

# **tfr2** : A Stata module for computing fertility rates and TFRs from birth histories<sup>1</sup>

Bruno Schoumaker

Centre de recherche en démographie et sociétés, UCL (Belgium)

PAA Meeting, San Francisco, May 3-5, 2012

DRAFT – April 19, 2012

## **1. Introduction**

Since the 1970s, birth histories have become a major way of collecting 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 freely available to researchers. Despite their wide availability, these data remain under-exploited. A possible reason for this is the fact that their use is not straightforward.

As often with longitudinal data, data organization represents a crucial step in the analysis of birth history data. 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>2</sup>. Other types of analyses (fertility differentials, reconstruction of fertility trends...) may also involve reorganization of data sets and statistical modeling that represent a barrier to the use of birth history data.

---

<sup>1</sup> `tfr2` can be obtained from Bruno Schoumaker ([bruno.schoumaker@uclouvain.be](mailto:bruno.schoumaker@uclouvain.be)). It will soon be available online.

<sup>2</sup> SPSS and SAS programs are provided by the Measure/DHS program to compute fertility rates from birth histories, but their adaptation is not straightforward if one wants to compute fertility rates on different time periods, or to use them with other types of surveys. The computation of correct standard errors of rates and TFRs, taking account of the clustering of observations, is also not implemented in these syntaxes. Other researchers (Rodriguez, 2006, Moultrie, 2011; Pullum, 2012) have also produced Stata programs to compute age-specific fertility rates. `tfr2` is designed to be more general (more possibilities that fertility rates) and user-friendly (a Stata command, rather than a program).

The `tfr2` module for Stata<sup>3</sup> 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 (e.g. WFS). In this paper, I present the way the `tfr2` module (and its companion `tabexp`) works, and discuss a few examples of analyses that can be done easily with `tfr2`. I illustrate its use with birth histories from DHS and WFS.

## 2. TFR2 in brief

`tfr2` is a stata command (.ado file) to analyze birth history data. Three types of analyses 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). The `tfr2` module is composed of two parts:

- 1) A tool to transform birth history data into a table of births and exposure. A Stata command called `tabexp` was created for this. `tabexp` is used by `tfr2`, but 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`.

## 3. Why a Stata module for fertility rates?

The idea of making a Stata module for analyzing birth history data 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 the Statcompiler for DHS). For instance, fertility rates are usually published for the 3 years before the survey, but a longer period (e.g. 5-year) may be preferable in some cases – notably when working on smaller populations. In some instances, rates need to be computed on calendar years instead of ‘years preceding the survey’.

---

<sup>3</sup> Stata is a software that is largely used among demographers, and 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.

- 2) *User-friendliness* - The computation of fertility rates as published in DHS reports is not straightforward. 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 allows computing classical indicators of fertility (rates, TFR), reconstructing fertility trends, and conducting multivariate analyses within the same framework. This in turn facilitates the transition from descriptive to multivariate analyses.

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 (DHS, WFS,...)
- 2) Improve evaluation of data quality through the computation of rates by single year of age, reconstruction of trends of fertility rates by year,...
- 3) Stimulate exploratory analysis of fertility trends and differentials

#### **4. Birth histories in DHS surveys and recode data files**

The `tfr2` module can be used with birth histories collected in various types of surveys. Given their wide availability, this paper will mainly use examples from DHS data. For this reason, I briefly explain the way birth histories are collected and organized in DHS data files. This presentation 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>4</sup>.

A birth history collects the dates of all the births a woman has had in her life, starting from her first birth until the time of the survey. In DHS, both the year and month of birth are recorded, and additional 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 married women 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

---

<sup>4</sup> These files can be downloaded from Measure/DHS website.

woman (whether or not she has ever given birth), and (3) the date of the survey. These dates allow 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 general fertility rates for all women when the sample was limited to ever married women (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>5</sup>. The v005 variable is the sampling weight variable<sup>6</sup>. The date of survey is recorded in v008, and is expressed in Century Month Code (CMC)<sup>7</sup>. The v011 variable is the date of birth of the woman, also recorded in CMC. Finally, all the births of the birth history are recorded (in 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 given birth (e.g. third woman). In such a case, dates of birth contain only missing values.

Table 1. Illustration of birth history data in 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	1241	1214	1130	1078	.	.	.	.	.	.
10	773970	1302	1036	1270	1250	.	.	.	.	.	.	.	.

### 5. Computing events and exposure from births histories

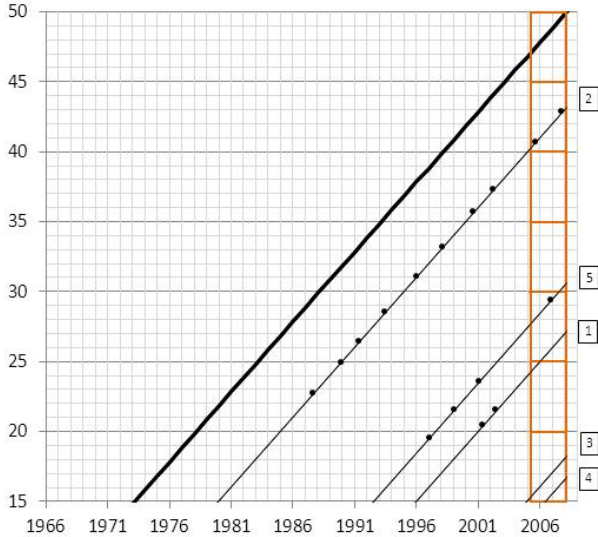
The computation of age-specific fertility rates requires counting events and measuring exposure in age groups during a defined period. In DHS reports, the rates are usually computed by 5-year age groups (between exact ages) for the 3

<sup>5</sup> For this example, I recoded this variable from 1 to n.  
<sup>6</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>7</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).

years preceding the survey (Rutstein and Rojas, 2006). The following example illustrates the computation of events and exposure in this situation, as implemented in `tfr2`.

