Documentation for

Children per Woman (Total Fertility Rate)

for countries and territories

Gapminder Documentation constitutes work in stepwise progress.
We welcome all sorts of comments, corrections and suggestions through e-mail to the author.

Gapminder Documentation 008
Version 2
Uploaded: 2010-11-04
Published by: The Gapminder Foundation,
Sweden, Stockholm, 2009

Author: Ferenc Ajus, edited by Mattias Lindgren
E-mail: mattias.lindgren (at) Gapminder.org
Gapminder is a non-profit foundation
promoting sustainable global development
by increased use and understanding of statistics.
www.gapminder.org

CC BY ND

1 Summary

Our "children per woman" indicator contains historic total fertility rate (TFR) data for 195 countries in the world, for the 1800-2008 period within their current borders. The main source of the data is the UN fertility dataset, which covers the period between 1950 and 2008. In addition, it has been complemented with data collected from statistical sources and various published articles. For almost all countries, the data now cover the period when fertility changed from high (e.g. 6 children per woman) to low levels (e.g. 2 children). For less developed countries, the UN dataset was enough for this, because the transition happened after 1950. For more developed countries, we could successfully extend the dataset backwards in time, to establish the pre-transition level of fertility. Based on this, we could also estimate the 1800 level of fertility for almost all countries.

The first release of the dataset was issued in 2009, followed by a revision in 2010. The two datasets are named as Gapminder children per woman 2009 (in short, Gapminder 2009) and Gapminder children per woman 2010 (in short Gapminder 2010) accordingly. The aim of the current document is to combine the studies that have been prepared throughout the project into a single coherent piece of documentation. The main changes in the 2010 release were that the estimates for the 19th century fertility level have been re-considered, especially in case of countries where there is no statistical information throughout the 19th century. In some cases, the estimates were deleted and in others more data were added from country-specific or regional literature.

The dataset may be used freely, but users are warned that it contains both good quality observations and rough estimates. However, it is easy to distinguish between them. Data quality is indicated in the dataset in the field "qual." for each record. Data with quality indicator 4-5 are estimates, and should not be used for research. Quality 1-2 are taken from a statistical source, published article or book, which is reliable in our view. Data quality 3 contains *crude birth rates* from a reliable source, but as TFRs were not available in the source, the authors estimated TFRs using a simple re-scaling conversion. You may also see the source of each data record in the dataset, cross-referenced to a list of literature used.

We hope that this work is only the beginning, we would be happy to receive any comments or suggestions for corrections. If you have additional fertility data please do not hesitate to them to send us. We are aiming to process these and include them in a later release of the dataset.

If you have any comments, please contact us using the e-mail address provided on our contact page on www.gapminder.org.

Contents

1 Summary 3

2 Introduction 6

3 Theoretical background 6

3.1 Theoretical issues 8

3.2 Regional levels of pre-transition fertility 9

3.2.1 Islamic fertility 9
3.2.2 Sub-Saharan Africa 10
3.2.3 Small Islands 12
3.2.4 Caribbean 12
3.2.5 Land availability and fertility 12

4 The explanation of the Gapminder fertility dataset 12

4.1 Data quality 12

4.2 Sources and results of data collection 15

4.3 Estimation and extrapolation methods used 17

4.3.1 TFR estimates from Princeton fertility indexes 17
4.3.2 TFR estimates from Crude Birth Rates 17
4.3.3 Extrapolating the 1950-1955 TFR backwards 18
4.3.4 Utilising additional sources to estimate “natural fertility” 18
4.3.5 Using regional averages or neighbouring countries 19
4.3.6 Extrapolation not possible 21

4.4 The structure of the dataset 21

5 Data quality review and corrections 22

5.1 Background studies prepared 23

5.2 Call for comments and additional data 23

6 Understatement of fertility in official statistics in Sri Lanka and Singapore – a critical review of sources 24

6.1 Some basic points about the demography of the region 24

6.2 Sri Lanka 26

6.2.1 Literature review - Sri Lanka 26
6.2.2 Crude fertility data in Sri Lanka 26
6.2.3 TFR data in Sri Lanka 29

6.3 Singapore 32

6.3.1 Literature review - Singapore 32
6.3.2 Fertility data in Singapore 33
6.3.3 Summary 36

7 Fertility in India: A review of estimates and theories 38

7.1 Literature review 38

7.1.1 India and the population debate 38
7.1.2 The determinants of fertility in India 39
7.1.3 The determinants and stages of fertility decline in India 39
7.1.4 Case study evidence for India 40
7.1.5 Stages of demographic transition in India 41
7.1.6 Diffusion was not balanced 41
7.1.7 Economic explanations 41

7.2 Fertility data and sources 43

7.2.1 Problems with the 2009 estimates 43
7.2.2 Newly found data and estimates 43
7.2.3 More details regarding the preferred estimate 46
7.2.4 The territorial range 47
7.2.5 Estimate for pre-transition fertility 47
7.2.6 Summary 49

8 Fertility in China: A review of estimates and theories 50

8.1 China in the 19th century 50

8.2 Use of model life table to see the CDR (crude death rate) level 50

8.3 Calculation for the estimates of fertility level 51

8.4 Conclusion 53

9 Smaller country reviews (data cleaning) 54

10 Reasonableness testing 63

10.1 Empirical approach to find the long term maximum death rate 63

10.2 Demographic reconstruction methods 66

10.3 Results 69

11 Literature 73

2 Introduction

At the request of Gapminder Foundation we have compiled a historic Total Fertility Rate (TFR)[1] dataset for 195 countries of the world. The purpose of this documentation is to explain the methods applied in some detail. The report will aim to:

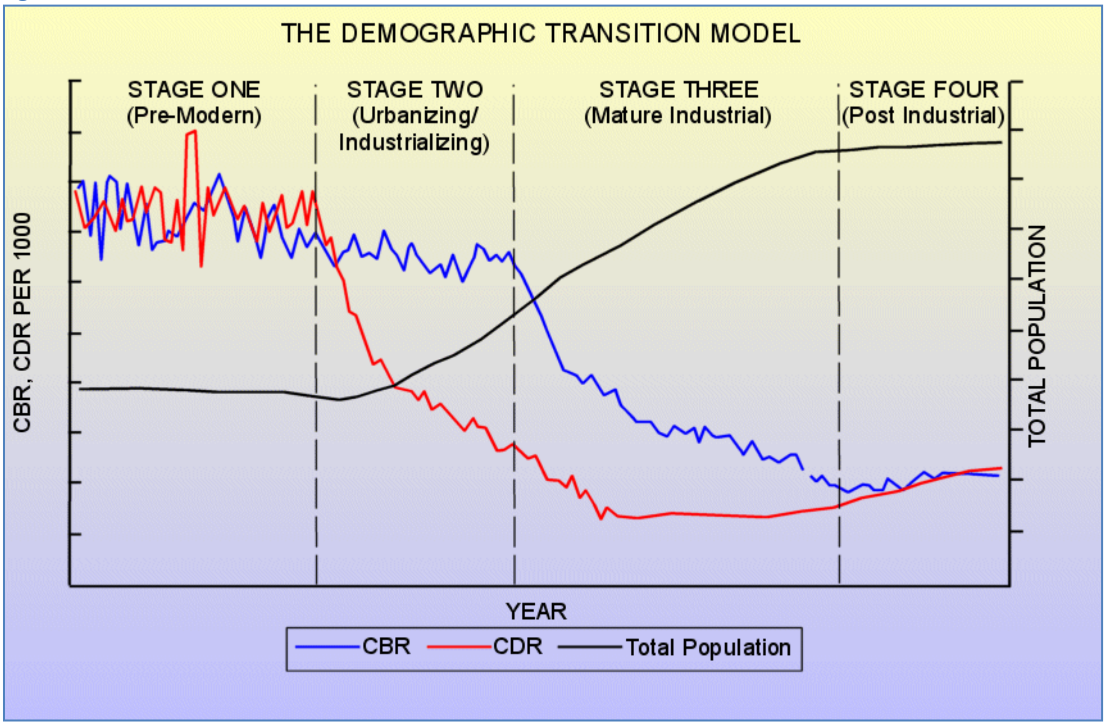
- Explain the approach (some theoretical background)
- Show the main sources of fertility data we have used
- Explain estimation and extrapolation methods used in case of missing or limited data
- Explain the structure of the dataset collected
- Show research background and references

Appendices include data sources, examples and the data themselves in graphs by region.

3 Theoretical background

Demographic transition theory played a key role in our work, especially in the estimations and data quality reviews. What do we mean by demographic transition? It is best illustrated with a chart (Figure 1).

