to 0 for the first period (in this case there is only one period). The second variable indicates the lower boundary of age groups, and the third and fourth variables contain the births and exposure. The fifth column is the central date of the period (2006.74).

```
tabexp, force
```

will replace the existing dataset with the table of events and exposure.

```
tabexp, nodis savetab(c:\table1.dta, replace)
```

will save the table in `c:\table1.dta` (and replace the file if it exists), and will not display results.

```
tabexp, cy
```

will produce the table of births and exposure for the three calendar years before the survey (Table 7).



**Table 7: Computing births and exposure by five-year age groups for the three calendar years preceding the survey, 2008 Bolivia DHS (computation using tabexp)**

<pre>. tabexp, cy weight variable is v005 Preparing table of events and exposure for 3 calendar year(s) preceding the year of the survey Period covered: 1/2005 to 12/2007 Central date is 2006.5 Number of cases (women): 16742 Number of person-years (weighted): 46968.188 Number of events (weighted): 5298.0869</pre>				
period	ageg	events	exposure	centry
0	15	879.955	9932.774	2006.5
0	20	1404.5	8023.217	2006.5
0	25	1352.31	7901.594	2006.5
0	30	841	6645.813	2006.5
0	35	581.961	6135.319	2006.5
0	40	210.387	5280.925	2006.5
0	45	27.972	3048.544	2006.5

### 5.1.2 Preparing a table of births and exposure for three calendar years

In order to compute the number of births and exposure by single calendar year in five-year age groups between 2005 and 2007 (Figure 2) the following command will be used:

```
tabexp [pweight=v005], length(3) ageg(5) bvar(b3_*) dates(v008)
wbirth(v011) trend(1) endy(2007)
```

The following command will produce the same results:

```
tabexp, trend(1) endy(2007)
```

Two additional options are used here (compared with the previous example): `trend(1)` indicates that births and exposure should be computed by one-year periods; `endy(2007)` indicates that the last year is 2007.

The option `rates` will display fertility rates for each age-period, as well as their standard errors.<sup>17</sup>

```
tabexp, trend(1) endy(2007) rates
```

**Table 8: Computing births and exposure five-year age groups for 2005, 2006 and 2007, 2008 Bolivia DHS (computation using `tabexp`)**

period	ageg	events	exposure	centry	rate	se_r
0	15	289.522	3127.123	2005.5	.092584	.0054412
1	15	322.19	3335.414	2006.5	.0965966	.0053815
2	15	268.243	3470.237	2007.5	.0772983	.0047196
0	20	478.809	2681.717	2005.5	.1785457	.0081596
1	20	467.016	2668.454	2006.5	.1750136	.0080985
2	20	458.677	2673.047	2007.5	.1715933	.0080121
0	25	434.192	2540.501	2005.5	.1709081	.008202
1	25	435.11	2630.474	2006.5	.1654111	.0079299
2	25	483.009	2730.618	2007.5	.1768862	.0080485
0	30	248.015	2172.252	2005.5	.1141743	.0072499
1	30	294.173	2208.166	2006.5	.1332206	.0077673
2	30	298.811	2265.395	2007.5	.1319025	.0076305
0	35	172.493	1979.69	2005.5	.0871311	.0066342
1	35	205.632	2049.459	2006.5	.1003349	.0069969
2	35	203.836	2106.17	2007.5	.0967805	.0067787
0	40	78.2486	1713.119	2005.5	.0456761	.0051636
1	40	76.1678	1769.379	2006.5	.0430478	.0049325
2	40	55.971	1798.427	2007.5	.0311222	.00416
0	45	10.8355	703.2753	2005.5	.0154072	.0046806
1	45	6.8539	1005.529	2006.5	.0068162	.0026036
2	45	10.2825	1339.74	2007.5	.007675	.0023935

<sup>17</sup> The rates are computed as  $[RATE=events/exposure]$  and their standard errors as  $[SE=rate/root(births)]$  (Keyfitz, 1966), equivalent to  $(SE= root(births)/exposure)$ .

Table 8 shows the output of `tabexp` for the Bolivia 2008 DHS. The list contains the same five variables as in Table 6, as well as rates and their standard errors. The period now varies from 0 to 2. The 0 period corresponds to year 2005 (central date is 2005.5). Births and exposure are computed for all the age groups and the three years.

### 5.1.3 Preparing a table using all women factors

As discussed before, when birth histories are only collected among women who have been married, individual exposure has to be multiplied by an all women factor that corrects for the fact that single women were not included in the birth histories (Rutstein and Rojas 2006). To indicate which variable contains the all women factors, the option `awf(varname)` is used:

```
tabexp, awf(awfactt)
```

This option computes the table of births and exposure by five-year age groups for the three years preceding, using the variable `awfactt` as the all women factor<sup>18</sup>.

The option

```
tabexp
```

will provide the same result: `tabexp` checks if the `awfactt` variable exists, and if its mean is different from 100, it is automatically used.<sup>19</sup> A message indicating that `awfactt` is used is displayed.<sup>20</sup> Table 9 shows the table of events and exposure using the all women factor in the 2008 Bangladesh survey.

---

<sup>18</sup> The `awfactt` variable is available in DHS data sets. It should be used when indicators are computed for the whole population. All women factors are available in DHS data files for analyses by place of residence, education or wealth quintile. Specific all women factors should be computed and used when working on other sub-populations.

<sup>19</sup> The all women factor is greater or equal to 100. If its mean is equal to 100, the all women factor can be ignored.

<sup>20</sup> Also displayed is a message warning that the correct all women factor should be used when producing tables for sub-populations.

**Table 9: Computing births and exposure by five-year age groups for the three years preceding the survey using all women factors, 2008 Bangladesh DHS (computation using `tabexp`)**

<pre>. tabexp By default the variable 'awfactt' is used with this data file. If you analyse fertility for sub-populations, use the correct all women factor. weight variable is v005 Preparing table of events and exposure for 3 year(s) preceding the survey Period covered: 5/2004 to 4/2007 Central date is 2005.878 Number of cases (women): 10986 Number of person-years (weighted): 36508 Number of events (weighted): 3579.926</pre>				
period	ageg	events	exposure	centry
0	15	1119.590	8869.827	2005.878
0	20	1203.510	6940.005	2005.878
0	25	724.397	5681.855	2005.878
0	30	351.901	5039.801	2005.878
0	35	145.180	4247.340	2005.878
0	40	32.688	3338.577	2005.878
0	45	2.661	2390.594	2005.878

## 6. Poisson regression to compute fertility rates

The second part of `tfr2` uses Poisson regression to compute fertility rates and to compute rate ratios or to reconstruct fertility trends from a table of births and exposure prepared by `tabexp`.<sup>21</sup>

Poisson regression is type of a generalized linear model in which the conditional distribution of the dependent variable is Poisson and the link function is logarithmic. It is used to analyze count data, such as number of births. By controlling exposure in an *offset* (a variable whose coefficient is equal to one), the model becomes a log-rate model (Powers and Xie 2000). This can be used to analyze birth histories in a flexible way (Schoumaker 2004). Births are the dependent variable, exposure is controlled with the offset, and independent variables include age groups and other types of covariates (such as time period and education).

<sup>21</sup> As shown in the previous section, fertility rates can be computed directly with `tabexp`. However, by using Poisson regression, `tfr2` offers a more general approach and allows for estimating multivariate models.