The Lexis diagram (Figure 1) shows the birth histories of the first five women of Table 1. The thick diagonal represents the line of the (hypothetical) oldest woman in the data set; rates can only be computed for ages and periods below that diagonal. Each thin diagonal line represents the life of one of the first five women from Table 1; births are indicated by dots. In this example, I consider that the survey was conducted in March 2008 (CMC equal to 1299) for all women<sup>8</sup>. Because the last month is incomplete, it is dropped for the computation of rates<sup>9</sup>. The period covered by rates for the last 3 years thus starts in March 2005 (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.

Figure 1. Illustration of birth history data on a Lexis diagram. Births and exposure in 5-year age groups for the 3 years preceding the survey.



A flexible approach to compute the number of births and exposure in the rectangles of the Lexis diagram consists in 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.

<sup>8</sup> This is correct for the 5 women represented on the Lexis diagram, but data collection usually extends over several months, so that the date of survey varies across women.

<sup>9</sup> Any birth occurring in month 1299 will be dropped, and exposure in that month will not be included.

**(1) Transforming the birth history into a person period data** file consists in splitting each observation in the original data file (woman) into one or several lines, where each line corresponds to a period in which the age group is constant. The number of births and exposure is computed in each period. We illustrate this in Table 2 with the five women of Figure 1.

In the 3 years preceding the survey, the first woman spent 9 months in the 20-24 age group, and 27 months in the age group 25-29<sup>10</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 age group 40-44, 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 8 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 the 5 women is represented by 7 lines in the data file. This approach will lead to a data file with a larger number of lines than the original data set.

Table 2. Illustration of transformation of birth history data into a person-period data file (births and exposure in 5-year age groups for the 3 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

Legend : expso\_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>11</sup>. This preserves all the information needed to compute age-specific fertility rates and their standard errors (the total number of births and

<sup>10</sup> She turns 25 in month  $972+300=1272$ . The 3-year period starts in month 1263 (March 2005). The woman thus spends 9 months (1272-1263) in the 20-24 age group, and 27 months (1298-1272+1) in the 25-29 age group.

<sup>11</sup> As discussed later, when sampling weights are used, it involves computing weighted sums, where weights are normalized so that their sum is equal to the sample size. When all women factors are used, it is necessary to multiply exposure (only) by the all women factor before summing exposure.

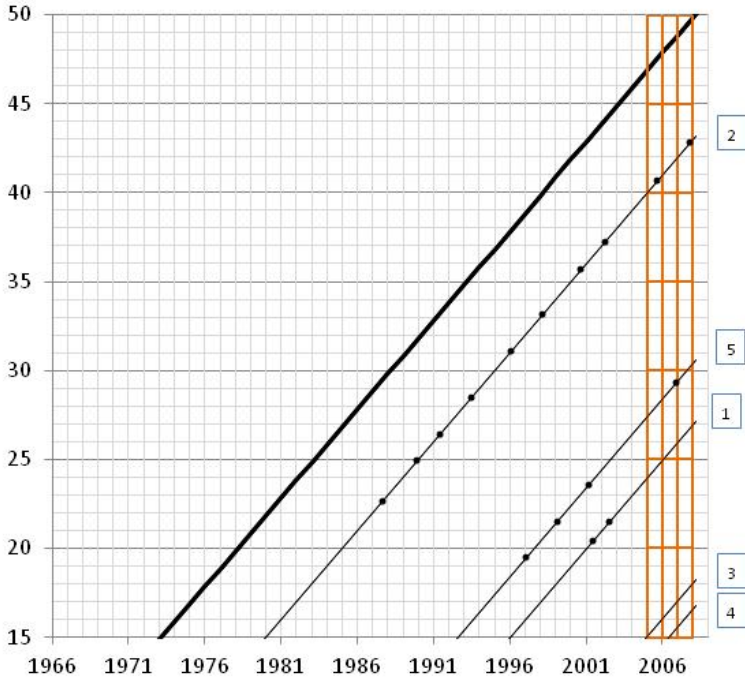
the total exposure in each age group). Both Table 2 and Table 3 can be analyzed with Poisson regression, and will lead to identical results. Poisson regression indeed provides equivalent results whether one works with individual data, person-period data or grouped data such as the tables of births and exposure (Powers and Xie, 2000; Rodriguez, 2007). For data storage and computation reasons, it is preferable to work with tables of events and exposure. This is why `tfr2` transforms birth history data into a table of events and exposure (using `tabexp`).

Table 3. Illustration of transformation of birth history data into a table of births and exposure in DHS survey (births and exposure in 5-year age groups for the 3 years preceding the survey).

age_g	births	expos_m	expos_y
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

The same principle can be used for any types of rates. Let us suppose that we want to compute fertility rates in 5-year age groups by calendar years for 2005, 2006 and 2007 (rectangles on Figure 2). The person period data file will be constructed by splitting observations for each change of age group or year.

Figure 2. Illustration of birth history data on a Lexis diagram. Births and exposure in 5-year age groups 2005, 2006 and 2007.



The person period data file will include 15 lines in this example (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 have births 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 turns 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 8 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 6 months in 2007. In July 2007, she turns 30, and spends 6 months in 2007 in the age group 30-34. She had a birth in 2006, when she was in the 25-29 age group.

Table 4. Illustration of transformation of birth history data into a person-period data file (births and exposure in 5-year age groups in 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

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.