Figure 1



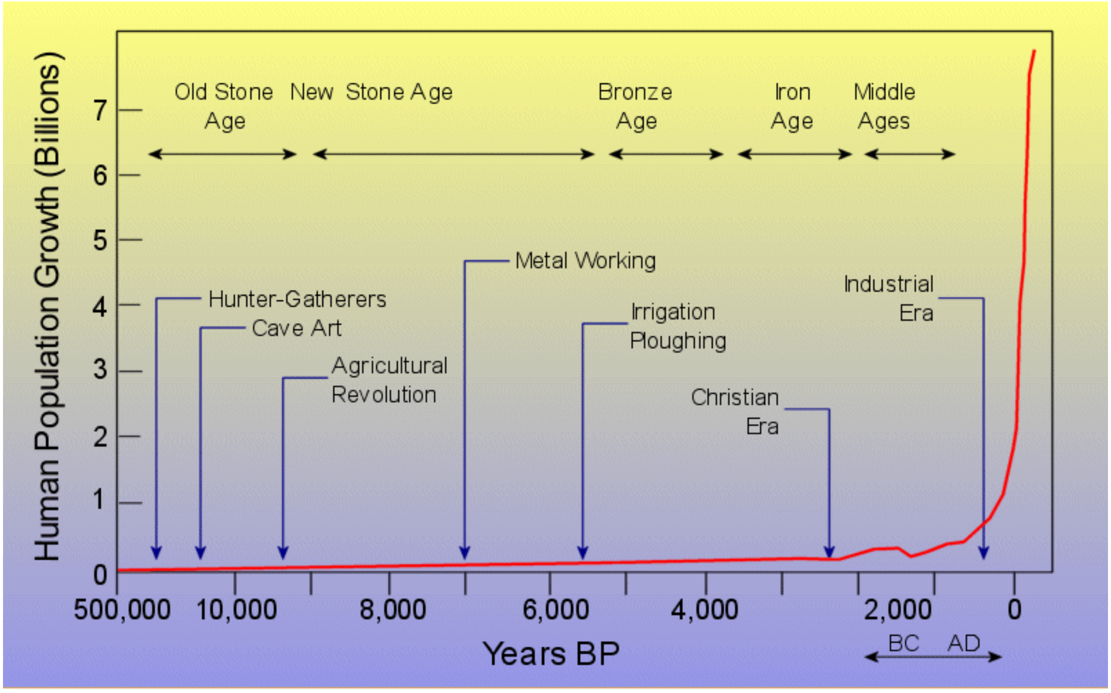
Source: http://cas.bellarmine.edu/tietjen/images/Human%20Population%20Growth.htm

The fertility level before the demographic transition is often referred to as natural fertility. Demographers tend to believe that fertility levels before the demographic transition have fluctuated

[1] The total fertility rate or TFR is the sum of age-specific fertility rates. It can be interpreted as the number of children an "average woman" would have in the given population. In other words TFR gives the average number of children that would be born to a woman over her lifetime if (1) she were to experience the statistically observed yearly age-specific fertility rates through her lifetime, and (2) she were to survive from birth through to the end of her reproductive life.

around a high level, say, six children per mother, and it was not consciously controlled. There was a balance because both fertility and mortality was high. In most countries mortality started to decline first, and fertility was only adjusted later, and this delay often caused a large increase in population.

Figure 2



Source: http://cas.bellarmine.edu/tietjen/images/Human%20Population%20Growth.htm

Throughout the history of mankind, population growth has been very moderate, but the delayed adjustment of fertility after the start of demographic transition has caused a massive population increase. The dramatic population increase is one of the main reasons why so many demographers have studied the demographic transition.

In our work we have mainly relied on the following sources:

The book of Chesnais (1992) was the second main source of data after UN. It includes a lot of useful data and analysis of various forms of Fertility Transition around the world. Caldwell (2006) includes a two page comprehensive table, which estimates the date of onset and speed of fertility decline in all major regions of the world and includes data by country as well. The country coverage is broader than in Chesnais, but the data are less exact. We have not used this table directly, because our dataset is more detailed. However, the classification of the countries was very informative, and we have implemented this in order to check our data. For example, Caldwell created a separate category for "European Population Overseas" for countries like the US, Australia and Chile, where the fertility transition started about as early as in Europe, while the decline in most non-European countries started much later, after the 1950s.

Mitchell (1998) is probably the best book on historical statistics; it comes in three large volumes. It included some fertility (crude birth rate) data, which overlapped with Chesnais (1992), but sometimes complemented it. Dyson (1985) includes data about the onset of the fertility decline in the World. Sardon (1996) provides the method and multiplier for converting Cole's indexes to TFR.

3.1 Theoretical issues

The main conclusion of our literature review work was that the assumption that preceding the fertility transition, fertility was fluctuating around a high plateau is usually correct, but sometimes incorrect. Our data and the model calculations show that pre-transition fertility was sometimes lower than the traditional thinking along the lines of "natural fertility" would suggest. Historical demographic studies show many examples from Asia and Africa of social controls, such as prolonged breast feeding, abstinence and prohibition of widow remarriage, which kept fertility low, despite marriage at a young age. Many of the countries that recently had very high fertility levels only had moderately high fertility levels in the 19th century, due to the application of social controls.

This also means that back-projecting the level that we see in the 1950s may sometimes lead to overstatement, because the 1950 level was a result of a pre-transition fertility increase. Unfortunately, for about every second country in the world, we have no fertility data for the 19th century.

Is there any way to estimate 19th century fertility with any scientific correctness? The first method we applied was to search for more data backwards in time that lead us back to the true pre-transition level. In the case of many countries this has been possible and we could estimate the fertility level for 1800 based on the early plateau that can be seen in the dataset. Sometimes it was also possible to find expert estimates in the literature specifically for the pre-transition fertility level in a given country, which we could apply in the dataset.

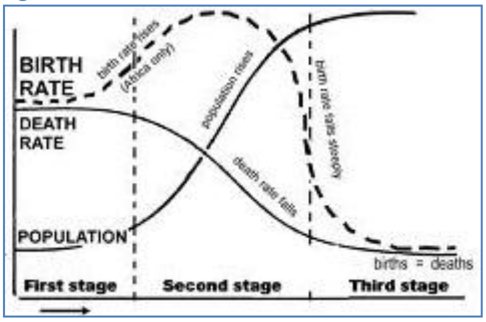
The third possible solution would be to create a regional typology based on historical demographic studies, existing data and stable population estimates. The historical demographic studies help us to understand the fertility system in a given region, using the example of a few villages. The stable population model requires an estimate of mortality and population growth to give a point estimate on TFR. The insights provided by the calculations can be combined with regional literature to create a typology, giving an approximate range for 19th century fertility for all major regions of the world. This would require a larger effort, which was out of scope for the current work.

3.2 Regional levels of pre-transition fertility

In this section we briefly summarize the literature review regarding the pre-transition level of fertility by region. In developing countries, it is often observed that modernisation had a positive effect on fertility. Fertility at first increased and only after that did it start to decline (Nag, 1980). In brief, modernization has both positive and negative effects on fertility. The positive effects are better health and fecundity, lower mortality (for example, the lower incidence of widowhood in fertile ages) and the disruption of the social norms that had previously kept fertility low, like decrease in lactation period and post-partum abstinence. The negative effect is that with increased education and broader social contacts birth control methods spread, causing marriage delay and a decline in marital fertility. The time trend of fertility is determined by the balance of the positive and negative effects. In Europe, the positive effects have only rarely become visible (e.g. in some regions of Transylvania 1900-1910 – Ajus 2009), but in Africa and Asia we often see a temporary fertility increase before the decline (Nag, 1980). The examples of Nag include many countries from Latin America, the Caribbean and Asia and Africa. [2]

If we assume that a pre-transition fertility increase took place, then we have to modify the view of demographic transition as shown in the following chart (Figure 3). If this figure applies to a country, then it is difficult to tell what the pre-transition level of fertility was, unless we have data for that period. It may be a mistake then to assume that the high fertility level that we see around 1950 can be extrapolated backwards even to the 19th century. For this reason we have reviewed all estimates for the 19th century, which assumed TFR of 7 or higher, provided that the estimate was only based on extrapolation from 20th century data. The review involved a literature review (see in the current section and appendix 9) and reasonableness test (see in the appendix 10) and the necessary corrections have been made in the Gapminder 2010 edition, mainly by dropping those estimates for which we could not find enough support.

