Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
http://www.demogr.mpg.de

MPIDR TECHNICAL REPORT 2010-007
NOVEMBER 2010

An "R" package for the production of cohort fertility tables

Edward Nash (nash@demogr.mpg.de)
Aiva Jasilioniene (jasilioniene@demogr.mpg.de)
Evgeny Andreev (andreev@demogr.mpg.de)

This technical report has been approved for release by: Vladimir Shkolnikov (shkolnikov@demogr.mpg.de), Head of the Laboratory of Demographic Data.

Third draft (r27), last modified 22/10/2010 10:47:00 by Edward Nash

An “R” package for the production of cohort fertility tables

by Edward Nash, Aiva Jasilioniene, and Evgeny Andreev

Abstract

The construction of life tables is standard practice in mortality research, and methods are widely standardised and used. The use of fertility tables is less widespread, and there is little agreement on the methodology of their construction. However, production of standardised fertility tables may facilitate new fertility research and make a significant contribution to comparative fertility analysis. As part of the Human Fertility Database project, a methodology for the construction of cohort fertility tables has been developed and implemented as a package for the free statistical computing language and environment “R”. This Technical Report introduces the software package for calculation of cohort fertility tables.

Keywords: cohort fertility table, Human Fertility Database, R

Background

In the Human Fertility Database (HFD)[1], a joint project of the Max Planck Institute for Demographic Research (MPIDR) and the Vienna Institute of Demography (VID), based at the MPIDR, fertility tables are produced by age and parity for both birth cohorts and periods. This report is devoted to cohort fertility tables. Period fertility tables may also be produced, but are in many respects more complicated, and a package for their production will be described in a further Technical Report.

Production of cohort fertility tables

The HFD considers the reproductive span between age 12- ($x_{\min}$) to age 55+ ($x_{\max}$) and birth data for orders 1…5+, allowing the consideration of parities 0…4+. Cohort fertility tables are produced for cohorts observed from age 15 or lower up to age 25 or higher. To enable the HFD methodology to be applied flexibly by other users who may wish to consider a greater or lesser age or parity range, the HFD Methods Protocol has been generalised in terms of age range and birth order in the description below: this generalised form is supported by the “R” package, although the values for many parameters default to those used by the HFD. The notation i_b^+ is used here for the highest (open-interval) birth order and i_l^+ for the highest (open-interval) parity, where $i_l^+ = i_b^+ - 1$.

The cohort fertility table includes a number of measures for each age and parity. These are summarised in Table 1. The first five measures are included once per order/parity, with $chi(x)$ only once: a complete cohort fertility table will therefore contain $1 + 5 \cdot i_b^+$ columns of indicators. Typically these columns are arranged by order/parity and then indicator ($b_1(x)$, $l_0(x)$, …, $b_2(x)$, $l_1(x)$, … $chi(x)$).

---

[1] http://www.humanfertility.org

Table 1. Summary of measures included in the HFD cohort fertility tables

Notation	*Range*	*Description*
$b_i(x)$	$i=1\ldots i_b^+$	Table number of births of order i at age interval $\left[x, x+1\right)$
$l_{i-1}(x)$	$i=1\ldots i_b^+$	Table population by parity i-1 at age x
$m_i(x)$	$i=1\ldots i_b^+$	Age- and parity- specific fertility rates[2] in age interval $\left[x, x+1\right)$
$q_i(x)$	$i=1\ldots i_b^+$	Conditional probability of giving i^{th} birth in age interval $\left[x, x+1\right)$
$Sb_i(x)$	$i=1\ldots i_b^+$	Cumulative births of order i by age x
$chi(x)$	i_l^+	Average number of children born by age x to women in the highest parity category i_l^+

The cohort fertility table is computed based on the schedule of unconditional age-specific fertility rates[3] by birth order, calculated by age and cohort (horizontal parallelograms): $f_i(x,c)$, usually calculated according to equation 1.

$$f_i(x,c) = \frac{B_i(x,t,t-x) + B_i(x,t+1,t-x)}{E(x,c)}, \tag{1}$$

where $B(x,t,c)$ is birth count by age, year and cohort, and $E(x,c)$ is population exposures for the respective cohort at age x (horizontal Lexis parallelogaram, see details in the HFD Methods Protocol). These data are available for each population in the HFD as file `XXXasfrVHbo.txt`, where `XXX` is the population code (e.g. `RUS` for Russian Federation).