Sampling weights were not used in these examples, but their use is straightforward. The weight that is attached to a woman in the births history data is also attached to a woman in the person-period data. The weights are used for the construction of the table of births and exposure, by computing weighted sums of births and exposure. The use of all-women factors is different. It is only



used to weight exposure, but the number of births is not affected by all women factors.

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

age_g	year	births	expos_m	expos_y
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

## 5.1 The `tabexp` command: computing events and exposure

`tabexp` was prepared as part of `tfr2`. This command can be used as a stand-alone command to produce tables of events and exposure as explained in the previous section. It is also used by `tfr2` to transform the birth history data into a proper table for the analysis with Poisson regression. Even though it is not necessary to use `tabexp` separately to use `tfr2` (it will be used automatically by `tfr2`), it is worth illustrating how `tabexp` works as it facilitates understanding `tfr2`.

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

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

`varlist` is used to include (categorical) covariates in the table of events and exposure. Using covariates will produce tables of births and exposure for all the categories of the variables.

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

The main options are used to indicate:

- The variables containing the relevant dates:
  - `dates(varname)` indicates the date of the survey (`v008` in DHS)

- `wbirth(varname)` indicates the date 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 (1 year or 5 years), and the minimum and maximum ages:
  - `ageg(#)` indicates the size of age groups : `ageg(5)` for 5-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 3-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 3-year periods over the last 15 years, the options `length(15) trend(3)` should be used (`length` must be a multiple of `trend`).
  - `cy` indicates that events and exposure should be computed by calendar years (by contrast with years before the survey),
  - `endy` can be used to indicate the last year of the period for the table. For instance, using options `length(3) endy(2007)` will prepare a table for years 2005 to 2007.
- The date of entry into the risk set:
  - `entry(varname)` can be used to indicate the date of entry of the individual in 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.
- `force` replaces the data file in memory by the data file of events and exposure.

- The option `rates` is used to display fertility rates (births divided by exposure). This is a quick way to calculate rates, without all the options available in `tfr2` (standard errors, trends...).
- `savetab(filename)` allows saving the table of events and exposure in 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 (in a similar way as when using covariates).

## 5.2 Examples of `tabexp`

A few examples are used to illustrate how `tabexp` works in typical situations.

### **Example 1: preparing a table of births and exposure for the last 3 years**

The first example produces a table of births and exposure for 5-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)
```

Typing 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 shows the output of `tabexp` for the Bolivia 2008 DHS. The output contains 5 variables. The first variable indicates the period. It is equal to 0 for the first period (there is only one period here); the second variable indicates the lower boundary of age groups, 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 by 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 3 calendar years before the survey.

Table 6. Computing births and exposure by 5-year age groups for the 3 years preceding the survey, 2008 Bolivia DHS (computation with tabexp)

period	ageg	events	exposure	centry
0	15	882.692	10078.72	106.7433
0	20	1391.94	8022.67	106.7433
0	25	1380.21	7970.151	106.7433
0	30	852.876	6675.862	106.7433
0	35	587.066	6186.429	106.7433
0	40	226.903	5304.668	106.7433
0	45	25.7305	3299.664	106.7433

### **Example 2: Preparing a table of births and exposure for 3 calendar years**

In order to compute the number of births and exposure by single calendar year in 5-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 simplified command will give the same results.

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

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

Table 7 shows the output of tabexp for the Bolivia 2008 DHS. The list contains the same 5 variables as in Table 6. 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.

Table 7. Computing births and exposure for the by 5-year age groups for 2005, 2006 and 2007, 2008 Bolivia DHS (computation with tabexp)

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

endy replaces date of survey

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
```

period	ageg	events	exposure	centry
0	15	289.704	3129.094	2005.5
1	15	322.393	3337.516	2006.5
2	15	268.413	3472.424	2007.5
0	20	479.111	2683.407	2005.5
1	20	467.31	2670.136	2006.5
2	20	458.966	2674.731	2007.5
0	25	434.466	2542.102	2005.5
1	25	435.384	2632.132	2006.5
2	25	483.313	2732.339	2007.5
0	30	248.172	2173.621	2005.5
1	30	294.359	2209.557	2006.5
2	30	299	2266.823	2007.5
0	35	172.601	1980.938	2005.5
1	35	205.762	2050.751	2006.5
2	35	203.965	2107.497	2007.5
0	40	78.2979	1714.199	2005.5
1	40	76.2158	1770.495	2006.5
2	40	56.0063	1799.56	2007.5
0	45	10.8424	703.7186	2005.5
1	45	6.85822	1006.162	2006.5
2	45	10.289	1340.585	2007.5

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

will also display the fertility rates for each age-period

```
. tabexp, minage(15) maxage(19) len(25) trend(5)
```

will compute births and exposure in the 15-19 age group by 5-year period over the 25 years preceding the survey.

***Example 3: Preparing a table using all women factors***

In some countries, birth histories are only collected among ever married women. In such cases, individual exposure has to be multiplied by an all-women factor that corrects for the fact that never married 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)
```

computes the table of births and exposure by 5-year age groups for the 3 years preceding, using the variable `awfactt` as the all-women factor<sup>12</sup>.