Figure 3



3.2.1 Islamic fertility

Many of the countries for which high 19th century fertility had been assumed in the Gapminder 2009 dataset are Arab and/or Islamic countries. We have reviewed some articles regarding fertility in the Arab world. Demographers often associate the Islamic countries with higher fertility; because this region was late in the demographic transition and in some cases fertility remained high despite growing levels of income. This is sometimes attributed to cultural norms in the Islamic world, which

[2] Additional reference: Dyson, T., and M. Murphy. 1986. Rising fertility in developing countries. In R. Woods and P. Rees (Eds.), Population structures and models: Developments in spatial demography. Boston, Massachusetts, USA, and London, England: George Allen and Unwin, pp. 68-94.

supposedly suggest higher family size and stronger resistance to modernization and change. Muslim society has for a long time “vigorously resisted fertility decline, largely as a result of the roles played by women's status and family structure” (Fargues 1989). Some recent articles aim to expel these myths (Eltigani 2005). Rashad (2000) agrees that the Arab countries had the highest level of pre-transition fertility in the world, but in his view this was not due to a special religious preference. Instead, the high pre-transition fertility was due to the absence of non parity-specific factors (i.e. social controls, like post-partum abstinence and long lactation, or late marriage). But in his view this does not prove that fertility aspirations were inherently higher. A closer examination reveals large differences between countries, which can be linked to economic circumstances rather than religion. On the whole, the demographic transition in the Arab countries was similar to that of other developing countries.

Interestingly, however, the debate only concerns the impact of Islam on fertility decline and there seems to be little doubt that pre-transition fertility was very high. According to Fargues (1989, pg 151) “the data are not sufficient for a proper comparison of the recent past in these countries. However, it seems probable that around 1960 total fertility was of the order of 7 to 8 children per woman in all of them (except Lebanon).” And Rashad (2000) also agrees with the notion of high pre-transition fertility. However, from these studies it seems that none of the authors could rely on any specific statistical information or survey for the pre-1950 period.

However, both my calculations (based on population growth and likely level of mortality) and the generalization from the Nag (1980) study make it likely that fertility at the 7-8 level was only temporary and that for earlier periods a lower estimate is more appropriate.

3.2.2 Sub-Saharan Africa

My view that very high fertility may not be assumed for a longer period gets much more support in the case of the Sub-Saharan countries. A very interesting study by Garenne, sponsored by US Aid and the French Institute of Research for Development (IRD), reviews the fertility history of the region country by country (Garenne, 2008).

The graphs in Garenne’s study show that in many cases there was a pre-transition fertility increase. Fertility in the first half of 20th century was probably lower, because of the high incidence of infecundity due to disease. To follow are two graphs taken from Garenne’s (2008) study, which show the general tendency, but the study also includes estimates by country (Figure 4 and Figure 5).

Sarah Walters (2008) reached similar conclusions in her study about the population history of Tanzania.

Figure 4 Overall fertility pattern observed in Sub-Saharan Africa (source: Garenne 2008, pg 5)

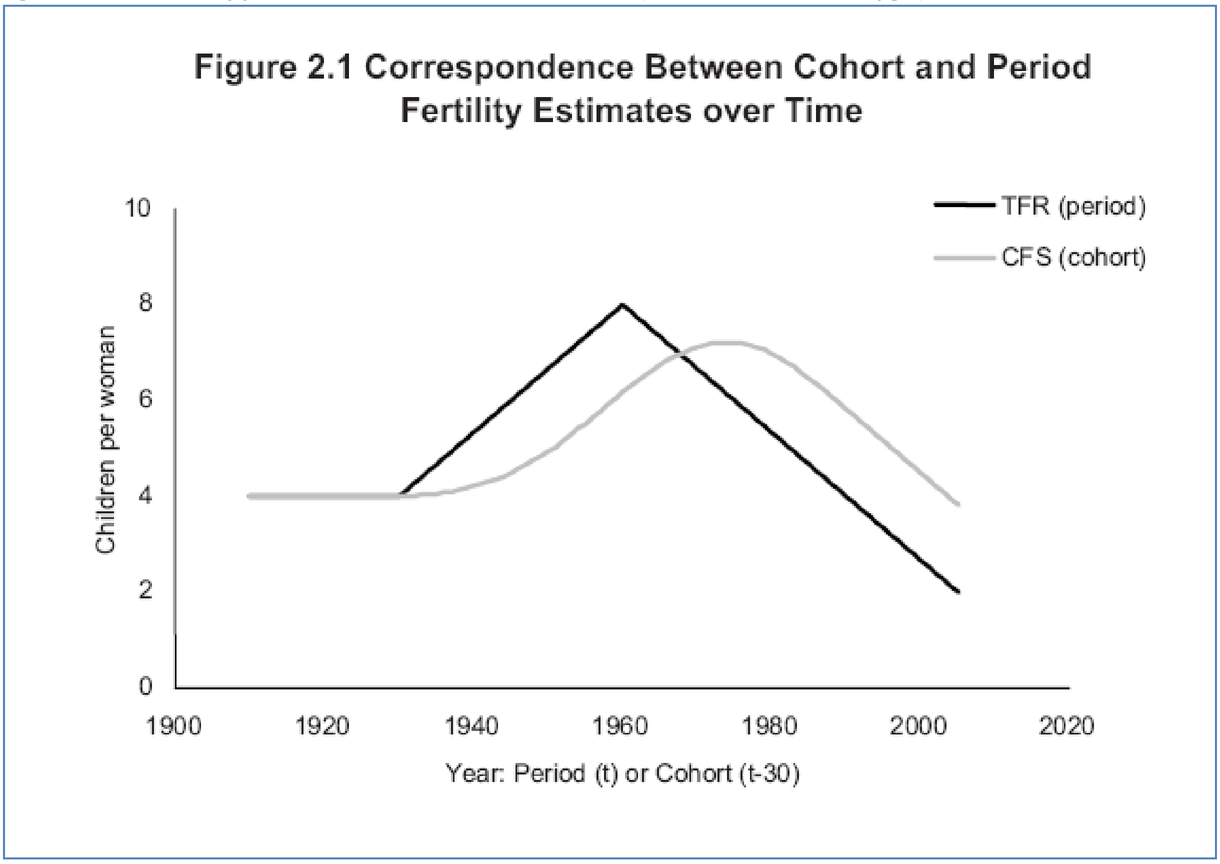
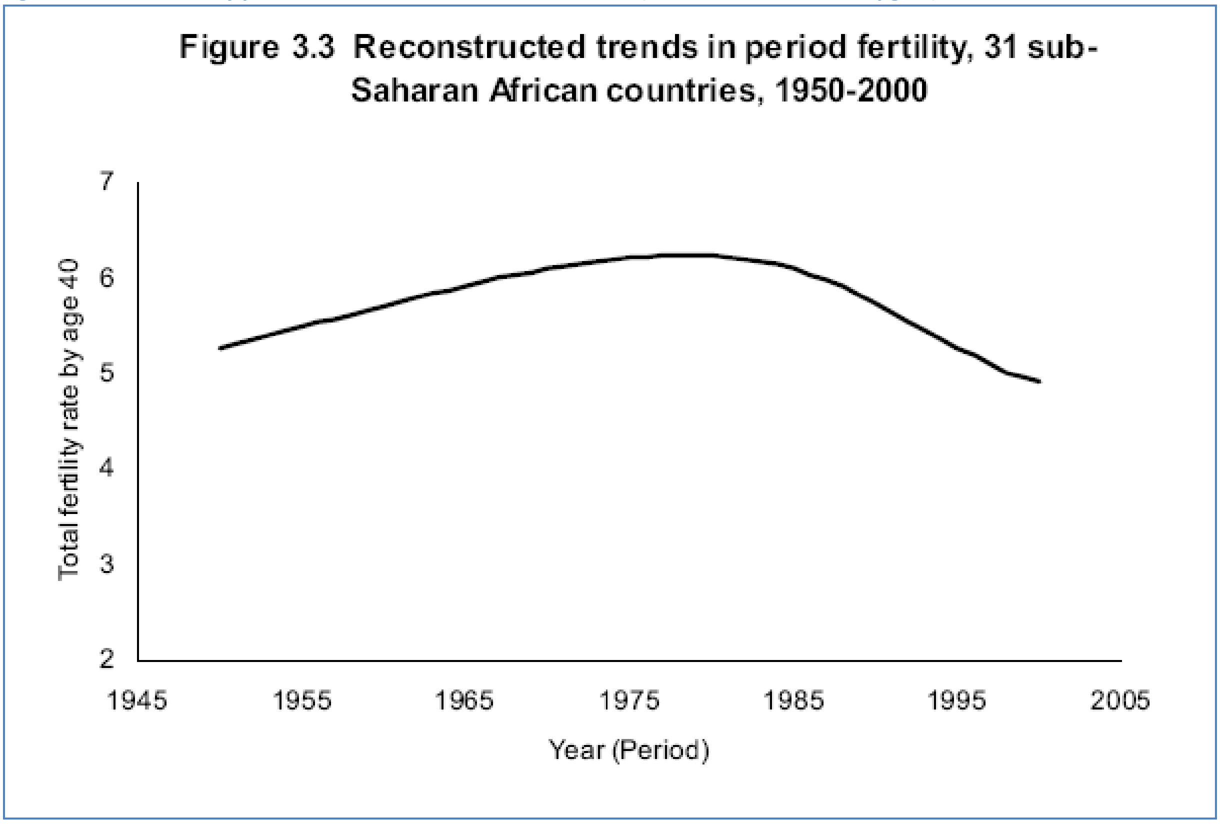


Figure 5 Overall fertility pattern observed in Sub-Saharan Africa (source: Garenne 2008, pg 11)



"Reconstructed trends in period fertility for the 31 countries in sub-Saharan Africa indicate a rise from a TFR of 5.3 children per woman in 1950 to 6.2 children in 1980, followed by a decline to 4.9 children in 2000." (Garenne 2008 pg 18.)