The cohort fertility table is constructed using a standardised table population (radix) – for the HFD this is 10,000 women. Since birth is considered a repeatable event, the radix remains constant across all ages whilst individuals move towards higher parities according to the given birth rates. $b_i(x)$ may thus be calculated using equation 2:

$$b_i(x) = radix \cdot f_i(x,c) \tag{2}$$

Initially, at $x_{\min}$, the entire table population is in parity 0 and progresses to higher parities at higher ages according to the schedule of births $b_i(x)$. The population by age and parity $l_i(x)$ is therefore calculated using equations 3 – 7.

$$l_0(x_{\min}) = radix \tag{3}$$

$$l_i(x_{\min}) = 0, \text{ for } i = 1\ldots i_l^+ \tag{4}$$

$$l_0(x) = l_0(x-1) - b_1(x-1), \text{ for } x > x_{\min} \tag{5}$$

---

[2] Also called conditional rates, occurrence-exposure rates, rates of the first kind, intensities, hazard rates or risks (Bongaarts and Feeney, 2006; Kohler and Ortega, 2002; Wunsch, 2006).

[3] Also called incidence rates, rates of the second kind, frequencies, densities or reduced rates (Bongaarts and Feeney, 2006; Kohler and Ortega, 2002; Wunsch, 2006).

$$l_i(x) = l_i(x-1) + b_i(x-1) - b_{i+1}(x-1)\text{, for } x > x_{\min} \text{ and } i = 1 \ldots i_l^+ - 1 \qquad (6)$$

$$l_{i+}(x) = l_{i+}(x-1) + b_i(x-1)\text{, for } x > x_{\min} \text{ and } i = i_l^+ \qquad (7)$$

At very young ages and high birth orders, the numbers of births by birth order $b_i(x)$ are very low and strongly fluctuating. In order to avoid negative values of $l_i(x)$, corrections are made during the calculation of the table population by age and parity to the table births parity calculated using equation 2 as shown in equations 8 – 10.

If $l_{i-1}(x) < 0$ then $l_{i-1}(x) = 0$ (8)

If $l_{i-1}(x) = 0$ then $b_i(x) = 0$ [4] (9)

If $l_{i-1}(x) < b_i(x)$ then $b_i(x) = l_{i-1}(x)$ (10)

The age- and parity-specific fertility rates $m_i(x)$ may then be calculated by relating births of order i at age x to person-years lived by the table population at this age at parity $i-1$ as shown in equations 11 – 13. In order to estimate these person years, $a(x)$, the average share of the age interval $[x, x+1)$ lived before giving birth to a child is needed. For the HFD, it is assumed that all $a(x)$ values are equal to 0.5 for any completed age x and birth order i. In the “R” package, $a(x)$ may additionally be specified by cohort, cohort and age or cohort, age and parity.

Note that for age-order/parity combinations for which $l_{i-1}(x)$ or $b_i(x)$ was corrected then it is assumed that $m_i(x)$ is infeasible and should therefore not be calculated[5].

$$m_i(x) = \frac{b_i(x)}{l_{i-1}(x) - a(x) \cdot b_i(x)}\text{, for } i = 1 \qquad (11)$$

$$m_i(x) = \frac{b_i(x)}{l_{i-1}(x) - a(x) \cdot b_i(x) + (1 - a(x)) \cdot b_{i-1}(x)}\text{, for } i = 2 \ldots i_b^+ - 1 \qquad (12)$$

$$m_i(x) = \frac{b_i(x)}{l_{i-1}(x) + (1 - a(x)) \cdot b_{i-1}(x)}\text{, for } i = i_b^+ \qquad (13)$$

The conditional probability of giving i^{th} birth at age x, $q_i(x)$, relates the births of order i at age x to the population at parity $i-1$ at that age, giving equation 14. Again, for age-order/parity combinations for which $l_{i-1}(x)$ or $b_i(x)$ was corrected then it is assumed that $q_i(x)$ is infeasible and should therefore not be calculated

$$q_i(x) = \frac{b_i(x)}{l_{i-1}(x)}\text{, for } i = 1 \ldots i_b^+ \qquad (14)$$

Cumulative births of order i by age x are simply the sum of births of order i from $x_{\min}$ up to $x-1$ (equation 15).

---

[4] Note that this implies that at age $x_{\min}$, there are no births of order >1.