```
. tabexp
```

will provide the same result: `tabexp` checks if the `awfactt` variable exists, and if its mean is different from one, it is automatically used. A message indicating that `awfactt` is used is displayed<sup>13</sup>. Table 8 displays results for the 2008 Bangladesh survey.

Table 8. Computing births and exposure by 5-year age groups for the three years preceding the survey using all-women factors, 2008 Bangladesh DHS (computation with `tabexp`)

```
. tabexp, awf(awfactt)

All-women factor should be used. By default the variable 'awfactt' is used.
If you analyse fertility for sub-populations, use the correct all-women factor
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
(0 real changes made)
0
period  ageg  events  exposure  centry
0       15   1119.58  8869.704  105.878
0       20   1203.49  6939.909  105.878
0       25   724.387  5681.776  105.878
0       30   351.896  5039.731  105.878
0       35   145.178  4247.281  105.878
0       40    32.6876  3338.531  105.878
0       45    2.66096  2390.561  105.878
```

## 6. Poisson regression to compute fertility rates

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

Poisson regression is a generalized linear model in which the conditional distribution of the dependant 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

<sup>12</sup> The `awfactt` variable is available in DHS data sets. It should be used when indicators are computed for the whole population. Specific all women factors should be computed and used when working on other sub-populations. All women factors are available in DHS data files for analyses by place of residence, education or wealth quintile.

<sup>13</sup> Another message warning that the correct all women factor should be used when producing tables for sub-populations is also displayed.

analyze birth histories in a flexible way (Schoumaker, 2004). Births are the dependant variable, exposure is controlled with the offset, and independent variables include age groups and other types of covariates (time period, education,...).

In 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  $\mu_i$  (Winkelman et al, 1994; Trussell and Rodriguez, 1990).

$$P(Y_i = y_i | \mu_i) = \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \quad \# \quad \# \quad \text{[Eq. 1]}$$

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

$$\mu_i = \lambda_i \cdot t_i \quad \# \quad \# \quad \# \quad \# \quad \text{[Eq. 2]}$$

Taking the logarithm of this expression, it becomes:

$$\log(\mu_i) = \log(t_i) + \log(\lambda_i) \quad \# \quad \# \quad \text{[Eq. 3]}$$

The regression model consists in modeling the logarithm of rates ( $\lambda_i$ ) as a linear combination of independent variables. In `tfr2`, independent variables include a function of age and possibly additional covariates:

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

The Poisson regression that is estimated can be written as follows:

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

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).

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

### **Age-specific fertility rates and TFR**

Classical indicators of fertility (rates, TFRs) can be obtained from the regression coefficients of a Poisson regression (as well as their standard errors) in which age groups are the only independent variables. In the example below (as in `tfr2`), 5-year age groups are included in the model as a series of dummy variables.

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

$\alpha$  is the constant term,  $A_{ki}$  are dummy variables for the 6 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:

$$\lambda_i = \exp \left[ \alpha + \sum_{k=20-24}^{45-49} \beta_k \cdot A_{ki} \right] \quad [\text{Eq. 8}]$$

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 just equal to the exponential of the sum of the constant and the coefficient of the 25-29 age group.

$$\lambda_{25-29} = \exp[\alpha + \beta_{25-29}] \quad [\text{Eq. 9}]$$

The total fertility rate (15-49) is equal to 5 times the sum of age-specific fertility rates.

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

Standard errors of the rates and of the TFR can be computed using the delta method.

### **Reconstructing fertility trends by year**

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

$$\log(\mu_{it}) = \log(t_{it}) + \alpha + \sum_{k=20-24}^{45-49} \beta_k \cdot A_{kit} + \sum_{h=2}^{15} \delta_h \cdot T_{hit} \quad [\text{Eq. 11}]$$

$\lambda_{it}$  is the fertility rate at age  $i$  and period  $t$ , and is equal to the product of an age effect and a time effect. We make the assumption that the age schedule of fertility is constant. Although this does not strictly hold, simulations show that this is a reasonable assumption for relatively short periods (e.g. 15 years).

Predicting the fertility rate for a single age group (e.g. age group 25-29) for a specific year (e.g. year 5) is also direct. 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 just 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.

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

The total fertility rate (15-49) for year h is equal to 5 times the sum of age-specific fertility rates, multiplied by the exponential of the regression coefficient of the dummy variable for year h.

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

### **Multivariate analyses of fertility**

Multivariate analyses of recent fertility can be performed in the same way as analyses of fertility trends: categorical independent variables are included in the model (using dummy coding), along with dummy variables for age groups. The exponential of the regression coefficients of the independent variables are interpreted as rate ratios (ratios of TFRs) for the categories of the independent variables compared to the reference category.

#### **6.1 tfr2 to analyze birth histories**

The syntax of `tfr2` is comparable to the syntax of `tabexp`, and includes a few additional features (graphical options, saving options...). The general syntax for the `tfr2` command is:

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

`varlist` is used to include categorical covariates. Rate ratios are computed for each category of the covariates, compared to the reference category.

`pweight` allows using sampling weights. Like 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), 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`: `dates(varname)`, `wbirth(varname)`, `bvar(varlist)`, `ageg(#)`,

minage(#), maxage(#), length(#), trend(#), cy, endy(#),  
entry(varname), cluster(varname), awf(varname).

Additional options include:

- `mac` computes mean age at childbearing
- `savetable(filename)` allows saving the table of events and exposure in a stata file.
- `saverates(filename)` saves ASFRs and TFR in a Stata file.
- `savetrend(filename)` saves 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 95% confidence intervals on graphs, or saves them in tables.

## 6.2 `tfr2` to compute fertility rates and TFR

This first series of examples illustrate the use of `tfr2` to compute classical indicators of fertility (rates and TFRs) in various situations.

### **Example 1. Age-specific fertility rates and TFR for the last 3 years**

This example shows how to compute fertility rates and TFRs for the 3 years preceding the survey, as published in DHS reports. This is done directly by typing

```
. tfr2
```

This is equivalent to typing

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

Results are displayed in Table 9 (fertility rates and TFR are in the ‘Coef.’ column). These rates are equal to those published in the DHS report (Table 10, column *Total 2008*). Standard errors of rates and of the TFR are provided. These standard errors are based on the assumption of a simple random sample.

Table 9. Age-specific fertility rates and TFR for the 3 years preceding the survey, 2008 Bolivia  
DHS (computation with tfr2)