3.2.3 Small Islands

There is an argument that fertility on small islands was also limited in the past by the amount of food the dwellers could produce. Islanders probably had relatively high life expectancy, as they had limited outside contacts and this reduced the incidence of epidemics. They probably controlled fertility in some traditional ways to limit population increase. For example, Pirie argues that "it has been observed repeatedly that densities in the Pacific islands never seemed to rise to the levels at which their carrying capacities were strained" (Pirie 2000 p 13). For this reason he argues that it is a myth that the recently achieved high fertility levels (TFR being around 7) were also sustainable in the past. In his view, pre-transition fertility was certainly lower. Although little is known about the exact level of pre-transition fertility, a 1950 anthropological study in Papua New Guinea indicates that the completed fertility was about 3.5 children per mother, due to social controls. Pirie believes that the situation in the Pacific Islands had probably been similar; he quotes the examples of Nukuoro and Eauripik, two islands in Micronesia, for which data exist over a longer period. (Please note that 3.5 completed fertility is not that low, the corresponding TFR can still be 4 – 7, depending on the level of infant and child mortality).

3.2.4 Caribbean

Our literature review regarding the Caribbean indicates that slave fertility was probably lower at the beginning of the slave trade era, improving gradually towards the end of the era. This is explained by improved living conditions for slaves and perhaps also fewer instances of physical separation of spouses. For this reason we may expect rising fertility in Caribbean countries during the end of 19^{th}/beginning of 20^{th} centuries, when 70-80% of the population were often slaves (Higmann, 1995; Engerman, 1997).

3.2.5 Land availability and fertility

There is a well known connection between fertility and land availability. For example high fertility has been observed in the US and Canada in frontier states, where land was abundant (Haines & Steckel 1997). On these grounds we have found the high 19^{th} century estimate for Ukraine reasonable, for example.

4 The explanation of the Gapminder fertility dataset

4.1 Data quality

The aim of the project was to collect a dataset that may be used to show the transition from high to low fertility in an interactive moving graph. This objective was achieved reasonably well, mainly using reliable sources, but partly also using estimates. The general tendency of demographic transition has been captured in a way that is in line with the current knowledge in historical demography (see also Caldwell, 2006 and Chesnais, 1992).

The data collection was guided by the principle that estimates should be avoided if at all possible and that looking for more data is preferable to searching for an estimation model. This approach builds on the knowledge what we have about fertility transition: as the explanations for fertility transition and pre-transition fertility levels are controversial, we do not have a model that would predict fertility with reasonable accuracy from other data. The transition from high to low fertility was usually quick; it occurred within a few decades. If the period of transition is not covered with data, then it is very difficult to make good estimates. However, pre-transition fertility (or in Henry's terms "natural fertility") was more stable and a few data points should be enough to be able to make a rough estimate.

Luckily, most of the time data availability coincides with modernization and modernization coincides with fertility transition. This means that with some exceptions, we have data for the period of fertility transition and we only have problems before that, which causes less inaccuracy because "natural fertility" is less volatile.

Keeping this in mind, during the second stage of data collection we identified and focused on those countries where the fertility transition started before 1950 (not being covered by UN data),.

Limitation 1: We know little about the level of "natural fertility"
While the data are suitable for showing the general tendencies of the transition from high to low fertility, the fluctuation and level of pre-transition fertility remain unknown for most countries and were out of scope for this project.

Limitation 2: The use of estimates
This dataset also includes some estimates, as explained in paragraph 4.3. Therefore, it should only be used for research after filtering out the estimates, that is, all the data with quality indicators 4 and 5. In addition, data rows with quality indicator 3 contain only rough estimates for TFR. It is therefore more advisable to use the original CBR values instead in this case. The following table explains the possible values of the data quality indicator:

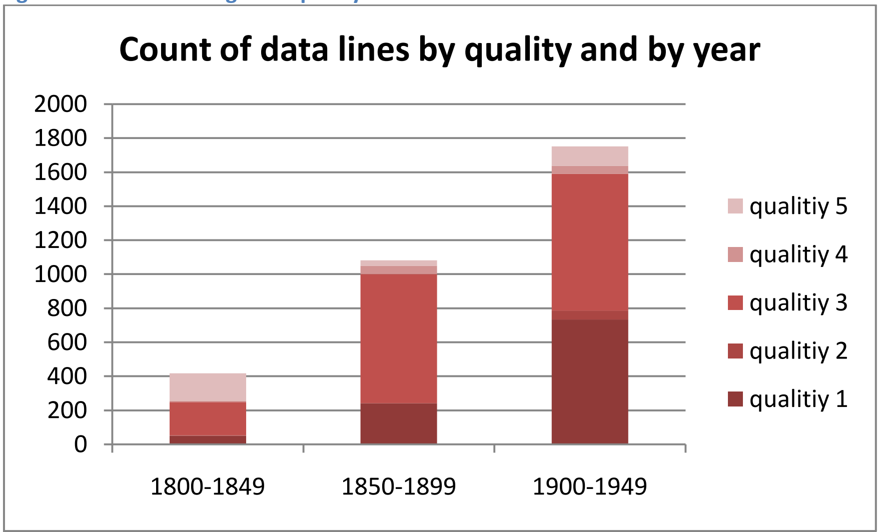
Table 0

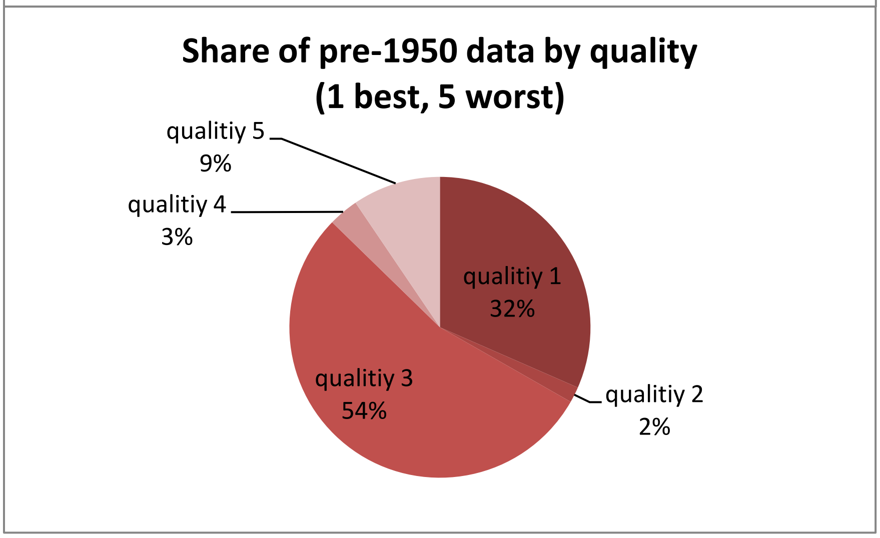
Quality indicator	Explanation	Suitable for research purposes?
1	Very good: TFR data taken from statistical source or published article	Yes
2	Good: mainly used for the UN data (1950-2008), which includes some estimates	Yes
3	Medium: crude birth rates (or Princeton If indexes) were taken from statistical source or published article, but TFR was estimated by the author (using a simple conversion)	Yes
4	Slightly uncertain: no country-wide data were available, but a published estimate was found. This estimate is usually based on historical demographic research for a smaller area.	No
5	Uncertain: estimate (or guess) by Gapminder	No

This means that only TFR data with quality 1-2 and crude rates with quality 1-3 may be used for research. We recommend that researchers who may wish to use this dataset filter out the values with 4-5 data quality indicators.

When we assessed data quality, in most cases we assumed that data from published sources are reliable and awarded quality indicator 1-3 without undertaking detailed source critique. The time series for a single country are often assembled from several sources. In some cases there was a contradiction between sources, which we had to clarify. In addition, the early statistics of some countries may be unreliable and may understate fertility. This issue is covered in the data review sections (6.4 and 6.5) in the appendix.