[5] Such infeasible or missing values are represented in the standard HFD output with a single period (.); we represent then internally in R using `NA`.

$$Sb_i(x) = \sum_{z=x_{\min}}^{x-1} b_i(z) \qquad (15)$$

Finally, the average number of children born by age x to women in the highest parity category i_l^+ may be calculated based on the population at the highest parity i_l^+ and the cumulative births at the highest birth order i_b^+ using equation 16.

$$chi(x) = \frac{i_l^+ \cdot l_{i_l^+}(x) + \sum_{z=x_{\min}}^{x-1} b_{i_b^+}(z)}{l_{i_l^+}(x)} \qquad (16)$$

The "R" package `hfdCohortFertilityTable`

"R" (R Development Core Team, 2010) is a language and free software system for statistical computing and graphics. The extensive core functionality may be extended using packages, which group together related functions and their documentation. R therefore provides an ideal environment for demographic data processing such as the production of life tables and a sound infrastructure for distributing software enabling such calculations to be used by others. All calculations for the Human Fertility Database are programmed in R; a number of the functions used which may be of more general interest are being made available as packages, of which `hfdCohortFertilityTable` is the first to be made publically available.

R is usually operated in a command-line environment with commands entered by the user at the "R prompt". In the following sections, input at the R prompt is shown in **`> bold Roman type`**, with output from R shown in *`oblique type`*.

Contents

`hfdCohortFertilityTable` contains seven individual calculation functions (see Table 2); one for each of the six indicators in the cohort fertility table plus one to produce the whole table by linking together the other six functions. Additionally, a set of age-specific fertility rate data (`someAsfr`) is included which serves in the function examples and may be used to experiment with the package. Furthermore, a utility function (`array2dataframe.cft`) is provided for reformatting the 3D array (Cohort × Age × Order) outputs from individual functions to a data frame analogous to the usual HFD ouptut formats. A description of these formats is provided in the next section.

More detailed descriptions of the parameters available for each function are provided in Table 3 and Table 4, as well as in the online package documentation which may be accessed by entering **`?hfdCohortFertilityTable`** at the R prompt.

Table 2. Calculation functions included in package `hfdCohortFertilityTable`

Function name	*Purpose*	*Main input data format*	*Output data format*
`calculate.cft`	Calculation of whole Cohort Fertility Table	– File or file connection (e.g. URL) in HFD standard format for XXXasfrVHbo.txt – Data frame with columns as for XXXasfrVHbo.txt – 3D array of age-specific fertility rates with named dimensions Cohort, Age and Order	– Data frame in HFD standard format for XXXcft.txt – 3D array of indicators with named dimensions Cohort, Age and Order – List containing each indicator in a separate array by Cohort, Age and (except for $chi(x)$) Order.
`calculate.b.cft`	Calculation of table births by cohort, age and birth order	– Data frame with columns as for XXXasfrVHbo.txt – 3D array of age-specific fertility rates with dimensions Cohort, Age and Order	– 3D array of births by Cohort, Age and Order with named dimensions

`calculate.l.cft`	Calculation of table population by cohort, age and parity	– 3D array of births by Cohort, Age and Order with named dimensions	– List containing
			  – 3D array of population by Cohort, Age and Parity with named dimensions
			  – 3D array of cells which were corrected to avoid implausible values

`calculate.m.cft`	Calculation of conditional fertility rates by cohort, age and birth order	– 3D arrays of births and population by parity by Cohort, Age and Order/Parity with named dimensions	– 3D array of conditional fertility rates by Cohort, Age and Order with named dimensions
`calculate.q.cft`	Calculation of conditional probabilities by cohort, age and birth order	– 3D arrays of births and population by parity by Cohort, Age and Order/Parity with named dimensions	– 3D array of conditional probabilities by Cohort, Age and Order with named dimensions
`calculate.Sb`	Calculation of cumulative births by cohort, age and birth order	– 3D array of births by Cohort, Age and Order/Parity with named dimensions	– 3D arrays of cumulative births by Cohort, Age and Order with named dimensions
`calculate.chi.cft`	Calculation of average number of children born to women in the highest parity category by cohort and age	– 3D arrays of births and population by parity by Cohort, Age and Order/Parity with named dimensions	– 2D arrays of average number of children born to women in the highest parity category by Cohrt and Age with named dimensions

Table 3. Parameters available for each function

Function name	*Parameters*
`calculate.cft`	`asfr, observeTo, observeFrom, radix, a, returnType, applyCorrections, ...`
`calculate.b.cft`	`asfr, observeTo, observeFrom, radix`
`calculate.l.cft`	`b, l, radix, applyCorrections`
`calculate.m.cft`	`b, l, a`
`calculate.q.cft`	`b, l`
`calculate.Sb`	`b`
`calculate.chi.cft`	`l, b`
`array2dataframe.cft`	`myData`

