

Cohort Fertility and Education Database

Methods Protocol

version 2.12.2014

Kryštof Zeman, Zuzanna Brzozowska, Tomáš Sobotka, Éva Beaujouan, Anna Matysiak

1. Introduction.....	1
2. Database structure & use in a nutshell.....	1
3. File system and labelling.....	2
4. Coding of education.....	3
5. Computing indicators.....	5
6. Terms of use.....	7
7. Acknowledgments.....	8

1. Introduction

The Cohort Fertility and Education (CFE) database provides a free access to aggregate data on completed cohort fertility and parity distribution by level of education. The data come from censuses and large-scale surveys conducted in the early 2000s in countries with generally high levels of education and relatively low fertility. As the database covers women and men past prime reproductive years who have completed or almost completed their family building, the data focus on people aged between 40 and 80, which corresponds to birth cohorts between the 1920s and 1960. In addition, for some countries we have collected data coming from earlier censuses, and thus encompassing women born in the late 19th century. We also plan to add data from censuses and surveys conducted in the 2010s.

2. Database structure & use in a nutshell

The database is available online at www.cfe-database.org and is freely accessible without registration. The data are structured by country and survey (one has to choose a country first and only then a survey). Having selected the survey of interest, the users can access a range of indicators, which are generated by the database using the formulas specified in section 5 of this protocol. Specifically, the database features:

- the completed fertility rate (CFR), specified by birth order (ranging from birth order 1 to 8+),
- the relative distribution of respondents by the number of children ever born (including childless respondents),
- parity progression ratios (PPR).

All indicators are specified by the highest achieved level of education at the time of the given census or survey and, when available, by sex and country of birth or citizenship. The source data are available by clicking the *Get dataset as CSV* button. They contain the absolute number of women (and, if available, men) by level of education, birth cohort, the number of children ever born and, if available, country of birth or citizenship (distinguishing between natives and foreigners). These source data are not standardised and include the unknown categories (unknown number of children, birth cohort, level of education and country of birth). For each country a brief description of each survey or census as well as the country's education system is provided in the Documentation file.

The CFE database supports enhanced possibilities to explore and visualise the data. Users can download a CSV file with the source data, but they can also compile and download tables defined by themselves (with a self-specified aggregation level, self-selected indicators and certain cohorts/education categories filtered out). In addition, the database allows users to generate interactive graphs with selected indicators for each country and survey, which provide first-hand visualisation of inter-cohort fertility trends, patterns and differences.

3. File system and labelling

The data are stored as comma-separated values (CSV) files, which you can easily read using the following command in the statistical package R:

```
read.csv('file_name.csv', header = TRUE)
```

If you open the file with Calc within the LibreOffice package, you should use the option 'separated with comma'; if you open it with Microsoft Office Excel and your operating system uses a delimiter other than comma (e.g. tabulation or semi-colon), mark the column with entries, go to *Data* on the ribbon, click *Text to Columns*, choose *Delimited* as the file type that best describes your data, press *Next*, tick off *Comma* as a delimiter, press *Next* and then *Finish*. [Here](#) you can find a PowerPoint presentation showing the procedure.

The first line contains a header, with the following categories listed:

- `cohort` - respondent's birth cohort either in single years or in five-year cohorts;
- `edu` - education coded as described in the next section;
- `sex` - F for women, M for men (when included);
- `origin` - Total for all places of birth combined; when available the data are specified by two broad categories of the place of birth/citizenship, further explained in country documentation (e.g., born in/outside Austria or Swiss/foreign);
- `women_total` - the total number of women (sex F) /men (sex M);
- `children_total` - the total number of children ever born to all respondents in a given cohort and educational group (and of a given origin status);
- `parity_0` - the number of childless women/men;
- `parity_i` ($i=1, 2, \dots, i_{\max}$) - the number of respondents with i children; the maximum-parity group i_{\max} differs across surveys and includes all respondents with a higher number of children $i+$;
- `parity_unknown` - number of respondents for whom the number of children is not known (not for all countries);
- `CFR`, `CFRi` ($i=1, 2, \dots, i_{\max}$) - the completed fertility rate, total and by birth order;
- `SHAREi` ($i=0, 1, 2, \dots, i_{\max}$) - proportion of women/men by parity i ;
- `PPRi-1i` ($i=1, 2, \dots, i_{\max}$) - parity progression ratio (see section 5 for more details).