Figure 6 Charts showing data quality





As our readers may have expected, we have many more observations as we approach the present time and the data quality is also improving. The second chart shows that only 12% of our data are estimates, all the rest comes from a statistical source or the literature.

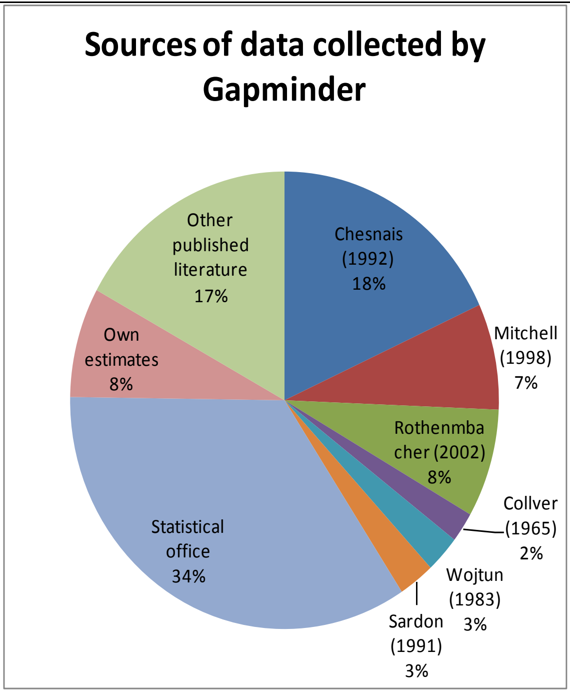
It may also be seen that more than half of the TFR values have been converted from CBR. The reason for this is that for many countries, until very recently, the official statistics do not contain enough information to calculate TFR, because the age of mothers at birth is not available. The method of converting CBR to TFR is only a crude one, called “empirical conversion”, and is nothing more than a mere re-scaling, as explained in section 4.3.

4.2 Sources and results of data collection

Between 2008 and 1950, the UN fertility dataset was used. Our work mainly concerned finding data sources before 1950, with a special focus on those countries where the transition started earlier than that. The following table shows the number of records collected by main sources:

Table 1. Count of records (observations) in the dataset by source

Sources of records	**TOTAL**
UN dataset (2008)	11166
Own collection	3722
Total data rows	14888
"Own collection" in more detail:	
Chesnais (1992)	675
Mitchell (1998)	280
Rothenmbacher (2002)	286
Collver (1965)	79
Wojtun (1983)	98
Sardon (1991)	101
Statistical office	1281
Own estimates	290
Other published literature	632
Total data rows, own coll.	3722

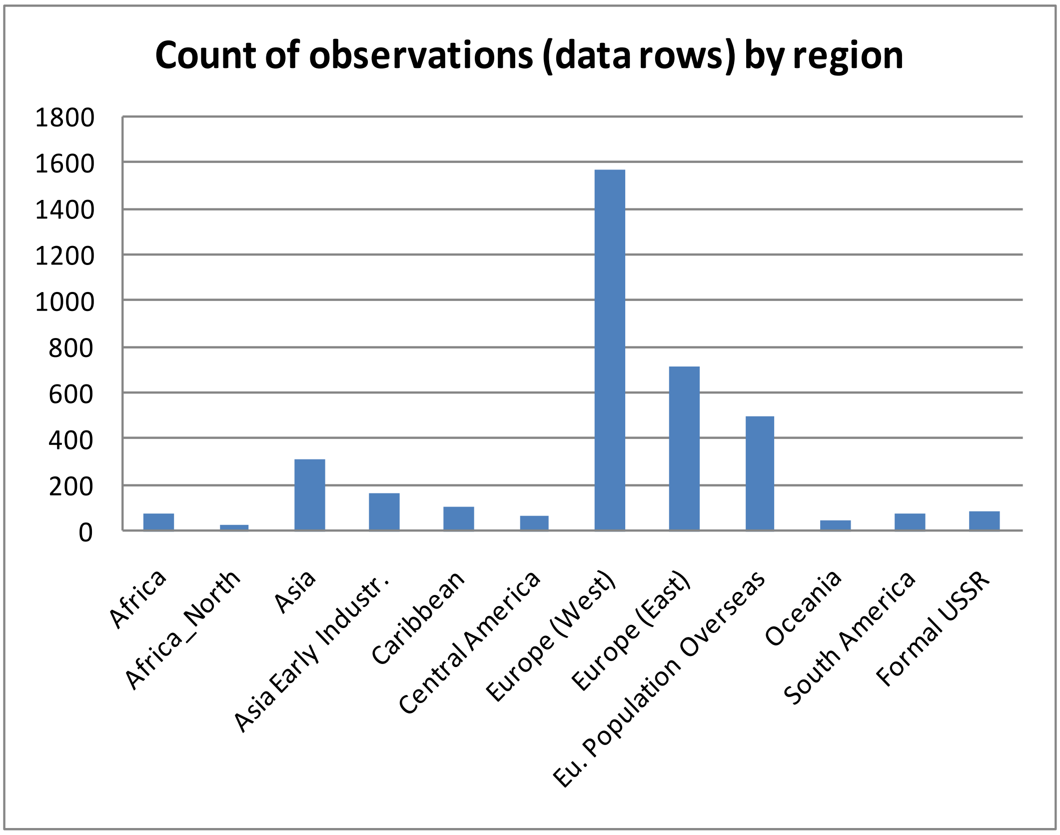


The "other" category includes the following:

- books about Historic Fertility and fertility transition
- search for articles about individual countries
- dataset provided by Tomas Sobotka
- data provided by Patrick Gerland
- Princeton European Fertility Project (I_f indexes were converted)
- results of data quality reviews

The full list of sources can be found in the dataset on sheet "Sources" and also in the reference list at the end of this document. The data collected may also be shown by region. We have used a demographically meaningful regionalization, as it was proposed by Caldwell (2006).

Figure 7



As you may observe on the graph, much more data were available for the more developed countries. We know a lot about the past of European and Asian populations and about some of the formal colonies, which were mainly populated by Europeans, but we know very little about the population history of the rest of the world.

4.3 Estimation and extrapolation methods used

The main challenge was to estimate the TFR values for 1800. This is the starting point of interpolation for the software, but for most countries we had no data so far back in time. We used a variety of methods described below to extend the data to before 1950.

4.3.1 TFR estimates from Princeton fertility indexes

Researchers within the framework of the European Fertility Project have used Princeton fertility indexes (I_f). These indexes compare the fertility of a given region with the fertility of the Hutterites, a population where no birth control was exercised due to religious reasons. The data have been published and they cover Europe on a county level, providing 4 observations between 1880 and 1940. These data can be converted to TFR with some accuracy because they include some information about age distribution. For example, we know the TFR of the Hutterites and we know that $I_f = 0.6$ means that fertility is 60% of the Hutterites. TFR can be determined from this. Sardon (1996) suggests 12.44 as a conversion key. We have applied this conversion to fill the data gap for a few European countries, like the Baltic countries and former Yugoslav states.[3]

4.3.2 TFR estimates from Crude Birth Rates

Age-specific birth rates are often unavailable from before 1950. In many of these cases we could estimate TFR from the crude birth rates. Here we used the method suggested by Bogue (1993). This textbook includes an empirical table of equivalent values between four different fertility measures. Among others, this includes conversion rates between Crude Birth Rate (CBR) and Total Fertility Rates (TFR). This empirical conversion was based on regression using UN data. The conversion is relatively exact, as the correlation is 0.989 between CBR and TFR. We did not use this table directly, but instead we have implemented the same regression method on our dataset. If possible, the conversion rate was determined by country using this method, but when there were not sufficient TFR data, we have applied a general conversion factor calculated from the UN dataset. (TFR=0.134*CBR)

[3] The Princeton dataset is available on-line at: http://opr.princeton.edu/archive/pefp/.

The following table gives the conversion rates for all countries, where a country-specific conversion was applied.

Table 2
Slope (conversion rate between crude birth rate and TFR) calculated for individual countries

Austria	0.1381
Belgium	0.1461
Denmark	0.1350
France	0.1500
Germany	0.1346
Hungary	0.1220
Japan	0.1478
Netherlands	0.1460
Norway	0.1439
Portugal	0.1300
Spain	0.1323
Sweden	0.1418
Tunisia	0.1490
United Kingdom	0.1358
United States	0.1193

In all cases where TFR was calculated from crude rates, both were included in the dataset in separate columns. As the conversion is only approximately correct, we have lowered data quality to “3” in these cases.