According to the Poisson model, the probability that the random variable  $Y_i$  is equal to the observed number of births ( $y_i$ ) is assumed to follow a Poisson distribution, with mean  $\mu_i$  (Winkelmann and Zimmermann 1994; Trussell and Rodríguez 1990).

$$P(Y_i = y_i | \mu_i) = \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \quad (1)$$

The mean  $\mu_i$  can be broken down into the product of fertility rate ( $\lambda_i$ ) and exposure ( $t_i$ ).

$$\mu_i = \lambda_i t_i \quad (2)$$

Taking the logarithm of this expression, it becomes:

$$\log(\mu_i) = \log(t_i) + \log(\lambda_i) \quad (3)$$

The regression model consists of modeling the logarithm of rates ( $\lambda_i$ ) as a linear combination of independent variables. In `tfrr2`, independent variables include a function of age and possibly additional covariates:

$$\log(\lambda_i) = \alpha + f(\text{age}) + g(\text{covariates}) \quad (4)$$

Replacing  $\log(\lambda_i)$  in Eq. 3. by Eq. 4, the Poisson regression that is estimated becomes:

$$\log(\mu_i) = \log(t_i) + \alpha + f(\text{age}) + g(\text{covariates}) \quad (5)$$

After fitting the model in Eq. 5, rates can be computed directly as the product of the exponentials of the functions of age and covariates (regression coefficients).

$$\lambda_i = \exp[f(\text{age})] * \exp[g(\text{covariates})] \quad (6)$$

## 6.1 Age-specific fertility rates and TFR

Classical indicators of fertility (rates, TFRs) — as well as their standard errors — can be obtained from the regression coefficients of a Poisson regression in which age groups are the only independent variable. In the example below, five-year age groups are included in the model as a series of dummy variables (the default option in `tfrr2`).

$$\log(\mu_i) = \log(t_i) + \alpha + \sum_{k=20-24}^{45-49} \beta_k A_{ki} \quad (7)$$

$\alpha$  is the constant term;  $A_{ki}$  are dummy variables for the six age groups from 20–24 to 45–49; the first age group (15–19) is the reference category.

The rate can be expressed in the following way:

$$\lambda_i = \exp[\alpha + \sum_{k=20-24}^{45-49} \beta_k A_{ki}] \quad (8)$$

Predicting fertility rates for a specific age group (e.g. 25–29 years) is straightforward. The dummy variable  $A$  is equal to 1 for the specific age group and 0 for the other age groups; the rate is then equal to the exponential of the sum of the constant and the coefficient of the corresponding age group (25–29).

$$\lambda_{25-29} = \exp[\alpha + \beta_{25-29}] \quad (9)$$

The total fertility rate (15–49) is equal to five times the sum of age-specific fertility rates.

$$TFR = 5 * (\exp[\alpha] + \sum_{k=20-24}^{45-49} \exp[\alpha + \beta_k]) \quad (10)$$

Standard errors of fertility rates and of the TFR can be computed from the standard errors of the regression coefficients using the delta method. In `tfr2`, a simple random sample is assumed by default (standard errors of the rates will in that case be identical to those computed by `tabexp`). Computing the standard errors for a two-stage sample is allowed by `tfr2` as well, using the jack-knife method and correcting for clustering. This is the same approach used in the DHS reports.

## 6.2 Reconstructing fertility trends

The Poisson model can also be used to reconstruct fertility trends from a birth history.<sup>22</sup> As in Eq. 7, age is controlled by a set of dummy variables. Calendar time is measured by dummy variables ( $T$ ) to model variations in fertility (annual variations in this example):

---

<sup>22</sup> This approach rests on the assumption of independence between mortality and fertility, and between migration and fertility. However, given that mortality is usually relatively low between 15 and 49 years of age (except in periods of high HIV prevalence with no antiretroviral treatment), this approach is reasonable even if this assumption does not hold.

$$\log(\mu_i) = \log(t_i) + \alpha + \sum_{k=20-24}^{45-49} \beta_k A_{ki} + \sum_{h=2}^{15} \delta_h T_{hi} \quad (11)$$

This model makes the assumption that the age pattern of fertility is constant (no interaction occurs between age and covariates). Although this does not strictly hold, simulations show that the assumption is reasonable for relatively short periods (e.g. 15 years).<sup>23</sup>

Predicting the fertility rate for a single age group (e.g. 25–29) for a specific year (e.g. year 5) is also straightforward. The dummy variables are equal to 1 for the specific age group and year (and 0 for the other age groups and years), and the rate is a function of the constant, the regression coefficient for the 25–29 age group, and the regression coefficient for the 5<sup>th</sup> year dummy variable.

$$\lambda_{25-29,5} = \exp[\alpha + \beta_{25-29}] * \exp[\delta_5] \quad (12)$$

The total fertility rate (15–49) for year h is equal to five times the sum of age-specific fertility rates for the reference year, multiplied by the exponential of the regression coefficient of the dummy variable for year h.

$$TFR_h = 5 * (\exp[\alpha] + \sum_{k=20-24}^{45-49} \exp[\alpha + \beta_k]) * \exp[\delta_h] \quad (13)$$

Several examples of reconstructed fertility trends are presented in sections 7.2.1, 7.2.2, and 7.2.3.

### 6.3 Multivariate analyses of fertility

Multivariate analyses of recent fertility can be performed in the same way as analyses of fertility trends. Instead of including time variables in the model, continuous or categorical independent variables are included (using dummy coding for categorical

---

<sup>23</sup> When mean age at childbearing decreases (fertility rates decline faster at higher ages), the consideration of a constant age pattern tends to slightly overestimate the TFR in recent years, and to underestimate it in earlier years. In the countries covered by DHS surveys, the decrease in the mean age at childbearing over a 15-year period rarely exceeds 1.2 years. In such cases, simulations indicate that the TFR is underestimated by about 3.5% in earlier periods (15 years before the survey), and is overestimated by about 1.5% at the time of the survey. In typical situations, however, the underestimation in earlier periods does not exceed 2%. When the mean age at childbearing increases (fertility rates decline faster at lower ages) — a much less common situation in developing countries — the method slightly underestimates the TFR in recent years, and overestimates it in earlier years. Again, the differences are relatively small.

variables), along with dummy variables for age groups.<sup>24</sup> The exponential of the regression coefficients of the independent variables are interpreted as rate ratios (ratios of TFRs). For categorical covariates, they represent the ratio of the TFRs for the categories of the independent variables compared to the reference category. For continuous variables, they represent ratios of TFRs associated with a one-unit increase of the explanatory variable. Currently, only time-constant variables can be included as covariates, and analyses including covariates should be limited to recent fertility.<sup>25</sup> Examples are presented in sections 7.3.1 and 7.3.2.

## 7. Birth histories analyzed by `tfr2`

The three types of model detailed in the previous section are estimated using `tfr2`. The syntax of `tfr2` is comparable to the syntax of `tabexp`, and includes a few additional features such as graphical options and saving options. The general syntax for the `tfr2` command is:

```
tfr2 [varlist] [if exp] [pweight = exp] [, options]
```

`varlist` is used to include covariates as in a regression model. By default, `tfr2` considers the covariates to be continuous. Categorical covariates are included using the `xi:` prefix. Rate ratios are interpreted as explained in section 6.3.

`pweight` allows using sampling weights. As for `tabexp`, it automatically ensures that they are normalized; `v005` is used as the default weight variable.

The options for the dates (births of children, birth of woman, date of survey, fractional months), the size of age groups, the definition of time periods and of entry time, the identifier of clusters, and the all women factor are the same as in `tabexp` (see section 5).

Additional options include:

- `mac`: computes mean age at childbearing.
- `savetable(filename)`: saves the table of events and exposure in a Stata file.

---

<sup>24</sup> The inclusion of both time variables and other independent variables has not yet been implemented in `tfr2`.

<sup>25</sup> Variables such as place of residence are not time-constant, and including place of residence as a time-constant covariate is thus not strictly correct. Despite the fact that it is an approximation, this is what is done in most survey reports when results are presented by place of residence. The influence of migration on fertility by place of residence will depend on the extent of migration, and on fertility differentials between in-migrants, out-migrants and non-migrants.



- `saverates(filename)`: saves ASFRs and TFR in a Stata file.
- `savetrend(filename)`: saves the reconstructed trend of TFRs in a Stata file.
- `grates`: displays a graph of ASFRs.
- `gtrend`: displays a graph of the reconstructed fertility trend.
- `se`: displays confidence intervals on graphs, and saves them in tables.
- `level(#)`: specifies the confidence level for confidence intervals.
- `input(wide|table)` indicates the format of the data. The default option is `wide`, and corresponds to the way the data are presented in standard recode data files in DHS. The `table` option is used if the data are in the same format as the data produced by `tabexp` (with the same variable names).

Several examples of birth history analysis using `tfr2` are presented below. Additional Stata do-files are available in appendix 4.

## 7.1 Fertility rates and TFR computed by `tfr2`

This first series of examples illustrates the use of `tfr2` to compute classical indicators of fertility (rates and TFRs) in various situations and for different types of surveys (DHS, WFS, MICS).