```
.tfr2

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
ASFRs and TFR
```

events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Rate_1519	.0875798	.0029478	29.71	0.000	.0818022	.0933574
Rate_2024	.1735013	.0046504	37.31	0.000	.1643867	.182616
Rate_2529	.1731727	.0046613	37.15	0.000	.1640368	.1823087
Rate_3034	.1277552	.0043746	29.20	0.000	.1191812	.1363292
Rate_3539	.0948959	.0039165	24.23	0.000	.0872196	.1025721
Rate_4044	.0427743	.0028396	15.06	0.000	.0372087	.0483399
Rate_4549	.0077979	.0015373	5.07	0.000	.0047849	.0108109
TFR	3.537385	.0492307	71.85	0.000	3.440895	3.633876

Table 10. Age-specific fertility rates and TFR for the 3 years preceding the survey, 2008 Bolivia  
DHS (source: 2008 Bolivia DHS report)

**Cuadro 4.1 Niveles de fecundidad para los tres años que precedieron la encuesta**

Tasas específicas por edad, tasas acumuladas de fecundidad y tasa bruta de natalidad para los tres años que precedieron la encuesta, por área de residencia, Bolivia 2008

Edad e indicador	Área de residencia		Total 2008	Total 2003	Total 1998
	Urbana	Rural			
<b>Edad</b>					
15-19	67	132	88	84	84
20-24	143	244	174	183	207
25-29	150	222	173	179	201
30-34	104	175	128	145	165
35-39	72	134	95	114	117
40-44	28	67	43	51	57
45-49	3	14	8	12	15
<b>Tasa</b>					
Tasa global de fecundidad <sup>1</sup>	2.8	4.9	3.5	3.8	4.2
Tasa general de fecundidad <sup>2</sup>	99	165	121	130	142
Tasa bruta de natalidad <sup>3</sup>	24	28	26	28	30

<sup>1</sup>La tasa global de fecundidad está expresada en nacimientos por mujer  
<sup>2</sup>La tasa general de fecundidad (nacimientos divididos por número de mujeres 15-44) está expresada en nacimientos por 1,000 mujeres  
<sup>3</sup>La tasa bruta de natalidad está expresada en nacimientos por 1,000 habitantes

tfr2 allows using jackknife 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)
```

Table 11. Age-specific fertility rates and TFR for the 3 years preceding the survey, standard errors computed using jackknife, 2008 Bolivia DHS (computation with tfr2)

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

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

ASFRs and TFR

events	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Rate_1519	.0875798	.0037312	23.47	0.000	.0802579 .0949017
Rate_2024	.1735013	.0056638	30.63	0.000	.1623869 .1846157
Rate_2529	.1731727	.0050979	33.97	0.000	.1631688 .1831766
Rate_3034	.1277552	.0052322	24.42	0.000	.1174879 .1380225
Rate_3539	.0948959	.0047672	19.91	0.000	.0855409 .1042508
Rate_4044	.0427743	.0036023	11.87	0.000	.0357053 .0498433
Rate_4549	.0077979	.0018346	4.25	0.000	.0041978 .011398
TFR	3.537386	.0730998	48.39	0.000	3.393939 3.680833

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. The computation of correct standard errors is a little more time consuming, but it is straightforward with tfr2.

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 3 years preceding the survey, 2008 Bolivia DHS (computation with tfr2)

```
. tfr2, mac norates notfr
```

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

ASFRs and TFR

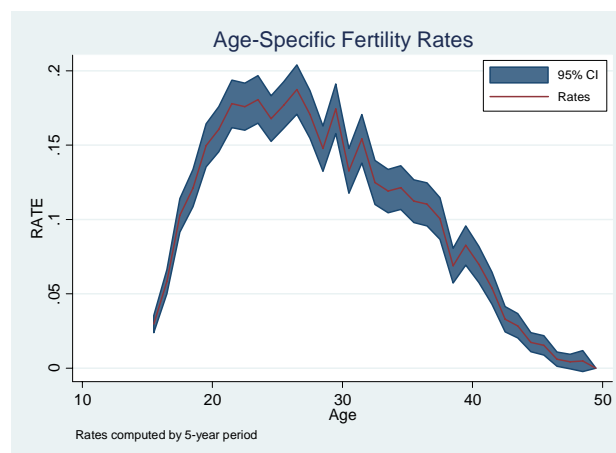
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
MAC	28.40745	.1081838	262.59	0.000	28.19542 28.61949

## Example 2. Fertility rates by single year of age for the last 5 calendar years

Like 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 5 calendar years, and to display these rates and their standard errors on a graph (Figure 3), the following command is used.

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

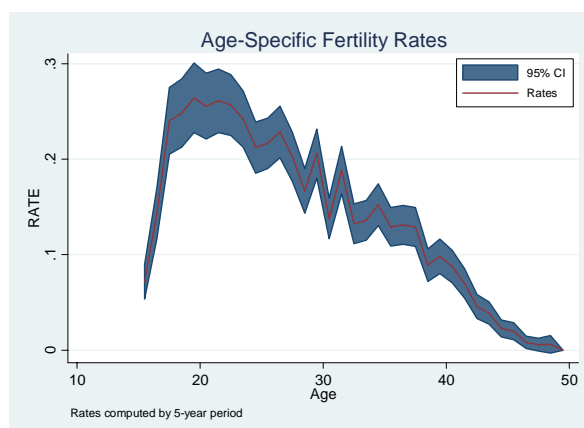
Figure 3 : Age-specific fertility rates for the 5 calendar years preceding the survey, Bolivia 2008 DHS (computation with `tfr2`)