The conversion factor is shown in the column “Conversion key” in the dataset when it differs from the general one (0.134).

4.3.3 Extrapolating the 1950-1955 TFR backwards

When none of the above were possible we sometimes assumed that the fertility observed in 1950 was natural, since the countries in question seemed to have passed their fertility transition after this year. Hence we assumed the same TFR in 1800 as in 1950. Accordingly, for some countries where the TFR was still high (above 5.5) in 1950, we attempted to find the level of natural fertility based on the data that we have. This involved calculating the average of the years 1950-55, or if there were earlier data, trying to find a “plateau” around which the data tend to fluctuate. Essentially, we assumed that the 1950-55 level of fertility is natural fertility and that fertility was similar in 1800.

In cases of countries where we had no data between 1800 and 1925, we repeated the 1800 estimate for 1925. This was done to show that we assume high natural fertility for the whole period. This was also needed to avoid a situation where data are automatically extrapolated between a high point in 1800 and a lower point in 1950. While the assumption of high natural fertility in general for these countries is in line with the literature, the year 1925 was selected in an arbitrary way, with this technical purpose in mind.

4.3.4 Utilising additional sources to estimate “natural fertility”

For many countries, the UN data from 1950 were not enough, since the fertility transition started earlier than 1950 in these countries. In most of these cases, efforts to cover fertility transition with at least some data were successful. Two different methods were used.

Firstly, in most cases we could find additional data for the earlier years and were able to look for a “plateau” in the data to identify the pre-transition fertility level. We have used this level as an estimate for fertility in 1800.

Secondly, for some countries we found an expert estimate of “natural fertility” in the literature, or some article which enabled us to estimate the pre-transition level of TFR. For example, Vishnevskij (2006) includes specific estimates for pre-transition TFR in ex-USSR countries; while Saito (2006) includes information on fertility level in two provinces of India, from which we could estimate the rough level of TFR for India as a whole.

4.3.5 Using regional averages or neighbouring countries

When none of the above was possible, we applied regional averages, or data from neighboring countries.

The full list of these countries is shown in Table 3. In case of a few other countries we had to make our estimates based on 1-3 observations, looking at the regional data at the same time. These are shown in Table 4. [4]

[4] Tables 3 and 4 contain all the exceptions. The estimate for the rest of the countries is more reliable, as they either had their transition during the period covered by UN data, or we could find 6 or more observations before 1950.

Table 3.

Countries where the initial phase of the decline is not covered with data

Estimates for 1800 were either (1) based on regional average plus a random factor, or (2) based on data of neighboring country.

Country	Besides the UN, how many observations we had	Estimated TFR (1800)	UN first TFR (1950)
Central African Republic	1	6.508	5.4
Gabon	2	6.500	4.1
Gambia, The	1	6.512	5.4
Sierra Leone	1	6.518	5.5
Hong Kong, China	1	5.47	4.5
Macao, China	2	5.51	5.3
Bahamas, The	2	5.90	4.0
Grenada	1	5.81	5.2
St. Lucia	2	5.53	5.3
Guam	1	6.37	5.4
New Caledonia	1	6.41	5.1
Reunion	1	6.51	5.5
French Guiana	1	6.27	5.1

The criteria used to select these countries: TFR in 1950< 5.5 and the count of observations without UN are less than 3.

Table 4

Countries where the initial phase of the decline is scarcely covered with data

Estimates for 1800 were based on the few existing country observations before 1950, using regional comparisons as well.

Country	Besides the UN, how many observations we had	Estimated TFR (1800)	UN first TFR (1950)
Channel Islands	3	5.07	2.1
Montenegro	5	5.90	3.5
Slovenia	5	4.93	3.1
Bosnia and Herzegovina	5	5.91	5.1
Moldova	4	6.39	3.5
Belarus	4	7.00	3.4
Georgia	4	7.80	3.1
Kazakhstan	3	6.58	5.0
Kyrgyz Republic	3	6.60	4.3
Ukraine	4	7.50	2.7

The criteria used to select these countries: TFR 1950< 5.5 and count of observations without UN are less than 6 but more than 2.

4.3.6 Extrapolation not possible

In some countries, fertility was very high in the 1950s and the assumption that the same fertility level prevailed for a long time (1800-1950) became unrealistic. We have carried out a formal reasonableness test to check the countries where TFR would have been above 7 in 1800 had we used extrapolated values. We excluded the observations before 1950 for the countries which failed the reasonableness test (see Appendix 10), and for which no other method was available.[5]

4.4 The structure of the dataset

We have followed the structure of the existing Gapminder life expectancy dataset. Please look at sheet “DATA” and “Sources”.

Rows: In the sheet DATA, rows are yearly observations, or estimates. The statistics about the count of rows (or records) in the dataset are shown in Table 1.

Columns: “Crude birth rate column (CBR)”: we show both crude birth rate and TFR in cases where TFR was converted by us using a rate. In other cases we usually do not fill the CBR column. There is a quality indicator from 1-5. The conversion key is shown in each line where a conversion from CBR to TFR was applied. The note and remark fields were used to show observation-specific comments, like basis of estimation, information about sources or secondary sources, reference years (e.g. when the source refers to a five year interval).

Sheets: We use a separate sheet to name major sources, and then cross-reference that to the list (“Chesnais” means that in Chesnais’ (1992) book there was a crude rate, which we have used to estimate the TFR). In addition, a web page address is entered directly into the DATA sheet if possible.

[5] These observations were originally included in the Gapminder 2009 dataset.

5 Data quality review and corrections

The purpose of the data quality review was to check the 2009 release of the Gapminder fertility dataset, with a view to improving data quality. When the Gapminder 2009 dataset was compiled, there was little possibility for source critique and in some cases the series broke when there was a new data source. For example, fertility in the 19th century is sometimes low in the statistics because the early censuses and vital registrations had large omissions. Only a more detailed examination could reveal the reasons behind these strange trends in the dataset.

We have therefore reviewed a number of countries in more detail where the data seemed to be inconsistent or incorrect. We have prepared a priority list, containing a number of countries that were due for (1) country reviews and (2) reasonableness testing.

Country reviews: those countries were selected where the Gapminder 2009 dataset shows a strange or unusual trend, or there is a gap in the data. The potential availability of additional information for the given country was also considered in the prioritisation. Four countries have been selected for a detailed country review. Twenty-two countries were selected for data cleaning.

For reasonableness testing we have selected a number of countries that had very limited fertility data in Gapminder 2009 before 1950 and the 1800 level of fertility was estimated to be above 7 children per woman (TFR). The purpose of our test was to see if it was a realistic assumption or not to attribute sustained high fertility levels to these countries throughout the 19th century. The reasonableness testing is done using stable population theory, with the Gapminder population and life expectancy data as an input.

5.1 Background studies prepared

The work has been summarized in a number of background studies, which are shown in the subsequent chapters.

The results of the larger country reviews are shown in chapters 6, 7 and 8. The results of the smaller country reviews ("data cleaning") are shown in chapter 9. The results of the reasonableness testing are shown in chapter 10. The literature quoted in the data cleaning and reasonableness testing sections has been merged with the literature of the other sections above and shown together in chapter 11.

The background studies were successful in finding the causes for suspected problems in the dataset and for some regions (the Caribbean, Latin America, Oceania), valuable additional data were found and processed. There is a considerable amount of regional literature about fertility and in the selected areas and countries we have had an opportunity to review this as well.

We have done reasonableness testing for all countries where the 1800 TFR estimate was 7 or above and if the test indicated a problem, we have deleted the estimates for the given country.

5.2 Call for comments and additional data

We hope that this work is only the beginning and we would be happy to receive any comments or suggestions for corrections. If you have additional fertility data, please do not hesitate to send them to us. We are aiming to process these and include them in a later release of the dataset.

6 Understatement of fertility in official statistics in Sri Lanka and Singapore – a critical review of sources

In this study we are selecting two countries, Singapore and Sri Lanka, where we see a strange trend in the previously published data: early fertility is very low. We have investigated a broader set of fertility literature to try and find the reason behind this and implemented some corrections in the Gapminder 2010 dataset accordingly, as detailed in the rest of this chapter.

6.1 Some basic points about the demography of the region

The pre-transition fertility level of some Asian countries was “moderate”, less than 40 per 1,000. Examples are China, Sri Lanka, Hong Kong and Japan. (Saito, 2006) This phenomenon is related to social practices in the region, such as prolonged breast-feeding, female infanticide, “semi-delayed marriage”. Indirect fertility control methods also included prohibition of widow remarriage and sexual abstinence during certain periods. (Chesnais, 1992, p. 107.)