### 7.1.1 Age-specific fertility rates and TFR for the last three years

This example shows how to compute fertility rates and TFRs for the three years preceding the survey, as published in DHS reports. This corresponds to the rates illustrated in Figure 1. This is done directly by inputting

```
tfr2
```

This is equivalent to inputting

```
tfr2 [pweight=v005], len(3) ageg(5) bvar(b3_*) dates(v008) wbirth(v011)
```

Results using the 2010 Cambodia DHS are displayed in Table 10 (fertility rates and TFR are in the “Coef.” column). These rates are strictly identical to those published in the DHS report.

**Table 10: Age-specific fertility rates and TFR for the three years preceding the survey, 2010 Cambodia DHS (computation using `tfr2`)**

```

. use KHIR61FL.DTA
. tfr2
weight variable is v005
Preparing table of events and exposure for 3 year(s) preceding the
survey
Period covered: 9/2007 to 8/2010
Central date is 2009.2415
Number of cases (women): 18698
Number of person-years (weighted): 52535.961
Number of events (weighted): 5050.3301

```

ASFRs - TFR						
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Rate_1519	.0460590	.002088	22.06	0.000	.0419666	.0501513
Rate_2024	.1734467	.0042919	40.41	0.000	.1650346	.1818587
Rate_2529	.1667040	.004148	40.19	0.000	.1585741	.1748339
Rate_3034	.1205480	.0047668	25.29	0.000	.1112053	.1298906
Rate_3539	.0706275	.0032368	21.82	0.000	.0642835	.0769715
Rate_4044	.0276796	.0020441	13.54	0.000	.0236733	.0316858
Rate_4549	.0039000	.0009642	4.04	0.000	.0020102	.0057899
TFR	3.044824	.0442416	68.82	0.000	2.958112	3.131536

In addition, `tfr2` computes standard errors of the rates and the TFR, as well as the confidence intervals (95% by default). Standard errors in Table 10 are based on the assumption of a simple random sample. The use of `tfr2` also allows for jackknifing to compute standard errors, correcting for the clustering of observations within primary sampling units. This is done with option `cluster(varname)`, where `varname` contains the identifiers of the clusters.

```
tfr2, cluster(v001)
```

Results are displayed in Table 11. In this example, the standard error of the TFR is approximately 50% greater when clustering is taken into account.<sup>26</sup> The computation of correct standard errors is more time-consuming than assuming a simple random sample, but it is rendered straightforward with `tfr2`.

<sup>26</sup> This standard error is equal to the one published in the appendix on sampling errors in the DHS report.

**Table 11: Age-specific fertility rates and TFR for the three years preceding the survey, standard errors computed using jackknifing, 2010 Cambodia DHS (computation using `tfr2`)**

```

.tfr2, cluster(v001)
weight variable is v005
Preparing table of events and exposure for 3 year(s) preceding the
survey
Period covered: 9/2007 to 8/2010
Central date is 2009.2415
Number of cases (women): 18698
Number of person-years (weighted): 52535.961
Number of events (weighted): 5050.3301

```

ASFRs - TFR						
events	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Rate_1519	.0460590	.0028501	16.16	0.000	.0404618	.0516562
Rate_2024	.1734467	.0052076	33.31	0.000	.1632198	.1836736
Rate_2529	.1667040	.0046731	35.67	0.000	.1575266	.1758813
Rate_3034	.1205480	.0054725	22.03	0.000	.1098007	.1312952
Rate_3539	.0706275	.0039837	17.73	0.000	.0628042	.0784509
Rate_4044	.0276796	.0024704	11.20	0.000	.0228281	.032531
Rate_4549	.0039000	.0009793	3.98	0.000	.0019769	.0058232
TFR	3.044824	.0631887	48.19	0.000	2.92073	3.168918

Mean age at childbearing and its standard error can also be reported with the option `mac`. Using the `norates` and `notfr` options will only display mean age at childbearing.

```
tfr2, mac norates notfr
```

**Table 12: Mean age at childbearing for the three years preceding the survey, 2010 Cambodia DHS (computation using `tfr2`)**

<pre>. tfr2, mac norates notfr   weight variable is v005   Preparing table of events and exposure for 3 year(s) preceding the   survey   Period covered: 9/2007 to 8/2010   Central date is 2009.2415   Number of cases (women): 18698   Number of person-years (weighted): 52535.961   Number of events (weighted): 5050.3301</pre>						
<b>Mean age at childbearing (MAC)</b>						
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
MAC	28.279	.0987503	286.37	0.000	28.08546	28.47255

### 7.1.2 Fertility rates by single year of age for the last five calendar years

As with `tabexp`, the definition of age groups and periods is flexible. For instance, to compute fertility rates by single year of age for the last five calendar years, and to display these rates and their 90% confidence interval on a graph (Figure 3), the following command is used:

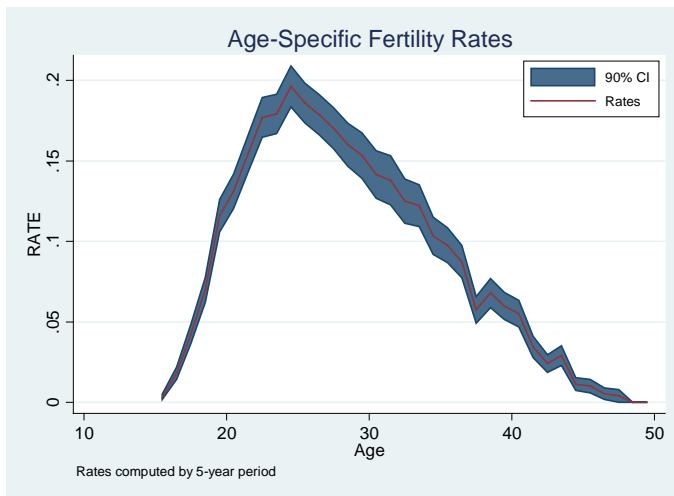
```
tfr2, ageg(1) length(5) cy gr se level(90)
```

### 7.1.3 Fertility rates for sub-populations

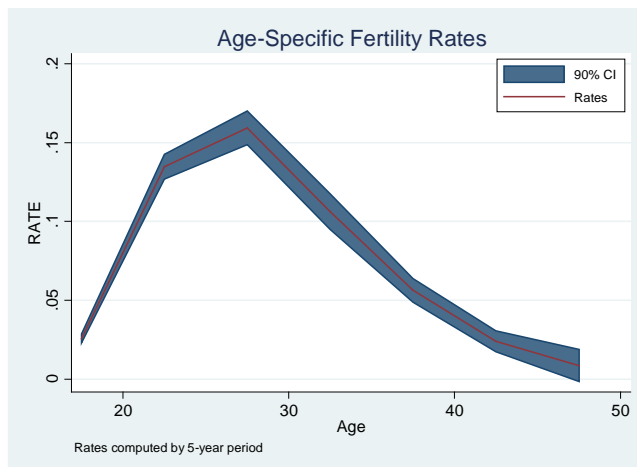
The same type of graph can be drawn for sub-populations by using the appropriate condition `if`. For example, fertility rates can be computed for women with secondary or higher education (`v106>=2`). Because the sample size is smaller, the rates are computed by five-year age groups in this example.

```
tfr2 if v106>=2, ageg(5) length(5) cy gr se level(90)
```

**Figure 3:** Age-specific fertility rates for the five calendar years preceding the survey, Cambodia 2010 DHS (computation using `tfr2`)



**Figure 4:** Age-specific fertility rates for the five calendar years preceding the survey among women with secondary or higher education, Cambodia 2010 DHS (computation using `tfr2`)

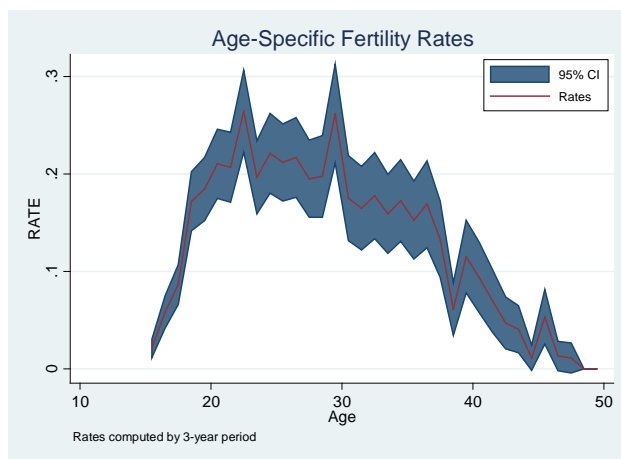


### 7.1.4 Using `tfr2` with WFS and MICS surveys

Data from other surveys can be processed using `tfr2`, as long as they are organized in a way similar to that of the DHS. For instance, WFS data can be very easily exploited with `tfr2`. The example below computes age-specific fertility rates in Colombia with the 1976 WFS data available on Germán Rodríguez's website,<sup>27</sup> and replicates his results. The data set (a selection of variables in Stata format) is downloaded from the website,<sup>28</sup> and appropriate variable names are used for the date of survey, the date of birth of the woman, and the dates of birth of the children.

```
use http://data.princeton.edu/eco572/datasets/cofertx, clear
tfr2, dates(v007) wb(v008) bvar(b0*2 b1*2) ageg(1) gr se
```