Table 4. Description of individual parameters to functions

Parameter name	*Default value*	*Description*
`asfr`	–	Age-specific data rates by Cohort and Age as a data frame, 3D array or (for `calculate.cft`) path to a file
`observeTo`	25	The minimum age to which cohorts must be observed (numeric)
`observeFrom`	15	The maximum age from which cohorts must be observed (numeric)
`radix`	10000	The size of the table population (numeric)
`a`	0.5	The average share of the age interval lived before giving birth (numeric or 2D or 3D array)
`returnType`	`"data.frame"`	How the CFT should be returned, one of "`data.frame`", "`array`" (for all indicators in a single 3D array) or "`list`" (for a list containing each indicator in a separate array)
`applyCorrections`	`TRUE`	Whether corrections as described in equations 8 – 10 should be applied
…	–	further arguments passed to `read.table` when `asfr` is to be read from file
`b`	–	A 3D array of table births with named dimensions Cohort, Age and Order
`l`	–	A 3D array of table populations by parity with named dimensions Cohort, Age and Parity. For `calculate.l` this may be used to supply initial parity distributions for cohorts not observed from the required minimum age in order to produce a “left-censored” CFT.
`myData`	–	A 2D or 3D array with named dimensions containing one of the indicators produced by the functions in this package.

Most of the default values for function parameters correspond to those used for the Human Fertility Database, but it is possible for users to specify the majority of parameters relevant to the processing, such as the age range through which cohorts must be observed and the size of the radix. Additionally, there is no restriction on the range of ages or birth orders which can be processed: the ranges from the input data are used, and so fertility tables may be produced for higher birth orders than 5, or for a lower range (e.g. to only 4+), or for a wider or narrower range of ages than that featured in the HFD.

Data formats and structures

For convenience, the package works with standard R data frames and arrays, meaning that conversions of data to/from specialised classes are avoided.

Table / data frame input format

The standard input to the package is a dataset of unconditional age- and order-specific fertility rates by horizontal Lexis parallelograms (Cohort and Age), e.g. the output files provided under the name `XXXasfrVHbo.txt` for each HFD population with code `XXX` (e.g. `USAasfrVHbo.txt` for United States). These files contain the columns Cohort, Age, ASFR, ASFR1, ASFR2, ASFR3, ASFR4, ASFR5p, where ASFR is the total age-specific fertility rate and columns ASFR1, …, ASFR5p the order-specific rates. Age may contain subscripts – and + to indicate open age intervals, but these have no effect on the calculation of the CFT.

Data rows are conventionally ordered by Cohort and then Age, e.g.:

```
> head(someASFR)

  Cohort Age ASFR ASFR1 ASFR2 ASFR3 ASFR4 ASFR5p
1   1904 12-   NA    NA    NA    NA    NA     NA
2   1904  13   NA    NA    NA    NA    NA     NA
3   1904  14   NA    NA    NA    NA    NA     NA
4   1904  15   NA    NA    NA    NA    NA     NA
5   1904  16   NA    NA    NA    NA    NA     NA
6   1904  17   NA    NA    NA    NA    NA     NA
```

```
> tail(someASFR)

     Cohort Age ASFR ASFR1 ASFR2 ASFR3 ASFR4 ASFR5p
3955   1993  50   NA    NA    NA    NA    NA     NA
3956   1993  51   NA    NA    NA    NA    NA     NA
3957   1993  52   NA    NA    NA    NA    NA     NA
3958   1993  53   NA    NA    NA    NA    NA     NA
3959   1993  54   NA    NA    NA    NA    NA     NA
3960   1993 55+   NA    NA    NA    NA    NA     NA
```

Data files from the HFD output database may be easily read into a `data.frame` in R which may then be used directly as input to the cohort fertility table functions using the command `read.table` with the options `header = TRUE, skip = 2, na.strings = "."`, e.g.:[6]

```
> usaASFR <- read.table("USAasfrVHbo.txt", header = TRUE, skip = 2,
  na.strings = ".")
```

Additionally, the function `calculate.cft` can read input directly from file.

---

[6] This and other examples assume that the file `USAasfrVHbo.txt` downloaded from http.//www.humanfertility.org is in the R working directory, which may be located by entering **`getwd()`**at the R prompt.

Users wishing to use `hfdCohortFertilityTable` with non-HFD data may reformat their data to a `data.frame` matching this HFD format with columns Cohort, Age, ASFR, ASFR1, …, ASFRi_b^+p and/or supply different arguments to `read.table`.