According to the world fertility survey, Sri Lanka had a relatively long period of breast-feeding, the mean duration being 21 months (Dyson and Murphy, 1985). There are indications that in addition to breast-feeding, poverty and malnutrition may have contributed to low pre-transition fertility in Sri Lanka (Puvanarajan, 2002).

Fertility in developing countries has generally increased before it declined. As living conditions started to improve, some social controls on fertility were relaxed, while individual controls started to gain strength a bit later (Dyson and Murphy, 1985; Saito, 2006). However, this increase is usually short-lived because we see a sharp decline in fertility as economic modernization continues.

In South and Southeast Asia, fertility decline typically started in the 1960s. In most countries the decline was very rapid, about twice as fast as it had been in Europe about a hundred years earlier. Caldwell’s study (2006) measures the speed of fertility decline by estimating the time that was required for an additional 30% drop in fertility after the onset (first 10% drop). According to Caldwell’s results, the same level of decline that took 30 years in England and Belgium was achieved in 20 years in Sri Lanka and 5 years in Singapore.

It is very clear that the main reason behind the speedy decline is the availability and quick adoption of contraceptives and consequently the reduction in marital fertility. A hundred years earlier in Europe, the cost of fertility regulation in a broad sense was much higher, partly because of the lack of contraceptive technology and partly because of moral barriers.

However, the second line of reasoning is less clear. What was the main reason behind the quick adoption of contraceptives? There is agreement in the literature regarding the main factors, but so far it has not been possible to sort out their relative importance convincingly (Hirschman, 2001). Some authors believe that government intervention had a dominant role and family planning programs were very successful in achieving their anti-natalist aims. Some others believe that rapid economic and social developments altered the cost of childbearing and there was a demand for family planning services that these programs filled well, but their importance was only secondary.

A number of characteristics specific to the region contributed to the speed and nature of the decline, in addition to the basic debate about the role of family-planning. Women have a relatively high status in Southeast Asia compared to other parts of Asia, their kinship network is more developed and equal importance is attached to the husband’s and the wife’s families (Hirschman, 2001). On the other hand, government intervention into family life is more readily accepted in Asia than in Europe, due to cultural reasons. There has been no opposition to family planning from Confucianism, Hinduism or Buddhism (Caldwell, 2006). The family planning programs could therefore include some quite drastic measures. For example, the Singaporean family planning program financially subsidized the sterilization of women who were less educated, while promoting the increase of fertility among female graduates (Thang, 2005).

In the literature about fertility transition both Sri Lanka and Singapore tend to appear as peculiar and unique examples. Sri Lanka has one of the lowest fertility rates among the poor countries of the world. Singapore, on the other hand was characterized by rapid modernization and very determined government intervention into fertility decisions at the same time. Albeit in a very different way, both examples suggest that family planning programs had secondary importance, while economic and social changes had a primary role in fertility decline.

6.2 Sri Lanka

6.2.1 Literature review - Sri Lanka

Sri Lanka produced an example of an early fertility decline in the absence of substantial modernization. The country is quoted as an example that government programs and advancement in schooling may trigger a fertility decline despite low economic development (Rammohan, 2004). Sri Lanka also stands out within Asia with its "marriage revolution". Like in Western Europe, the fertility decline started with the postponement of marriage (Caldwell, 1989). In other words, Sri Lanka is an eastern example of John Hajnal's "Western European marriage pattern".

Yapa and Siddhisena (1998) carried out a regional analysis of the relationship between contraceptive use and fertility decline in the 1980s in Sri Lanka. Although it confirms the primary importance of contraception, within that they stress the importance of traditional methods, while modern temporary methods (like pills and condoms) play only a minor role. While 58% of the respondents have used some sort of contraception method, 29.4% use sterilization and 19.2% traditional methods. Only 9.4% of respondents mentioned the use modern temporary methods - mainly young people in their 20s. This gives an indication that in the initial phase of the fertility decline, traditional methods may have been more important than new methods promoted by fertility planning programs.

There is one ethnic group, which showed a quite peculiar fertility development compared to other groups: the Indian Tamils. Puvanarajan, P. (2002) shows that the crude birth rate of Indian Tamils declined very early. It dropped from 41.2 to 33.0 between 1945 and 1953, while the fertility of other social groups increased and remained very low until the 1970s. Puvanarajan puts this down to extreme poverty and malnutrition.

The fertility decline in Sri Lanka could initially be an economic reaction to poverty, using traditional methods and marriage postponement, while later the government played an increasing role especially through progressing the education of women, which made contraception more acceptable and contributed to the increased use of modern contraceptive methods.

6.2.2 Crude fertility data in Sri Lanka

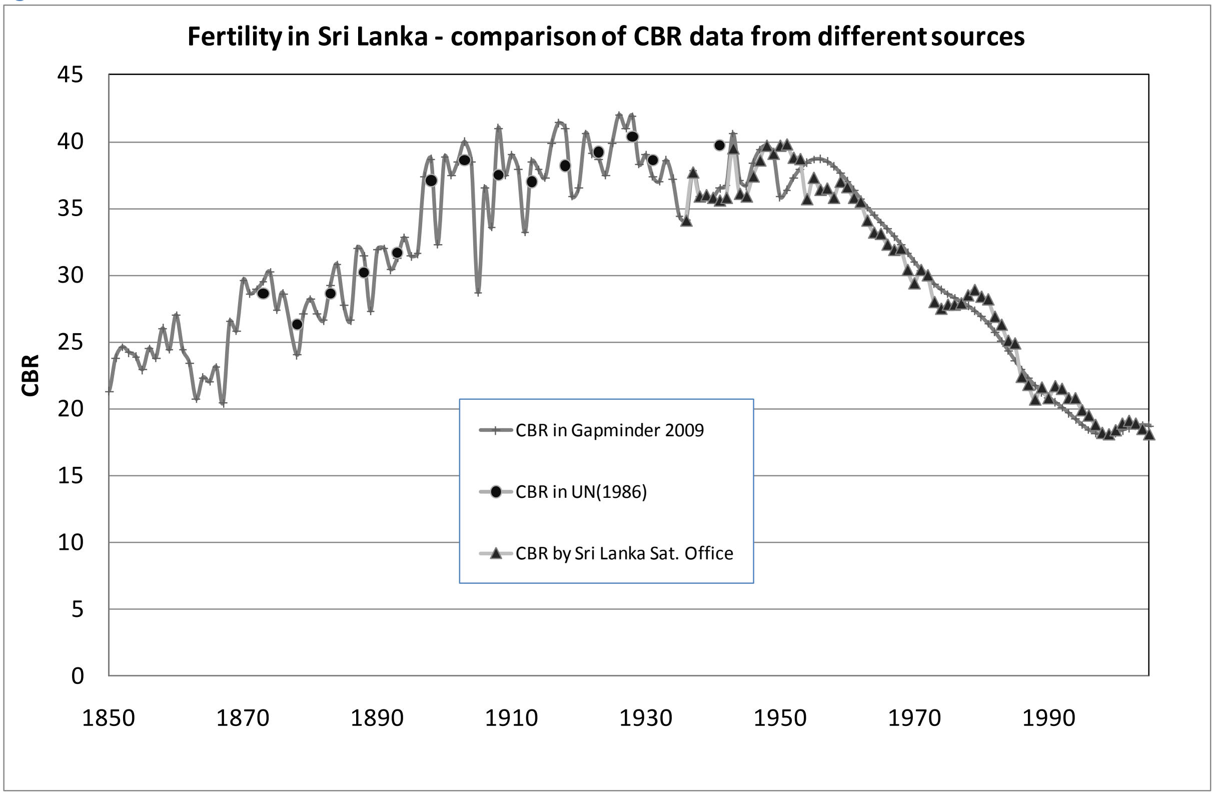
Birth registration in Sri Lanka started in colonial times, in 1867, but until the turn of the century it was unreliable. According to Fernando, the first introduction of birth registration was a failure due to low levels of literacy, communication problems and resistance due to cultural factors (Fernando, 1981). A UN Sri Lanka Study (1986) publication includes crude rates from 1871 for five yearly periods and it also states that crude birth rates before the turn of the century are "in all probability" underestimated. A more recent publication by Langford (2001) includes data for the period after 1900. From 1936, crude rates were available from the web site of the Statistical Office of Sri Lanka.

As a first step we have used several sources to check the Gapminder 2009 dataset crude rate data. We have plotted the following:

- Crude birth rates from UN Sri Lanka Study(1986) five yearly averages
- From 1936, official statistics (Sri Lanka Sat. Office)

The comparison is shown in Figure 8.

Figure 8



The graph indicates that the data published match the new sources reasonably well. There are two problems:

1. Now we know that the pre-1900 data underestimate fertility, we should try to find the omission rate and adjust the data accordingly.
2. The Gapminder 2009 dataset shows a slight deviation in 1950-52, which is likely to be an error in the UN data. These years should be adjusted upwards in my view. We have adopted the figures from the official statistics instead of the smoothed UN data.