**Figure 5:** Age-specific fertility rates for the three years preceding the survey, Colombia 1976 WFS (computation using `tfr2`)



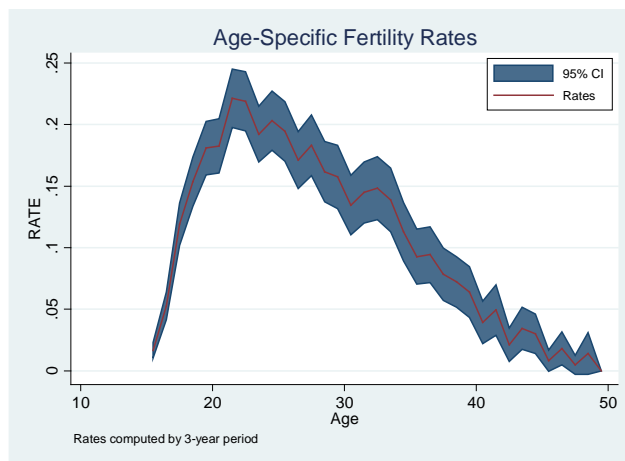
Some MICS surveys, in which birth histories were collected, can also be analyzed using `tfr2`. Contrary to DHS and WFS data, the birth history data file and the women data file need to be merged before using `tfr2`, and transformed into a format similar

<sup>27</sup> <http://data.princeton.edu/eco572/asfr.html>.

<sup>28</sup> WFS data are not directly available in Stata format. However, they can be downloaded from Princeton's Office of Population Research website and easily converted to Stata. See the example using Ghana WFS in section 7.2.3 and Appendix 2. The "Read ISI" package in R (<http://cran.r-project.org/web/packages/Read.isi/>) can also be used to convert WFS data into SPSS format, which can then be converted into Stata format.

to the DHS/WFS format. MICS data files are also not as highly standardized as the DHS or WFS, and variable names vary from one survey to the other. Despite these differences, `tfr2` can facilitate the analysis of birth histories in MICS surveys. Figure 6 (below) shows age-specific fertility rates computed with `tfr2` from the 2009 Zimbabwe MICS. The Stata syntax in Appendix 2 shows the data transformation and the use of `tfr2` with that MICS survey.

**Figure 6:** Age-specific fertility rates for the three years preceding the survey, Zimbabwe 2009 MICS (computation using `tfr2`)



## 7.2 Reconstructing fertility trends using `tfr2`

The TFR (15–49) can also be reconstructed using `tfr2` over a period of about 15 years. Figure (e) in appendix 1 shows the Lexis diagram illustrating the computation of births and exposure for reconstructing fertility trends by year. Although birth histories are truncated, and rates can not normally be estimated for ages and periods above the diagonal line corresponding to the oldest woman, using Poisson regression and making the assumption of a constant age pattern of fertility allows the reconstruction of the TFR for the 15–49 age group in the past (see section 6.2).

This approach can be used to evaluate data quality. Birth histories in DHS may be affected by various types of errors, for instance displacement and omissions of births (Schoumaker 2011). The lengthy health module, which is usually restricted to births

from January five years before the survey, may encourage interviewers to displace and/or omit births to avoid asking the questions in the health module. The reconstruction of TFR by calendar year offers a useful check of data quality: a sudden drop in the TFR at the cut-off year of the health module provides evidence of displacements and/or omissions of births.

### **7.2.1 Reconstructing the TFR (15–49) over 15 years**

The reconstruction of the TFR for the last 15 calendar years is illustrated below with the 2003 Mozambique survey (Table 13, Figure 7), and is done with the following commands.

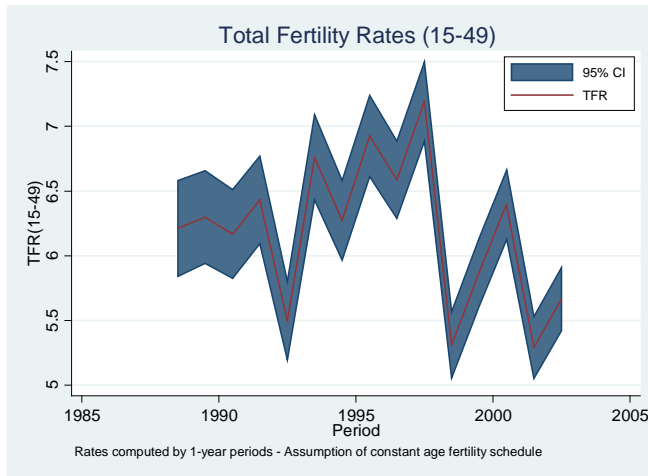
```
use mzir41f1.dta, clear

tfr2, len(15) trend(1) cy gt se
```

The figure below shows a sudden drop in the TFR (from 7.2 to 5.3) at the start of the health module (year 1998), which clearly suggests displacements and/or omissions of births.



**Figure 7: Total fertility rate (15–49) for the 15 calendar years preceding the survey, Mozambique 2003 DHS (computation using `tfr2`)**



### 7.2.2 Reconstructing adolescent fertility over 30 years

Trends can also be reconstructed for a specific age group; this allows reconstruction of trends over long periods for young age groups. For instance, the command below is used to reconstruct fertility for the 15–19 age group (partial total fertility rate) over the last 30 calendar years with the 2010 Colombia DHS.

```
use coir60fl.dta

tfr2, len(30) trend(1) minage(15) maxage(19) gt se cy
```

**Table 13: Total fertility rate (15–49) for the 15 calendar years preceding the survey, Mozambique 2003 DHS (computation using tfr2)**

```

. tfr2, len(15) trend(1) cy gt se

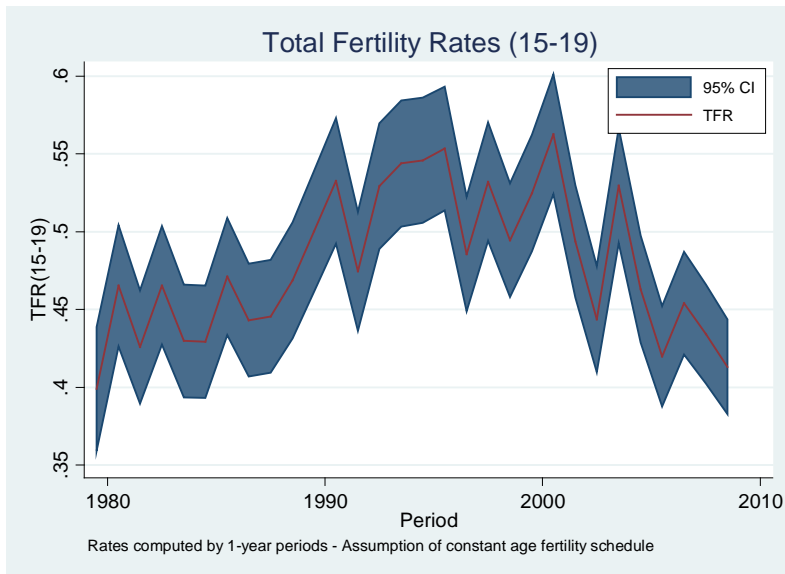
weight variable is v005
Preparing table of events and exposure for 15 calendar year(s)
preceding the year of the survey
Period covered: 1/1988 to 12/2002
Central date is 1995.5
Number of cases (women): 11978
Number of person-years (weighted): 125555.71
Number of events (weighted): 26415.5
    
```