Array input format

Alternatively, the ASFR dataset may be supplied formatted as a 3D array with named dimensions of Cohort, Age and Order, e.g. as is produced by the following command:

```
> randomASFR <- array(runif(1860), dim = c(20, 31, 3), dimnames =
  list(Cohort = 1950:1969, Age = c(15:45), Order = c(1:3)))
```

This variation may in many cases be more convenient for use with arbitrarily formatted external data. Note that the names of the dimensions `Cohort` and `Age` must contain only numeric values (cohort birth years and age respectively), although for `Age` any `+` or `–` subscripts will be accepted and propagated to the output: such open age categories are however treated identically to closed single-year age groups in the production of the CFT. The names of the `Order` dimension are ignored: it is assumed that the natural order (1, 2, 3, …) is followed and that no total fertility rate data is included. An array without named dimensions will result in an error.

Output format

The majority of the functions return results as named 3D arrays of data, and these are used internally for the calculations. For convenience, a utility function `array2dataframe.cft` is provided to convert such arrays to a data frame representation analogous to the usual HFD output tables, i.e. with columns Cohort, x, Ix1, Ix2, …, Ixi_b^+p (where I is the indicator) and data rows arranged by Cohort then Age. Note that column x (Age) will contain numeric data only if age subscripts were not present in the input, otherwise character data. An example of how this may be converted to numeric data is given at the end of the next section.

Installation and basic usage

`hfdCohortFertilityTable` is written purely in R and may be obtained as a source package or as a binary package for Windows. The package archive is included with this Technical Report.

Installation from the archive may be performed using the command line, or from within R[7]:

```
> install.packages("hfdCohortFertilityTable_1.0-742.tar.gz",type = "source")
> library(hfdCohortFertilityTable)
```

Once the package is installed, the example data may be loaded using R’s `data` function:

```
> data(someAsfr)
```

It is then possible to directly produce the cohort fertility table with one further function call:

---

[7] Note that the package `abind` is required by `hfdCohortFertilityTable` and must therefore be available, and that installation of source packages on Windows may require additional software. The package version number may need adjusting to match the name of the file you have received.

```
> someCFT <- calculate.cft(someAsfr) # default returns a data.frame
```

Alternatively, the path to a file compatible with the standard HFD format for ASFR by horizontal Lexis parallelograms (cohort/age, VH), may be used directly:

```
> usaCFT <- calculate.cft("USAasfrVHbo.txt")
```

It is also possible to produce cohort fertility tables using different numbers of birth orders to that used in the HFD:

```
> usaASFR <- read.table("USAasfrVHbo.txt", header = TRUE, skip = 2,
  na.strings = ".")
> usaASFR$ASFR4 <- usaASFR$ASFR4 + usaASFR$ASFR5p # combine 4+
> names(usaASFR)[7] <- "ASFR4p" # rename to reflect new content
> usaASFR <- usaASFR[, 1:7] # drop ASFR5p
> newCFT <- calculate.cft(usaASFR) # CFT is now only for birth orders 4+
> tail(names(newCFT)) # verify that there are no 5px columns
```

```
[1] "b4px"  "l3px"  "m4px"  "q4px"  "Sb4px" "chix"
```

Further parameters may also be adjusted, e.g. to change the radix size and to assess the effect of not making the corrections to births and parity data to prevent implausible results and to return the result in a different format:

```
> someCFTuncorr <- calculate.cft(someAsfr, radix = 100000,
  applyCorrections = FALSE, returnType = "list")
> any(someCFTuncorr$l < 0) # are there any negative populations?
```

```
[1] TRUE
```

If an array output has been produced, this may be converted to a data frame in the HFD output table style using the function `array2dataframe.cft` – the data frame format may be more convenient for displaying and working with:

```
> someB <- calculate.b.cft(someAsfr)
> class(someB)
```

```
[1] "array"
```

```
> head(someB)
```

```
[1] NA NA NA  0  0  0
```

```
> someB <- array2dataframe.cft(someB)
> head(someB)
```

```
  Cohort   x   b1x  b2x  b3x b4x b5px
1   1944 12-    NA   NA   NA  NA   NA
2   1944  13    NA   NA   NA  NA   NA
3   1944  14    NA   NA   NA  NA   NA
4   1944  15  76.1  6.3  0.3 0.0  0.0
5   1944  16 203.2 28.7  2.7 0.2  0.1
6   1944  17 387.6 82.8 11.8 1.2  0.2
```

Since the ages have subscripts indicating open age intervals, the column x (Age) currently contains character data and not numeric data. This may however be converted to numeric data (losing the subscripts) as shown:

```
> class(someB$x)
```

```
 [1] "character"
```

```
> someB <- transform(someB, x = as.numeric(sub("[+-]", "", x)))
> class(someB$x)
```

```
[1] "numeric"
```

```
> head(someB$x) # note that subscript to 12 is now lost
```

```
[1] 12 13 14 15 16 17
```

It is of course also possible to produce further outputs, plots, etc. from the results of the functions: some basic examples may be found in the online documentation for the package, a further example is given in the next section.