For the number of persons with unknown or not specified characteristics (cohort, education, country of origin etc.) the label `Unknown` is used. The database lists totals by cohort and education, and, if available, by sex and place of origin. Total can always be computed as a sum of all specified cases.

The source data are extracted directly from the survey or census records, with very few, if any, computations (only with translating the educational categories into the ISCED97 levels).

4. Coding of education

The highest attained education (the `edu` column in the CSV files) is coded according to the International Standard Classification of Education 1997 (ISCED97). There are 7 broad levels of education, starting from 0 and ranging up to 6, further distinguishable into C/B/A sublevels (see Table 1). In the documentation files for each country there is a table translating the original educational categories (as given in the questionnaire) to the ISCED97 levels.

Table 1 The 1997 ISCED levels and their short description

ISCED level	Description
0	Early childhood education; no education
1	Primary education
2C-2A	Lower-secondary education (Second stage of basic education)
3C	Upper-secondary education: programmes designed to lead directly to labour market, not to ISCED 5A or 5B (e.g. Apprenticeship, Secondary technical school)
3B	Upper-secondary education: programmes designed to provide direct access to ISCED 5B (e.g. Vocational training)
3A	Upper-secondary education: programmes designed to provide direct access to ISCED 5A (e.g. Higher general secondary/Higher technical secondary school)
4B	Post-secondary non-tertiary education not giving access to level 5 (primarily designed for direct labour market entry, e.g. Schools for medical services, Schools for nursing)
4A	Post-secondary non-tertiary education programmes that prepare for entry to ISCED 5 (e.g. follow-up courses, language schools)
5B	First stage of tertiary education (not leading directly to an advanced research qualification; e.g. Higher technical school, polytechnic)
5A	Bachelor's, Master's or equivalent level
6	Doctoral or equivalent level

For detailed information on ISCED 1997 visit <http://www.uis.unesco.org/Library/Documents/isced97-en.pdf>

Users can aggregate the educational levels manually or they can use the three- or the four-level EURREP classification. The four-level classification is specified as follows:

Level 1: lowest compulsory education; encompasses ISCED levels 0, 1 and 2; also labelled basic or primary education;

Level 2: usually ISCED3C¹;

Level 3: upper-secondary and post-secondary education; includes ISCED3A, ISCED3B and ISCED4 levels;

Level 4: ISCED5 and ISCED6 levels; tertiary or university education.

The three-level EURREP classification merges all ISCED3 (3A, 3B and 3C) and ISCED4 levels, with the first and the last levels remaining the same as in the four-level EURREP classification. In some cases (e.g. Switzerland), the Medium-low and Medium-high cannot be distinguished, and only the three-level EURREP classification is specified (Level 1, 2, 3).

¹ In some countries a different categorisation is used to reflect the underlying differences in educational systems and to achieve maximum comparability. For example for Austria ISCED3B (apprenticeship and vocational training) is included into the Level 2, while for the Czech Republic ISCED3B is included into the Level 3.

5. Computing indicators

This section specifies all the indicators provided in the CFE database and the equations used to compute them.

- *CFR* is the completed fertility rate;
- *SHARE* is the share of women/men by parity *i*, including the childless and ranging up to the maximum open-ended group *i_max*, which also includes all women/men with a higher number of children;
- *PPR* is the parity progression ratio.

Generally, the highest parity/birth order group considered in the *CFR* and *SHARE* indicators is an open-ended group 8+, if not explicitly stated otherwise (for instance, when the original data contain a narrower set of available categories). For parity progression ratios, the highest progression rate computed is from 6th to 7th child (*PPR67*), unless specified otherwise.