ASFRs and TFR (average over the period)						
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Rate_1519	.1815951	.0023229	78.18	0.000	.1770423	.1861478
Rate_2024	.2638362	.0030127	87.58	0.000	.2579315	.2697409
Rate_2529	.2487996	.0032550	76.44	0.000	.2424198	.2551794
Rate_3034	.2148552	.0033979	63.23	0.000	.2081955	.2215148
Rate_3539	.16426	.0036188	45.39	0.000	.1571673	.1713527
Rate_4044	.0984417	.0038788	25.38	0.000	.0908394	.106044
Rate_4549	.0539136	.0057114	9.44	0.000	.0427196	.0651077
TFR	6.128507	.0493388	124.21	0.000	6.031804	6.225209

TFRs by 1-year periods - Assumption of constant age fertility schedule						
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
TFR_0	6.211032	.1881744	33.01	0.000	5.842217	6.579847
TFR_1	6.298593	.1826907	34.48	0.000	5.940526	6.65666
TFR_2	6.168181	.1740966	35.43	0.000	5.826958	6.509404
TFR_3	6.432248	.1724056	37.31	0.000	6.094339	6.770156
TFR_4	5.497554	.1543557	35.62	0.000	5.195022	5.800086
TFR_5	6.759918	.1674616	40.37	0.000	6.431699	7.088137
TFR_6	6.273411	.1565448	40.07	0.000	5.966589	6.580233
TFR_7	6.924626	.1603500	43.18	0.000	6.610345	7.238906
TFR_8	6.587153	.1522134	43.28	0.000	6.28882	6.885486
TFR_9	7.191719	.1564049	45.98	0.000	6.885171	7.498266
TFR_10	5.312160	.1303988	40.74	0.000	5.056583	5.567737
TFR_11	5.872705	.1343576	43.71	0.000	5.609369	6.136041
TFR_12	6.394282	.1371103	46.64	0.000	6.125551	6.663014
TFR_13	5.292314	.1217961	43.45	0.000	5.053598	5.53103
TFR_14	5.664699	.1239884	45.69	0.000	5.421686	5.907711

**Figure 8: Total fertility rate (15–19) for the 30 calendar years preceding the survey, Colombia 2010 DHS (computation using `tfr2`)**



Overall, there is an upward trend in 1980s and 1990s, followed by a downward trend since the late 1990s (Figure 8). These trends may reflect true changes, or they may be influenced by data quality problems (e.g. displacements and omissions of births, misreporting of women's ages).

### 7.2.3 Comparing reconstructed fertility trends from successive surveys

Comparison of fertility trends from successive surveys is also facilitated by use of `tfr2`. This can be used to reconstruct long-term fertility trends, as well as to evaluate data quality.

The first example below compares fertility trends from the two DHS in Mozambique (1997 and 2003). Using `tfr2`, this is done with a few lines of syntax: `tfr2` is used for each survey, results from `tfr2` for the two surveys are appended, and a graph of fertility trends from successive surveys is drawn.

```
cd "c:\DHS\"
local listDHS mzir31f1 mzir41f1
```

```
foreach survey of local listDHS {  
  use `survey'.dta, clear  
  tfr2, len(15) trend(1) cy savetr(trend_`survey'.dta, replace)  
  use trend_`survey'.dta, clear  
  rename TFR1 TFR_`survey'  
  sort date  
  save, replace  
}  
  
use trend_mzir31fl.dta, clear  
append using trend_mzir41fl.dta  
  
twoway (line TFR_* date, sort)
```

**Figure 9:** Total fertility rate (15–19) for the 15 calendar years preceding each survey, Mozambique 1997 and 2003 DHS (computation using `tfr2`)

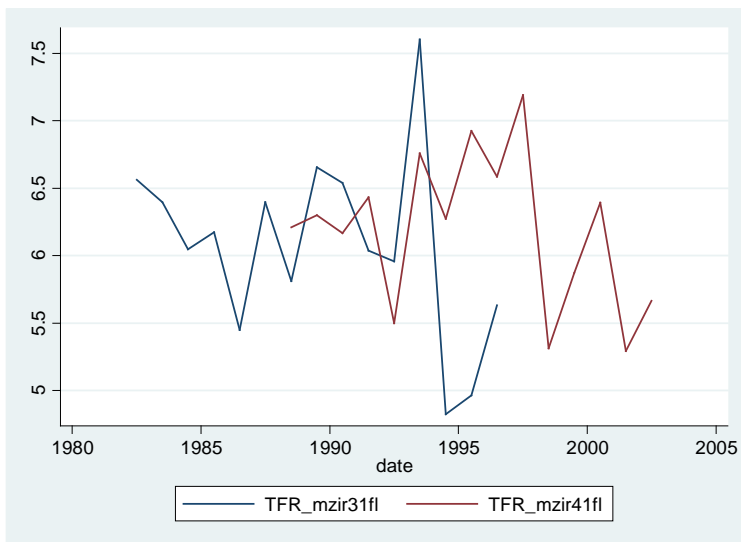


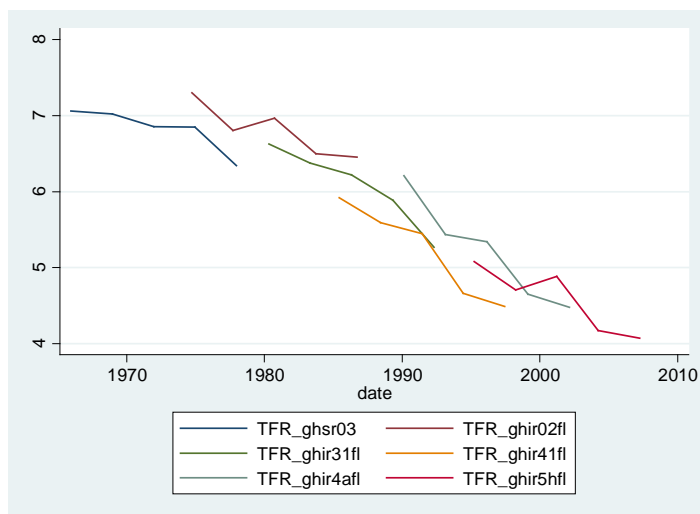
Figure 9 clearly illustrates the discrepancy between recent fertility in the 1997 survey, and fertility in the same period estimated from the 2003 survey, suggesting serious displacements and/or omissions of births in the 1997 survey.

The second example compares fertility trends from six surveys in Ghana, combining the 1979–80 WFS and five DHS (1988, 1993, 1998, 2003, and 2008). TFRs

are computed by three-year periods over the 15 years preceding each survey. The syntax (provided in Appendix 3) for combining the six surveys is a little longer than for the Mozambique example (mainly because the WFS data need to be imported from an ASCII file), but is still relatively short.

Figure 10 shows the decline in fertility that started in the late 1970s and early 1980s. Despite the clear downward trend, this figure also illustrates the relatively important discrepancies across surveys, again suggesting potential data quality problems in some of the surveys. For instance, recent fertility in the 1998 survey (TFR\_ghir41fl) seems to be underestimated, possibly reflecting omissions of recent births. Fertility in the 1988 DHS (TFR\_ghir02fl) is also higher than fertility in the 1980 WFS and in the 1993 DHS, indicating possible differences in sample composition.

**Figure 10:** Total fertility rate (15–19) for the 15 years preceding each survey (by three-year periods) in Ghana — 1979–80 WFS, 1988 DHS, 1993 DHS, 1998 DHS, 2003 DHS and 2008 DHS (computation using tfr2)



### 7.3 Rate ratios computed using tfr2

Finally, `tfr2` can be used with one or several covariates. This approach relies on the assumption that the age pattern of fertility is fairly similar across the values of the explanatory variables. Under that assumption, rate ratios are interpreted as ratios of

TFRs.<sup>29</sup> The way covariates are included is similar to what is done with regression models in Stata.

### 7.3.1 Fertility differentials by education

The table below (Table 14) shows the results for educational differentials in the 2008 Bolivia DHS. `v106` is used as a categorical covariate in this way:

```
xi: tfr2 i.v106
```

---

<sup>29</sup> This is a common assumption in regression models where no interaction between age and other covariates are included, and is similar to the proportional hazards assumption of event history models. Although it is not necessarily a correct assumption, fertility differentials (rate ratios) are not very sensitive to this assumption, as illustrated by the example below.