Further examples

Identifying corrected cells in the calculation of $l_{i-1}(x)$

According to equations 8 – 10, where the calculation of table population by parity $l_i - 1(x)$ based on the matrix of births $b_i(x)$ calculated from the age-specific fertility rates $f_i(x,c)$ produces negative populations or infeasible birth rates or probabilities, the corresponding cells of $l_{i-1}(x)$ and $b_i(x)$ will be corrected. The following script produces the graphical visualisation of the corrections for a cohort shown in Figure 1:

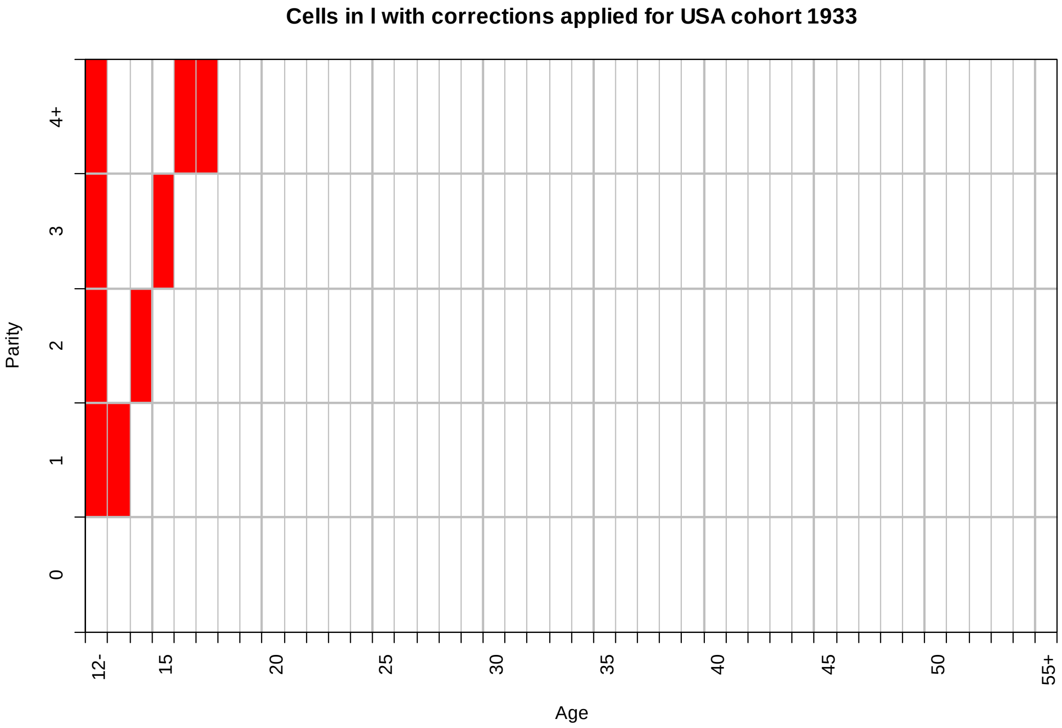

Figure 1: Visualisation of cells requiring correction in calculation of $l_{i-1}(x)$ for USA cohort 1933. Data from Human Fertility Database.

```
> # Step 1: read data
> usa.asfr <- read.table("USAasfrVHbo.txt", header = TRUE, skip = 2,
  na.strings = ".")
> # Step 2: get the array of corrections
> usa.b <- calculate.b.cft(usa.asfr)
> usa.l.with.corrections <- calculate.l.cft(usa.b)
> usa.corrections <- usa.l.with.corrections$corrections
> # Step 3: render the corrections for one cohort as a grid
> image(12.5:55.5, 0.5:4.5, usa.corrections["1933", , ], xlab = "Age", ylab
  = "Parity", col = c("white", "red"), xaxt = "n", yaxt = "n", main = "Cells
  in l with corrections applied for USA cohort 1933")
> # Step 4: improve the aesthetics of the graphic
> abline(v = 13:55, col = "grey")
> abline(v = seq(15, 55, 5), h = 1:4, lwd = 2, col = "grey")
```

```
> axis(side = 2, at = 0.5:4.5, labels = c(0:3, "4+"), tick = FALSE)
> axis(side = 2, at = 0:5, labels = FALSE)
> axis(side = 1, at = c(12.5, seq(15.5, 55.5, 5)),labels = c("12-",
  seq(15,50,5), "55+"), las = 2, tick = FALSE)
> axis(side = 1, at = 12:56, labels = FALSE)
> box()
```