The CFE database uses the following notations:

c...birth cohort

e...educational level

o...country of origin/citizenship

s...sex

i...birth order of child or parity of the mother

i_max...highest birth order considered (used as open interval *i_max+*)

parity_i...number of respondents with *i* children

women_total (*men_total*)...total number of respondents

children_total...total number of children

SHARE_i...proportion of respondents with *i* children

CFR_i...completed fertility rate for birth order *i*

PPR_{i-1,i}...parity progression ratio; conditional probability of having another (*ith*) child among respondents achieving parity *i-1*

For each cohort, educational level, origin and sex (not shown in the formulas below), the following indicators are computed:

Completed fertility rate by birth order:

$$CFR_i = \frac{\sum_i parity_i}{women_total} \quad (1)$$

Completed fertility rate for all women/men:

$$CFR = \frac{children_total}{women_total} \quad (2)$$

The share of women/men by the number of children ever born (summing up to 1):

$$SHARE_i = \frac{parity_i}{women_total} \quad (3)$$

The parity progression ratio to first birth and for higher-order transitions:

$$PPR_{0,1} = CFR_1 \quad (4)$$

$$PPR_{i-1,i} = \frac{CFR_i}{CFR_{i-1}}, \text{ for } i > 1 \quad (5)$$

Indicators for the highest birth order/parity i_{max} are computed as open intervals in the case of CFR_{8+} and w_{8+} and as a single-parity progression rate in the case of $PPR_{7,8}$ (if CFR_8 is available):

$$CFR_{8+} = CFR - \sum_{i=1}^7 CFR_i \quad (6)$$

$$SHARE_{8+} = \frac{parity_8p}{women_total} \quad (7)$$

$$PPR_{7,8} = \frac{CFR_8}{CFR_7} \quad (8)$$

In most cases respondents with unknown number of children are disregarded in computations of all fertility indicators:

$$women_total \sim women_total - parity_unknown \quad (9)$$

In practice this gives the same result as if these respondents were proportionally redistributed according to the relative parity distribution of respondents for whom the number of children ever born has been recorded. In special cases, where there is a good reason to presume that respondents with unknown number of children are in fact childless (e.g. in the censuses for the Czech Republic and Slovakia), the number of respondents with unknown number of children has been added to the number of childless respondents:

$$parity_0 \sim parity_0 + parity_unknown \quad (10)$$

These country or survey-specific computations are specified in country documentation files.

6. Terms of use

The data in the Cohort Fertility and Education database are provided free of charge to all interested users. We kindly ask you to comply with the following requests and restrictions:

1. CFE data are not to be used for commercial purposes. The user may copy and retain data solely for scholarly, educational or research purposes or for personal use. The original tabulations of absolute number of respondents by cohort, education and number of children ever born should not be used for commercial gain or re-published in any form without the explicit permission of the data owners.
2. In all published work and presentations, please acknowledge the Cohort Fertility and Education database as either the source or the intermediary of the data. When downloading the data, you should note the date for future reference.

When using the database, please cite the full name of the database (Cohort Fertility and Education database, CFE database) and refer to:

Zeman, K., Z. Brzozowska, T. Sobotka, E. Beaujouan, A. Matysiak. 2014. *Cohort Fertility and Education database. Methods Protocol*. Available at www.eurrep.org (accessed on [date]).

3. Please do not pass your copy of these data to other users. Rather refer them to the www.eurrep.org or www.cfe-database.org website, where they may download the same data for themselves. Since these data are updated on a regular basis (including corrections if needed), this practice helps to prevent the circulation of multiple outdated or incorrect versions. It also ensures that each user has full access to complete information about the data, methodology, citation procedures, etc.
4. All data contained in the Cohort Fertility and Education database have been carefully checked to avoid errors. However, the database accepts no responsibility for any harm suffered by the user as a result of using these data, even if such harm results from errors on our part. We ask you kindly to inform us of any errors that you may identify, so that they can be corrected in a timely fashion.

7. Acknowledgments

The Cohort Fertility and Education database is a part of the EURREP project, which is funded by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement n° 284238.

We are grateful for the support we have received from individuals and institutions providing the data or agreeing with their use. You find the list of our collaborators in the ABOUT section of the database, under ACKNOWLEDGMENTS.