**Table 14: Fertility rates and rate ratios by level of education for the three years preceding the survey, Bolivia 2008 DHS (computation using `tfr2`)**

```

. xi : tfr2 i.v106
i.v106          _Iv106_0-3          (naturally coded; _Iv106_0
omitted)
Explanatory variables :_Iv106_1 _Iv106_2 _Iv106_3
weight variable is v005
Preparing table of events and exposure for 3 year(s) preceding the
survey
Period covered: 3/2005 to 2/2008
Central date is 2006.7433
Number of cases (women): 16912
Number of person-years (weighted): 47524.059
Number of events (weighted): 5345.8384

```

<b>ASFRs and TFR for the reference category</b>						
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Rate_1519	.1659365	.0113901	14.57	0.000	.1436123	.1882607
Rate_2024	.3320382	.0212218	15.65	0.000	.2904443	.3736322
Rate_2529	.3036214	.0187993	16.15	0.000	.2667755	.3404674
Rate_3034	.2121667	.0136655	15.53	0.000	.1853829	.2389506
Rate_3539	.1504648	.0101284	14.86	0.000	.1306134	.1703161
Rate_4044	.0647875	.0053820	12.04	0.000	.0542389	.0753361
Rate_4549	.011172	.0022605	4.94	0.000	.0067414	.0156025
TFR	6.200936	.3566947	17.38	0.000	5.501827	6.900044

<b>Rate ratios of explanatory variables - Assumption of constant age fertility schedule</b>	
Variable	Rate_ratios
_Iv106_1	.75309453***
_Iv106_2	.48176432***
_Iv106_3	.29574879***

Note: \* p<.1; \*\* p<.05; \*\*\* p<.01.

Age-specific fertility rates and TFRs are computed for the reference category (uneducated women,  $v106=0$ ), and rate ratios are displayed for the other categories of education (primary, secondary and higher education). This shows that fertility among the highly educated ( $v106=3$ ) is equal to 30% of the fertility among uneducated women ( $v106=0$ ). As mentioned before, this estimate relies on the assumption of proportionality of rates (constant age pattern of fertility).

Another way to measure fertility differentials with `tfr2` is to do stratified analyses. This is done with the `by` prefix. The following example computes the TFR by

categories of the v106 variable (rates not shown). This offers a quick way to compute fertility rates by level of education without making the assumption of proportionality of rates.

```
by v106, sort : tfr2, norates
```

**Table 15: TFRs by level of education for the three years preceding the survey, Bolivia 2008 DHS (computation using tfr2)**

```
. by v106, sort : tfr2, norates
```

---

```
-> v106 = no educa
weight variable is v005
Preparing table of events and exposure for 3 year(s) preceding the
survey
Period covered: 4/2005 to 3/2008
Central date is 2006.7808
Number of cases (women): 749
Number of person-years (weighted): 2236.8735
Number of events (weighted): 298.63803
```

<b>TFR</b>						
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
TFR	6.116523	.5006628	12.22	0.000	5.135241	7.097804

```
-> v106 = primary
weight variable is v005
Preparing table of events and exposure for 3 year(s) preceding the
survey
Period covered: 4/2005 to 3/2008
Central date is 2006.7738
Number of cases (women): 6832
Number of person-years (weighted): 19357.043
Number of events (weighted): 2719.7637
```

<b>TFR</b>						
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
TFR	4.748195	.0919421	51.64	0.000	4.567992	4.928398

```
-> v106 = secundar
weight variable is v005
Preparing table of events and exposure for 3 year(s) preceding the
survey
Period covered: 3/2005 to 2/2008
Central date is 2006.7277
Number of cases (women): 6075
Number of person-years (weighted): 16230.327
Number of events (weighted): 1632.4358
```



**Table 16: (Continued)**

<b>TFR</b>						
<b>events</b>	<b>Coef.</b>	<b>Std. Err.</b>	<b>z</b>	<b>P&gt; z </b>	<b>[95% Conf. Interval]</b>	
TFR	3.039606	.0827214	36.75	0.000	2.877475	3.201737
<pre> -&gt; v106 = higher weight variable is v005 Preparing table of events and exposure for 3 year(s) preceding the survey Period covered: 3/2005 to 2/2008 Central date is 2006.6996 Number of cases (women): 3256 Number of person-years (weighted): 9761.3428 Number of events (weighted): 657.90955 </pre>						
<b>TFR</b>						
<b>events</b>	<b>Coef.</b>	<b>Std. Err.</b>	<b>z</b>	<b>P&gt; z </b>	<b>[95% Conf. Interval]</b>	
TFR	1.881887	.0770486	24.42	0.000	1.730875	2.032899

Table 15 shows that the TFR among women with higher education is equal to 1.88, and among uneducated women is equal to 6.11. The ratio (0.307) is very close to the ratio estimated with the assumption of proportionality of rates (0.296; see Table 14). The same conclusion applies for other categories of the education variable, illustrating the minor impact in this case of the assumption of constant age pattern of fertility.

### 7.3.2 Multivariate model of recent fertility

A major advantage of using regression is, of course, the possibility of including several covariates. The following command evaluates the net effect of education (v106) in Bolivia, controlling for standard of living quintiles (v190) and place of residence (v025). In the previous example, the TFR of women with higher education was equal to 30% of the TFR of the uneducated women. Controlling for place of residence (urban/rural) and standard of living, the rate ratio is equal to 0.52 (category 3 of v106). In other words, the net effect of education on recent fertility is diminished when controlling for these two variables, but remains very strong and significant. The rate ratio of the women in the richest households compared to the poorest is 0.38, and urban women's fertility is 11% lower than their rural counterparts.

**Table 17: Fertility rates and rate ratios by level of education, standard of living and place of residence for the three years preceding the survey, Bolivia 2008 DHS (computation using tfr2)**

```

. xi: tfr2 i.v106 i.v190 i.v025
i.v106      _Iv106_0-3      (naturally coded; _Iv106_0 omitted)
i.v190      _Iv190_1-5      (naturally coded; _Iv190_1 omitted)
i.v025      _Iv025_1-2      (naturally coded; _Iv025_1 omitted)
Explanatory variables :_Iv106_1 _Iv106_2 _Iv106_3 _Iv190_2 _Iv190_3
_Iv190_4 _Iv190_5 _Iv025_2
weight variable is v005
Preparing table of events and exposure for 3 year(s) preceding the
survey
Period covered: 3/2005 to 2/2008
Central date is 2006.7433
Number of cases (women): 16912
Number of person-years (weighted): 47524.059
Number of events (weighted): 5345.8384

```

ASFRs and TFR for the reference category						
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Rate_1519	.2092753	.017182	12.18	0.000	.1755991	.2429515
Rate_2024	.4187188	.0325961	12.85	0.000	.3548316	.4826059
Rate_2529	.3958473	.0302613	13.08	0.000	.3365361	.4551584
Rate_3034	.2830199	.0223033	12.69	0.000	.2393062	.3267335
Rate_3539	.2026897	.0164534	12.32	0.000	.1704416	.2349378
Rate_4044	.0889283	.0084206	10.56	0.000	.0724242	.1054325
Rate_4549	.0154882	.0032154	4.82	0.000	.0091861	.0217903
TFR	8.069837	.5885886	13.71	0.000	6.916224	9.22345

Rate ratios of explanatory variables - Assumption of constant age fertility schedule	
Variable	Rate_ratios
_Iv106_1	.86432753**
_Iv106_2	.70679347***
_Iv106_3	.52205846***
_Iv190_2	.7464624***
_Iv190_3	.60725341***
_Iv190_4	.47019752***
_Iv190_5	.37893004***
_Iv025_2	.88987724**

Note: \* p<.1; \*\* p<.05; \*\*\* p<.01.

## 8. Conclusion

`tfr2` is a user-friendly and flexible tool for analyzing birth histories with the Stata software. By facilitating the descriptive and exploratory analyses of birth history data, this tool may contribute to getting the most out of the large amount of fertility data collected since the 1970s. It is particularly useful for computing fertility rates for specific populations (not published in survey reports), reconstructing fertility trends over relatively long periods and facilitating the evaluation of data quality. The estimation of standard fertility indicators and multivariate models of fertility within the same framework also makes the link between these approaches more explicit.