Displaying a surface of conditional age-specific fertility rates for birth order 4+

In this example, we will see that the functions included in `hfdCohortFertilityTable` can be used with data that does not match the range of ages and birth orders used in the Human Fertility Database: we will start by modifying the USA data such that it covers the ages 15- – 45+ and birth orders 1 – 4+ only and then use this as the basis for calculations. This also serves to illustrate how HFD data may be manipulated to match the age and birth order ranges from other data to enable comparisons between the HFD and other sources. For the plotting, the “Trellis” graphics function `wireframe` from the `lattice` package is used, producing the graphic shown in Figure 2.

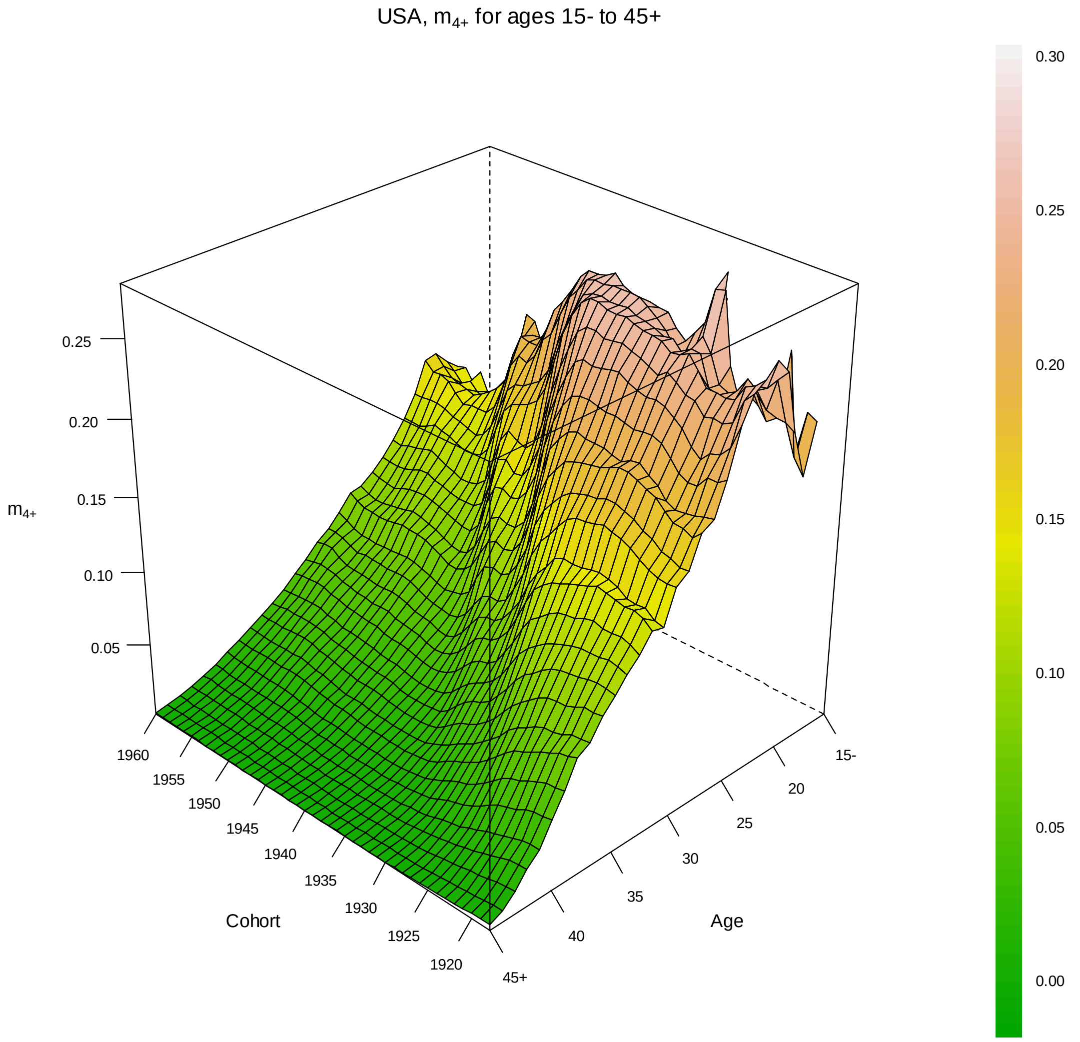

Figure 2. Visualisation of surface of m_{4+} for USA in ages 15- – 45+. Data from Human Fertility Database.

```
> # Step 1: aggregate birth orders to 4+
> usa.asfr <- read.table("USAasfrVHbo.txt", header = TRUE, skip = 2,
  na.strings = ".")
> usa.asfr$ASFR4 <- usa.asfr$ASFR4 + usa.asfr$ASFR5p
> names(usa.asfr)[7] <- "ASFR4p"
> usa.asfr <- usa.asfr[, 1:7]
> # Step 2: remove +/- from the Age column and convert to numeric
> levels(usa.asfr$Age) <- gsub("[+-]", "", levels(usa.asfr$Age))
> usa.asfr$Age <- as.numeric(levels(usa.asfr$Age))[usa.asfr$Age]
> # Step 3: Drop data for any cohorts not observed in the range [15,45)
> usa.asfr <- usa.asfr[!usa.asfr$Cohort %in% usa.asfr$Cohort[(usa.asfr$Age
  == 45 & is.na(usa.asfr$ASFR)) | (usa.asfr$Age == 15 &
  is.na(usa.asfr$ASFR))], ]
> # Step 4: Aggregate ages 15- and 45+
> attach(usa.asfr)
> for (ct in unique(Cohort)) { usa.asfr[Cohort == ct & Age == 15, 3:7] <-
  colSums(usa.asfr[Cohort == ct & Age <= 15, 3:7], na.rm = TRUE);
  usa.asfr[Cohort == ct & Age == 45, 3:7] <- colSums(usa.asfr[Cohort == ct &
  Age >= 45, 3:7], na.rm = TRUE) }
> usa.asfr <- usa.asfr[Age >= 15 & Age <= 45, ]
> detach(usa.asfr)
> # Step 5: Subscript the new open age intervals
> usa.asfr$Age[usa.asfr$Age == 15] <- "15-"
> usa.asfr$Age[usa.asfr$Age == 45] <- "45+"