The same graph can be drawn for sub-populations by using the appropriate condition (`if`). For example, this can be done for women with no or primary education (`v106<=1`).

```
. tfr2 if v106<=1, ageg(1) length(5) cy gr se
```

Figure 4 : Age-specific fertility rates for the 5 calendar years preceding the survey among women with no or primary education, Bolivia 2008 DHS (computation with tfr2)



### **Example 3. Using tfr2 with WFS data**

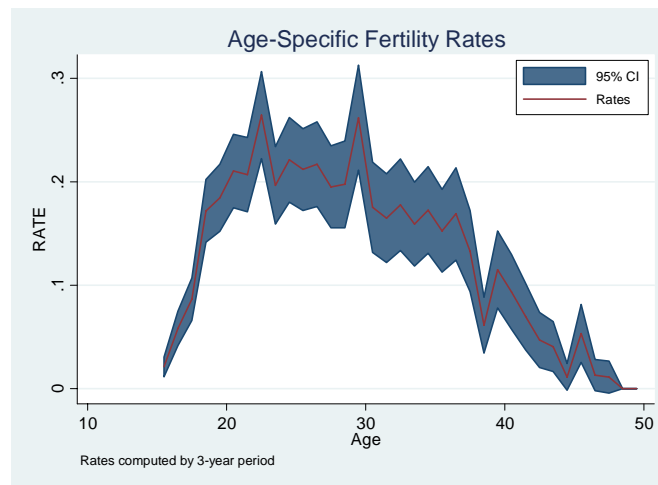
tfr2 can also be used with data from other surveys, as long as they are organized in a similar way as the DHS. For instance, WFS data can be very easily be exploited with tfr2. The example below computes age-specific fertility rates in Colombia with the 1976 WFS data available on German Rodriguez' website<sup>14</sup>, and replicates his results. The data set (a selection of variables in Stata format) is downloaded from the website<sup>15</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
```

<sup>14</sup> <http://data.princeton.edu/eco572/asfr.html>

<sup>15</sup> Contrary to this file for Colombia, WFS data are not directly available in Stata format. However, they can be downloaded from Princeton's Office of Population Research website, and easily imported in Stata.

Figure 5 : Age-specific fertility rates for the 3 years preceding the survey, Colombia 1976 WFS  
(computation with `tfr2`)



### 6.3 `tfr2` to reconstruct fertility trends

Assuming that the age pattern of fertility is fairly stable, `tfr2` can be used to reconstruct TFR (15-49) over 10 to 20 years. One of the use of this approach is to evaluate data quality. Birth histories in DHS may be affected by various types of errors, notably displacement and omissions of births (Schoumaker, 2011). The lengthy health module, which usually is restricted to births from January five years before the survey, may encourage interviewers to displace and/or omit births to avoid asking these questions. The reconstruction of TFR by calendar years 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.

#### Example 1. Reconstructing the TFR (15-49) over 15 years

We illustrate this with the 2003 Mozambique survey:

```
. use mzir41f1.dta, clear
. tfr2, len(15) trend(1) cy gt se
```

This computes and displays TFRs for the last 15 calendar years (Table 13, Figure 6). This figure shows a sudden drop in the TFR (from 7.2 to 5.3) at the start of the health module (year 1998). This clearly suggests displacements and/or omissions of births as a result of the health module.

Table 13. Total fertility rate (15-49) for the 15 calendar years preceding the survey, Mozambique  
2003 DHS (computation with tfr2)

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

ASFRs and TFR

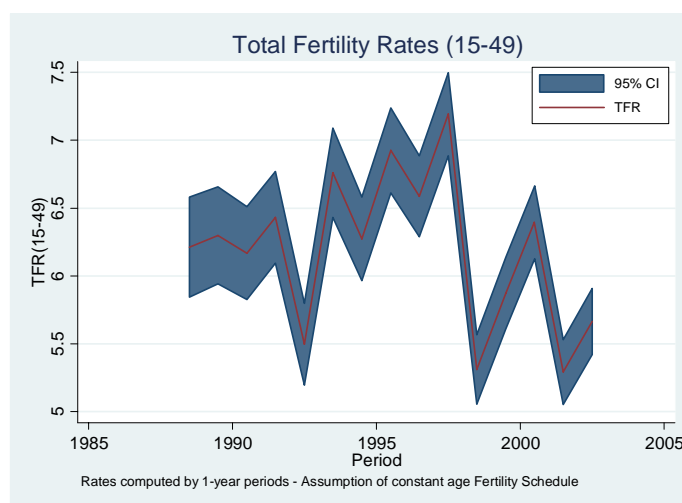
events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Rate_1519	.1815951	.0023213	78.23	0.000	.1770455	.1861447
Rate_2024	.2638362	.0030106	87.64	0.000	.2579356	.2697368
Rate_2529	.2487996	.0032528	76.49	0.000	.2424242	.255175
Rate_3034	.2148552	.0033955	63.28	0.000	.2082001	.2215103
Rate_3539	.16426	.0036163	45.42	0.000	.1571722	.1713479
Rate_4044	.0984417	.0038761	25.40	0.000	.0908446	.1060387
Rate_4549	.0539136	.0057074	9.45	0.000	.0427273	.0651
TFR	6.128507	.049305	124.30	0.000	6.031871	6.225143

events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
TFR_0	6.211032	.1880453	33.03	0.000	5.84247	6.579594
TFR_1	6.298593	.1825654	34.50	0.000	5.940771	6.656415
TFR_2	6.168181	.1739772	35.45	0.000	5.827192	6.50917
TFR_3	6.432248	.1722874	37.33	0.000	6.094571	6.769925
TFR_4	5.497554	.1542499	35.64	0.000	5.19523	5.799878
TFR_5	6.759918	.1673467	40.39	0.000	6.431924	7.087912
TFR_6	6.273411	.1564374	40.10	0.000	5.966799	6.580023
TFR_7	6.924626	.1602401	43.21	0.000	6.610561	7.238691
TFR_8	6.587153	.152109	43.31	0.000	6.289025	6.885281
TFR_9	7.191718	.1562976	46.01	0.000	6.885381	7.498056
TFR_10	5.31216	.1303093	40.77	0.000	5.056758	5.567562
TFR_11	5.872705	.1342655	43.74	0.000	5.60955	6.135861
TFR_12	6.394283	.1370163	46.67	0.000	6.125736	6.662829
TFR_13	5.292314	.1217126	43.48	0.000	5.053762	5.530866
TFR_14	5.664699	.1239034	45.72	0.000	5.421852	5.907545



Figure 6 : Total fertility rate (15-49) for the 15 calendar years preceding the survey, Mozambique 2003 DHS (computation with tfr2)

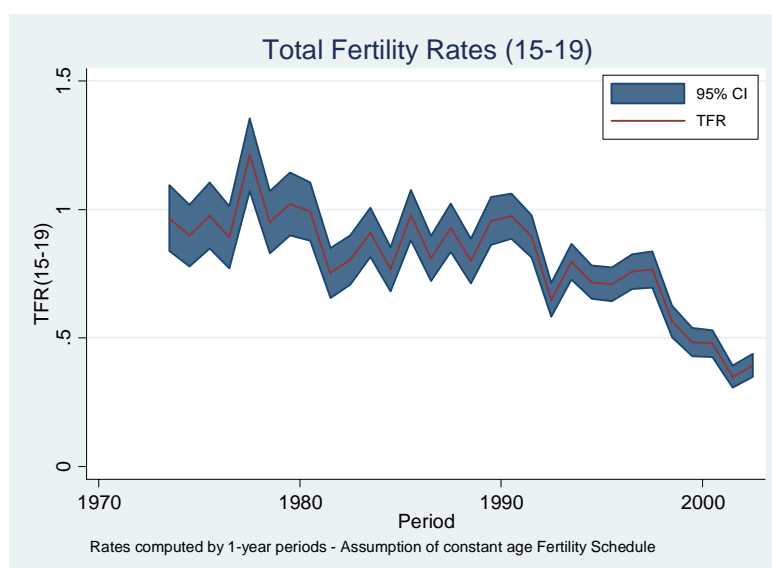


### Example 2. Reconstructing the adolescent fertility over 25 years

Trends can also be reconstructed for a specific age-group; this allows reconstructing trends over long periods for young age groups. For instance, the command below is used to reconstruct fertility rates between 15 and 19 (partial total fertility rate) over the last 30 calendar years with the 2011 Ethiopia DHS. Overall, there is a downward trend over the last 10 years. Like in the previous example, some of this decline may reflect data quality issues (Figure 7).

```
. use etir60f1.dta
. tfr2, len(30) trend(1) minage(15) maxage(19) gt se cy
```

Figure 7 : Total fertility rate (15-19) for the 25 calendar years preceding the survey, Ethiopia 2011 DHS (computation with tfr2) (note : dates are recorded in Ethiopian calendar).



### Example 3. Comparing reconstructed fertility trends from successive surveys

This tool can also be used to compare reconstructed fertility trends from successive surveys. This can be used to reconstruct long term fertility trends, as well as to evaluate data quality. The example below compares fertility trends from two surveys in Mozambique. Using `tfr2`, this is done with only a few lines of syntax: `tfr2` is used for each survey, results from `tfr2` for the various surveys are appended, and a graph of fertility trends from successive surveys is drawn.

```
.cd "c:\DHS\  
.local listDHS mzir31fl mzir41fl  
  
.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_mzir31fl date, sort) (line TFR_mzir41fl date, sort)
```

Figure 8 : Total fertility rate (15-19) for the 15 calendar years preceding each survey, Mozambique 1997 and 2003 DHS (computation with `tfr2`)

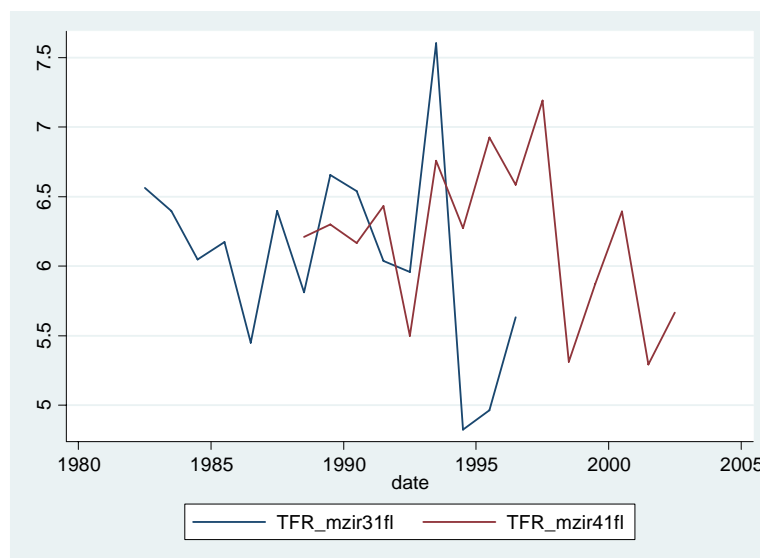


Figure 8 shows the results for the 1997 and 2003 Mozambique surveys, clearly illustrating the discrepancy between recent fertility in the 1997 survey, and fertility at the same time estimated from the 2003 survey.

## 6.4 tfr2 to compute rate ratios

Finally, `tfr2` can be used to compute rate ratios of categorical variables. If the age pattern of fertility is fairly similar across categories, rate ratios can be interpreted as ratios of TFRs. One or several covariates can be included in `tfr2`.

### Example 1. Fertility differentials by education

The table below shows the results for educational differentials in Bolivia: `v106` is used as a covariate in this way:

```
. tfr2 v106
```

Table 14. Fertility rates and rate ratios by level of education for the 3 years preceding the survey, Bolivia 2008 DHS (computation with `tfr2`)

```
. tfr2 v106
Explanatory variables :i.v106
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
ASFRs and TFR for reference category
```

events	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Rate_1519	.1659365	.0113884	14.57	0.000	.1436156 .1882574
Rate_2024	.3320382	.0212186	15.65	0.000	.2904504 .373626
Rate_2529	.3036214	.0187965	16.15	0.000	.266781 .3404619
Rate_3034	.2121667	.0136634	15.53	0.000	.1853869 .2389466
Rate_3539	.1504648	.0101269	14.86	0.000	.1306163 .1703132
Rate_4044	.0647875	.0053812	12.04	0.000	.0542405 .0753345
Rate_4549	.011172	.0022602	4.94	0.000	.0067421 .0156018
TFR	6.200936	.3566418	17.39	0.000	5.501931 6.899941

```
-----
Rate ratios of explanatory variables
-----
Variable | Rate_ratios
-----+-----
_Iv106_1 | .75309453***
_Iv106_2 | .48176432***
_Iv106_3 | .29574878***
-----
legend: * p<.1; ** p<.05; *** p<.01
```

Age-specific fertility rates and TFRs are computed for the reference category (uneducated women), and rate ratios are displayed for the other categories of education. This shows that fertility among the highly educated (`v106=3`) is equal to 30% of the fertility among uneducated women (`v106=0`). 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 using the `by` prefix.

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

This offers a quick way to compute fertility rates by level of education without making the assumption of proportionality of rates. Results (not shown) indicate that the TFR among the highly educated is equal to 1.88, and is 6.11 among women without education. The ratio is 0.307, very close to the ratio estimated with the assumption of proportionality of rates.

### **Example 2. Multivariate model of recent fertility**

A major interest of regression is of course the possibility of including several covariates. The following command evaluates the net effect of education, controlling for standard of living quintiles (`v190`) and place of residence (`v025`). The TFR of the highly educated is half that of the uneducated, controlling for place of residence and standard of living. In other words, the net effect of education on recent fertility is smaller when controlling for these two variables, but it remains very strong.

Table 15. Fertility rates and rate ratios by level of education, standard of living and place of residence for the 3 years preceding the survey, Bolivia 2008 DHS (computation with tfr2)

```

. tfr2 v106 v190 v025
Explanatory variables :i.v106 i.v190 i.v025
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
ASFRs and TFR for reference category
-----
      events |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
Rate_1519 |   .2092753   .0171795    12.18   0.000   .1756041   .2429465
Rate_2024 |   .4187188   .0325912    12.85   0.000   .3548411   .4825964
Rate_2529 |   .3958473   .0302568    13.08   0.000   .3365449   .4551496
Rate_3034 |   .2830199    .0223      12.69   0.000   .2393127   .3267271
Rate_3539 |   .2026897    .016451    12.32   0.000   .1704464   .234933
Rate_4044 |   .0889283    .0084194    10.56   0.000   .0724266   .10543
Rate_4549 |   .0154882    .0032149     4.82   0.000   .009187    .0217893
      TFR |   8.069837   .5885014    13.71   0.000   6.916396   9.223279
-----
Rate ratios of explanatory variables
-----
      Variable |      Rate_ratios
-----+-----
   _Iv106_1 |   .86432752**
   _Iv106_2 |   .70679347***
   _Iv106_3 |   .52205846***
   _Iv190_2 |   .74646236***
   _Iv190_3 |   .6072534***
   _Iv190_4 |   .4701975***
   _Iv190_5 |   .37893003***
   _Iv025_2 |   .88987723***
-----
legend: * p<.1; ** p<.05; *** p<.01

```

## 7. 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 valorizing the large amount of fertility data collected since the 1970s. It may notably be used to reconstruct fertility trends over long periods, to facilitate the evaluation of data quality, to compute fertility rates for specific populations (not published in survey reports...).

tfr2 covers a limited set of analyses, mainly descriptive. Other important indicators (parity progression ratios, parity-specific fertility rates,...) are not currently implemented in tfr2, but similar tools could complement this one.

## 8. References

- Moultrie TA. (2011). *Direct estimation of fertility from survey data containing birth histories*. In Moultrie TA, RE Dorrington, AG Hill, KH Hill, IM Timæus and B Zaba (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*, Academic Press, San Diego, 305 p.
- 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*. Available at <http://data.princeton.edu/eco572/>.
- Rodríguez, G. (2007). *Lecture Notes on Generalized Linear Models*. Available at <http://data.princeton.edu/wws509/notes/>.
- Rutstein S.O, Rojas G. (2006), *Guide to DHS Statistics*, Demographic and Health Surveys, ORC Macro, Calverton.
- Schoumaker Bruno (2004), "A person-period approach to analyzing birth histories", *Population-E*, vol.59, pp. 689-702.
- Schoumaker, B. (2011). "Omissions of recent births in DHS birth histories. Measurement and determinants", *Paper presented at the PAA Meeting*, Washington DC.
- Trussell J. and Rodriguez G. (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.
- Winkelmann R. and Zimmermann (1994), "Count data models for demographic data", *Mathematical Population Studies*, 4(3), pp. 205-221.