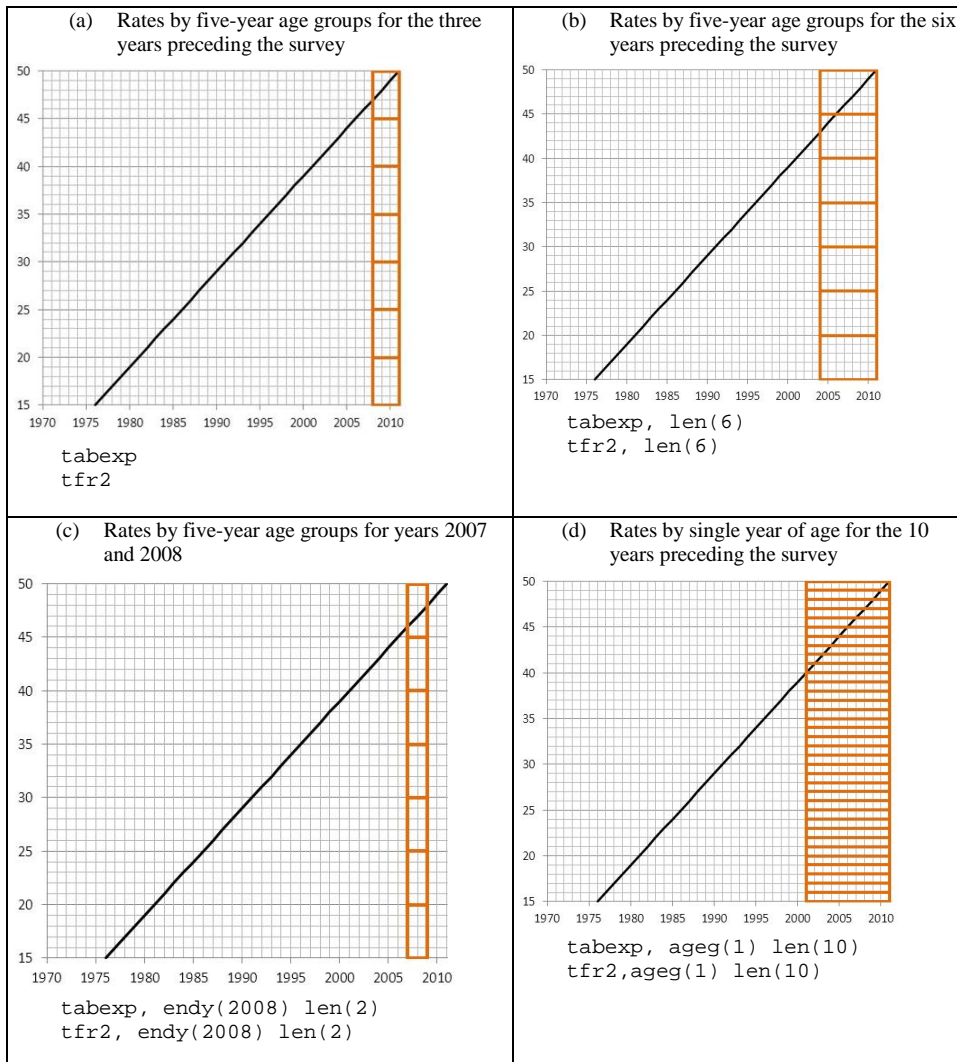
The analyses performed by `tfr2` essentially focus on the computation of age-specific fertility rates (and TFRs), their variations over time, and their variation across covariates. Other useful fertility indicators (e.g. parity progression ratios, parity-specific fertility rates) are not implemented in `tfr2`, but similar tools could be developed to complement `tfr2`.

## References

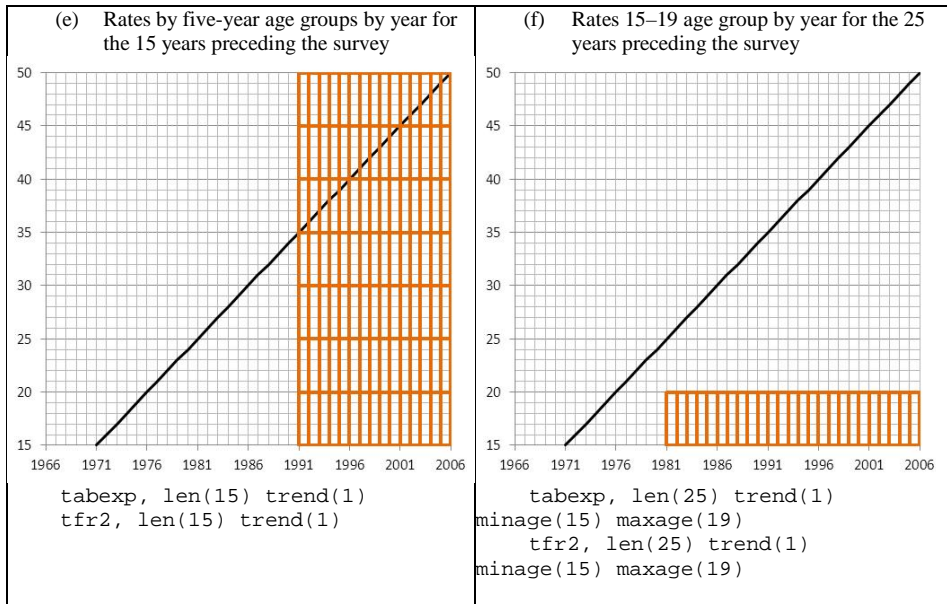
- Keyfitz, N. (1966). Sampling variance of standardized mortality rates. *Human Biology* 38(3): 309–317.
- Moultrie, T.A. (2012). Direct estimation of fertility from survey data containing birth histories. In: Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K.H., Timæus, I.M., and Zaba, B. (eds.). *Tools for Demographic Estimation*. <http://demographicestimation.iussp.org/content/direct-estimation-fertility-survey-data-containing-birth-histories>.
- Powers, D. and Xie, Y. (2000). *Statistical methods for categorical data analysis*. San Diego, CA: Academic Press.
- Pullum, T. (2012). *Program to produce ASFRs, TFR, GFR for specific windows of time, with covariates*. Personal communication.
- Rodríguez, G. (2006). *Lecture notes on research methods in demography*. <http://data.princeton.edu/eco572/>.
- Rodríguez, G. (2007). *Lecture Notes on Generalized Linear Models*. <http://data.princeton.edu/www509/notes/>.
- Rutstein, S.O. and Rojas, G. (2006). *Guide to DHS Statistics*. Demographic and Health Surveys, Calverton, MD: ORC Macro.
- Schoumaker, B. (2004). A person-period approach to analyzing birth histories. *Population (English Edition)* 59(5): 689–701. doi:10.2307/3654922.
- Schoumaker, B. (2011). *Omissions of recent births in DHS birth histories: Measurement and determinants*. Paper presented at the PAA annual Meeting, Washington, DC. <http://paa2011.princeton.edu/papers/112255>.
- StataCorp (2011). *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.
- Trussell, J. and Rodríguez, G. (1990). Heterogeneity in demographic research. In: Adams, J., Lam, D., Hermalin, A., and Smouse, P. (eds.). *Convergent Issues in Genetics and Demography*. New York: Oxford University Press: 111–132.
- Winkelmann, R. and Zimmermann, K. (1994). Count data models for demographic data. *Mathematical Population Studies* 4(3): 205–221. doi:10.1080/08898489409525374.