> # Step 6: Calculate the CFT and convert Cohort and x to numeric
> usa.cft <- calculate.cft(asfr = usa.asfr, observeFrom = 15, observeTo =
  45)
> usa.cft$Cohort <- as.numeric(usa.cft$Cohort)
> usa.cft$x <- as.numeric(gsub("[+-]", "", usa.cft$x))
> # Step 7: Plot m4px as a 3D surface using Trellis (lattice) graphics
> library(lattice)
> dev.new()
> trellis.par.set("axis.line", list(col = "transparent")) # no plot border
> wireframe(m4px ~ Cohort * x, usa.cft, col.regions = terrain.colors(256),
  drape = TRUE, aspect = 1, screen = list(z = 135, x = -60) , zoom = 0.9,
  scales = list(arrows = FALSE, col = "black", x = list(tick.number = 10), y
  = list(at = seq(15,45,5), labels = c("15-", seq(20, 40, 5), "45+"))), xlab
  = "Cohort", ylab = "Age", zlab = expression(m["4+"]), main =
  expression(paste("USA, ", m["4+"], " for ages 15- to 45+")))
```

Summary

This Technical Report has summarised the methodology used in the Human Fertility Database for the production of cohort fertility tables and introduced an "R" package, `hfdCohortFertilityTable`, which allows users easy access to the HFD methods. The functions included in the package have been described, together with examples of their usage. Further details may be found in the online help pages included in the package, and by examining the source code.

References

Bongaarts, J. and Feeney, G. (2006). The Quantum and Tempo of Life-Cycle Events. *Vienna Yearbook of Population Research*, 2006: 115-151.

Jasilioniene, A., Jdanov, . A., Sobotka, ., Andreev, E. M., Zeman, K., Nash, E. J. and Shkolnikov, V. M. (with contributions of Goldstein, J., Philipov, D. and Rodriguez, G.) (2010). Methods Protocol for the Human Fertility Database. URL http://www.humanfertility.org.

Kohler, H.-P. and Ortega, J.A. (2002). Tempo-adjusted period parity progression measures, fertility postponement and completed cohort fertility. *Demographic Research* 6(6), pp. 91–144.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Wunsch, G. (2006). Rates, Frequencies, and Probabilities; in Caselli, G., Vallin, J., and Wusch, G. (Eds.) *Demography and Synthesis*, Vol. 1. Amsterdam et al.: Elsevier, pp. 79-87.