**Appendix 1:**

**Figure A1: Syntaxes of `tabexp` and `tfr2` for various types of rates, illustrated with Lexis diagrams**



**Figure A1: (Continued)**



## Appendix 2:

### Stata syntax for using tfr2 with MICS data (Zimbabwe 2009)

```
cd "G:\MICS\zimbabwe\"

* women data file
use wm.dta, clear
keep v1 v2 v3 v4 v5 v6 v251 v252 v248
rename v4 cluster
rename v5 hhd
rename v6 line
sort cluster hhd line
save woman, replace

* birth history data
use bh.dta, clear
keep v60 v3 v4 v5
rename v3 cluster
rename v4 hhd
rename v5 line
sort cluster hhd line

merge cluster hhd line using woman
cap drop _m

* creation of identifier for women
egen ident=group(cluster hhd line)
* creation of birth number
by ident (v60), sort: gene numb=_n
* transformation of the file into the wide format
reshape wide v60, i(ident) j(numb)

tfr2 [pw=v248], wb(v252) dates(v251) bvar(v60*) leng(3) gr se ageg(1)
```

## Appendix 3

### Stata syntax for comparing fertility trends from successive surveys in Ghana (WFS and DHS surveys).

```
cd "G:\wfs-dhs\"

local wfs ghsr03

* reading WFS data

infix v004 23-26 v005 27-30 v006 31-34 v007 35-38 v008 39-42 /*
*/ b012 182-185 b022 192-195 /*
*/ b032 202-205 b042 212-215 b052 222-225 b062 232-235 /*
*/ b072 242-245 b082 252-255 b092 262-265 b102 272-275 /*
*/ b112 282-285 b122 292-295 b132 302-305 b142 312-315 /*
*/ b152 322-325 b162 332-335 b172 342-345 b182 352-355 /*
*/ b192 362-365 b202 372-375 b222 382-385 b232 392-395 /*
*/ b242 402-405 /*
*/ using `wfs'.dat, clear

* using tfr2 to estimate fertility trends by 3 year periods over 15
  years (WFS survey)

tfr2[pw=v006], dates(v007) wb(v008) bvar(b*) len(15) trend(3)
savetr(trend_`wfs'.dta, replace)
use trend_`wfs'.dta, clear
rename TFR1 TFR_`wfs'
save, replace

local listDHS ghir02f1 ghir31f1 ghir41f1 ghir4af1 ghir5hf1

* using tfr2 to estimate fertility trends by 3 year periods over 15
  years (DHS surveys)

foreach survey of local listDHS {
  use `survey'.dta, clear
  tfr2, len(15) trend(3) savetr(trend_`survey'.dta, replace)
  use trend_`survey'.dta, clear
  rename TFR1 TFR_`survey'
  sort date
  save, replace
}

clear
local listsurv ghsr03 ghir02f1 ghir31f1 ghir41f1 ghir4af1 ghir5hf1

* data appended
```



```

foreach survey of local listsurv {
  append using trend_`survey'.dta
}

twoway (line TFR_* date, sort) /*

```

## Appendix 4

Stata syntaxes illustrating various uses of `tabexp` and `tfr2` (syntaxes available as separate do-files)

*.do file 1 (replication DHS.do)*

---

```

*****
** Replication of results of 2011 Uganda DHS
** http://www.measuredhs.com/pubs/pdf/FR264/FR264.pdf
*****

*****
* This do file replicates results typically published in DHS reports
* The Uganda 2011 DHS report is used.
* This syntax can be used with other surveys to produce comparable
  indicators
*****

cd "G:\DHS\"

use UGIR60FL.DTA, clear

**----- Content of Table 5.1 (p.57)
*Age-specific fertility rates and TFRS by urban-rural & total

by v025, sort : tfr2
tfr2

**----- First colum of Table 5.2 (p.59)
*TFRS by urban-rural, region, education, and wealth index

local listvar v025 v024 v106 v190
foreach var of local listvar {
  by `var', sort : tfr2, norates
}

**----- Table 5.3.1 (p.60)
*Fertility trends by age groups and 5-year periods

quietly {
  preserve

  tabexp, trend(5) len(20) rates force
  drop events expos se_r centry
  reshape wide rate, i(ageg) j(period)
}

```

```
}
    list

**----- Table 5.3.1 (p.60)
*Fertility in the three years preceding each survey (2000-01, 2006, 2011)

local listDHS UGIR41FL UGIR51FL UGIR60FL

foreach survey of local listDHS {
    use `survey'.dta, clear
    tfr2
}

**----- Figure 5.2 (p.61)
*Fertility in the three years preceding each survey (2000-01, 2006, 2011)

local listDHS UGIR41FL UGIR51FL UGIR60FL

foreach survey of local listDHS {
    use `survey'.dta, clear
    tfr2, notfr saver(ASFR_`survey', replace)
    use ASFR_`survey', clear
    rename rates1 rate_`survey'
    save, replace
}

foreach survey of local listDHS {
    merge 1:1 age using ASFR_`survey'.dta
    cap drop _m
    sort age
}

tway (line rate_* age, sort)
```

*.do file 2 (marital fertility rates.do)*

---

```
*****
** Computing age-specific marital fertility rates
*****

/*
This syntax computes age-specific marital fertility rates
and presents them on a graph. v509 is used as the date of
entry into the risk set.
*/

cd "G:\DHS\"

use mzir41fl.dta, clear
tfr2, entry(v509) gr level(90) se
```

*.do file 3 (Rates with fractional months.do)*


---

```

*****
** Estimating fertility rates with fractional months
*****

/*
This syntax computes fertility rates using fractional months.
The frm option replaces the dates of birth of women and children
(measured in years and months, expressed in CMC), by an 'exact' date.
This is done
by adding a random value between 0 and 1 to the CMC dates.
The seed is set in order to replicate the result if necessary.
*/

cd "G:\DHS\"
set seed 100
use KHIR61FL.DTA, clear
tfr2, frm

```

*.do file 4 (Fertility trends in successive surveys.do)*


---

```

*****
** Comparing annual fertility trends (15 years) in successive surveys****
*****

/*
This syntax computes total fertility rates (15-49)
over a 15-year period in 4 successive surveys, and shows
these trends on a graph.
*/

cd "G:\DHS\"

local listDHS  mlir01fl mlir32fl mlir41fl mlir52fl

foreach survey of local listDHS {
use `survey'.dta, clear
tfr2, len(15) trend(1) cy savetr(trend_`survey'.dta, replace)
use trend_`survey'.dta, clear
rename TFR1 TFR_`survey'
sort date
save, replace
}

foreach survey of local listDHS {
append using trend_`survey'.dta
}

twoway (line TFR_* date, sort)

```

*.do file 5 (Adolescent fertility trends in successive surveys.do)*

---

```
*****
** Trends in adolescent fertility in successive surveys in Colombia
*****

/*
This syntax computes partial total fertility rates (15-19)
over a 25-year period in 6 successive surveys, and shows
these trends on a graph.
Set maxvar may need to be changed to accomodate the large data files.
*/

cd "G:\DHS\"
local listDHS  coir01fl coir22fl coir31fl coir41fl coir52fl coir60fl

foreach survey of local listDHS {
use `survey'.dta, clear
tfr2, len(25) trend(1) maxage(19) savetr(trend_`survey'.dta, replace)
use trend_`survey'.dta, clear
rename TFR1 TFR_`survey'
sort date
save, replace
}

foreach survey of local listDHS {
append using trend_`survey'.dta
}

tway (line TFR_* date, sort)
```

*.do file 6 (Fertility before and after.do)*

---

```
*****
** Comparing fertility before and after an event*****
*****

/*
This syntax computes the TFR for the a 3-year period before a
specific date, and for the 3-year period after that date.
This is done by creating an artificial survey date 3 years after
that date.
The event occurred in CMC month 1176; the date of survey is considered
as 1176+37 months (the last month is removed). Fertility trends are
computed by 3-year periods
over a 6-year period. A graph with 90% confidence interval is shown.
*/

cd "G:\DHS\"
use mzir41fl.dta, clear

gene cutoff=1176

gene dates2=cutoff+36+1

tfr2, dates(dates2) len(6) trend(3) gt level(90) se
```

*.do file 7 (Graph ASFRS by period.do)*

---

```
*****
** Graphs of age-specific fertility rates by 5-year periods
*****

/*
This syntax computes age-specific fertility rates by
5-year periods and presents rates on a graph.
It uses tabexp to compute age-specific fertility rates.
The file is reshaped to have rates presented in columns.
*/

cd "G:\DHS\"

use ugir60fl.dta, clear
preserve
tabexp, trend(5) len(15) rates force
drop events expos se_r centry
reshape wide rate, i(ageg) j(period)
tway (line rate* ageg, sort)
```

*.do file 8 (Graph ASFRS by categories.do)*

---

```
*****
** Graphs of age-specific fertility rates by categories
*****

/*
This syntax computes age-specific fertility rates by
categories of the V106 variable (education) and presents
rates on a graph. It uses tabexp to compute age-specific
fertility rates for the 3 years preceding the survey.
The file is reshaped to have rates presented in columns.
*/

cd "G:\DHS\"

use ugir60fl.dta, clear
preserve
tabexp v106, force rates
drop events expos se_r
reshape wide rate, i(ageg) j(v106)
tway (line rate* ageg, sort)
```