Regarding the first point, luckily, in the UN Sri Lanka Study we have found some additional data that help to estimate the degree of under-registration. Firstly, the document suggests CBR=38 as an estimate for the years before 1900 (UN 1968, page 19). Secondly, the same source suggests that child-woman ratios from three censuses may be used to estimate the level of fertility more correctly, despite the high degree of omission in the vital statistics.

We have used the child-woman ratios (CWR) from the three censuses to find out the degree of omission and to adjust our data. Essentially, the level of fertility is estimated from censuses, but the fluctuation is taken by transferring the vital statistics. The data are adjusted in the following steps:

1. For census years (1881, 1891, 1901), we take child-woman ratios from censuses, and convert them to CBR. The conversion is based on empirical conversion factors published by Bogue (1993). Bogue suggests an empirical conversion which is based on the regression of fertility indicators of 152 nations for 1985 on a UN database. As the various fertility indicators are highly correlated (r=0.979 between CWR and CBR), the conversion is possible despite the

conceptual difference between the measures. Bogue calculated the regression coefficients between CWR and CBR and published the result. We use this result in Table 5.

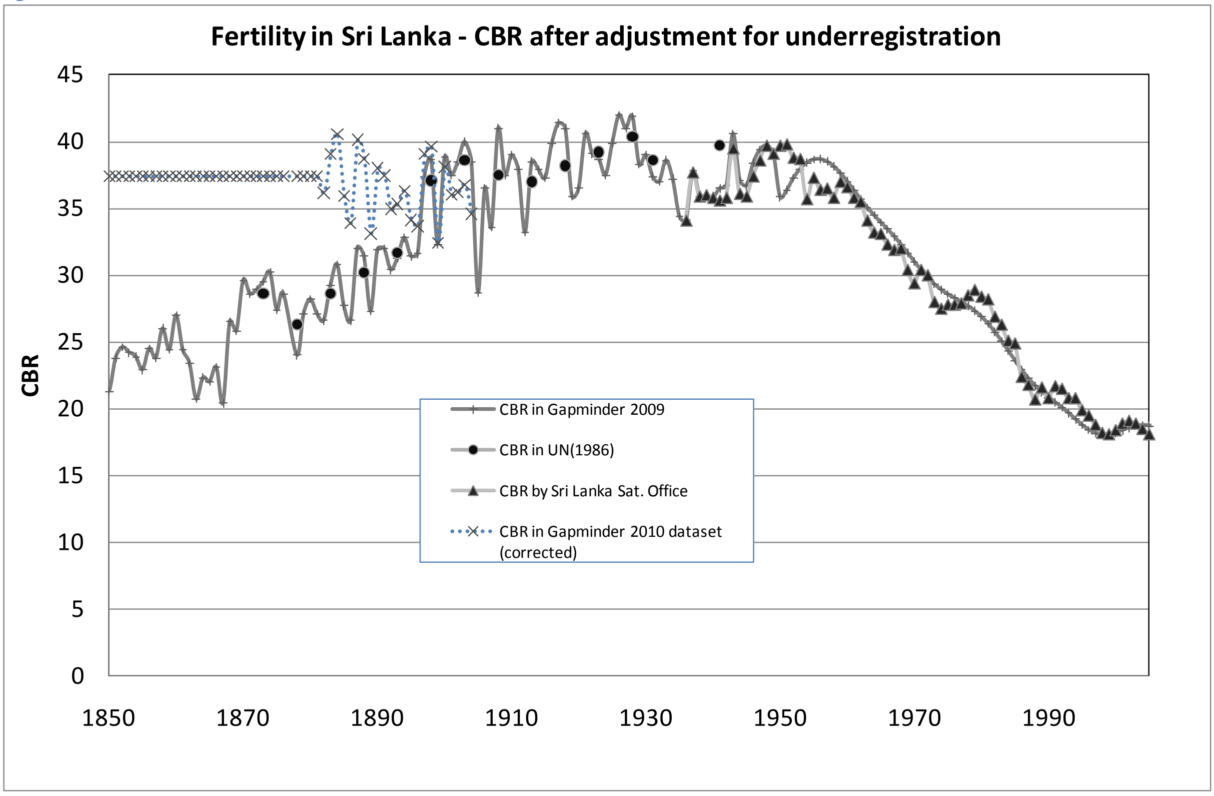
2. Next, we need to estimate the degree of omission. We calculate correction factors for the census years, dividing the estimate based on census (CBR II) by the fertility rates in vital statistics (CBR). These factors are used to correct for the known omission in the vital statistics.
3. We have calculated corrected TFR values from the original data using these correction factors.
4. As the correction factors are roughly in a line, we may assume that the rate of under-registration follows a linear trend (it was gradually improving). We have calculated yearly correction factors using linear interpolation between 1881 and 1901 and extrapolation until 1904, when the correction factor became nearly one.
5. Between census years we have re-scaled the fertility rates (both TFR and CBR) using the correction factors.
6. Before 1881 we did not try to adjust the data, but assumed a constant rate, being equal with the 1881 figure. The reason for this is that we do not have any firm basis on which to adjust the data and the vital statistics were probably very unreliable before 1881.

Table 5

Estimate for rate of omission, using Child-Woman ratios						
CWR-CBR Regression coefficients from Bogue, pg 11.51						
	slope coefficient	intercept				
	b=	*a=*				
	0.05711	-1.89875				
Time of census	Child-woman ratio	Converted to CBR using the above regression coefficients	Original	Correction factor estimate	Original	Corrected
t	*CWR(t)*	*II.CBR*	*CBR*	*conv(t)*	*TFR*	*II. TFR*
calculation		*=a+CWR(t)*b*		*II.CBR /CBR*		*TFR(t)*conv(t)*
1881	776	42.4	27.1	1.57	3.63	5.68
1891	761	41.6	32.0	1.30	4.29	5.57
1901	749	40.9	37.5	1.09	5.03	5.48

The result is shown graphically in Figure 9.

Figure 9

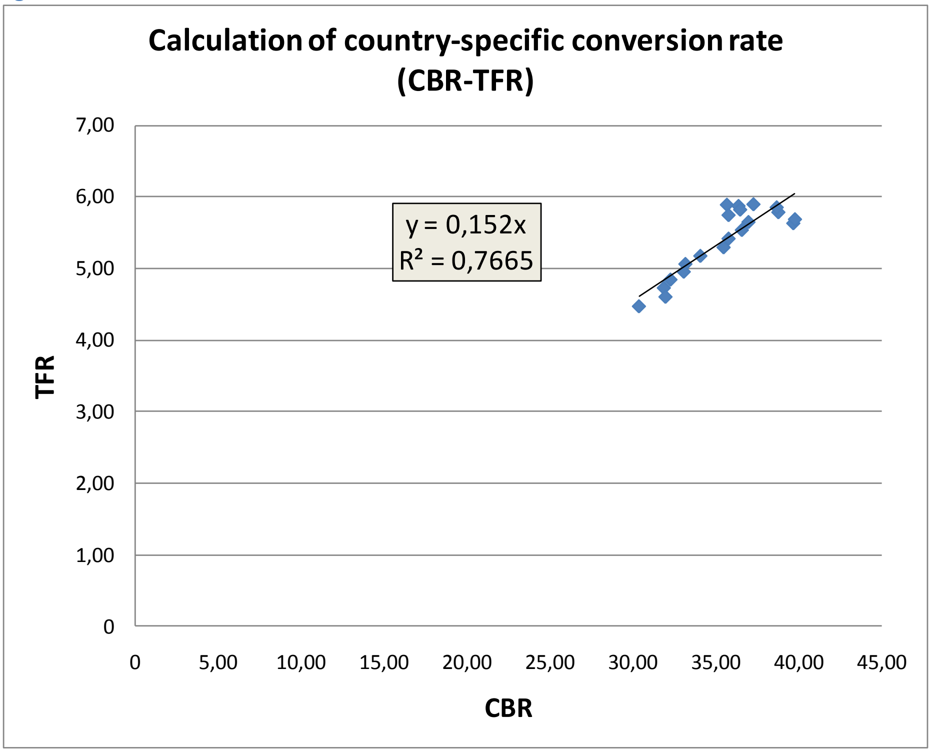


6.2.3 TFR data in Sri Lanka

When turning to TFR, there is a second problem: in 1950 there is a jump in TFR, despite the continuity in CBR. This is due to a different conversion factor (UN around 0.15, Gapminder 2009: 0.134). This can be corrected by calculating a country-specific conversion rate from the UN data instead of the general one that we used in the Gapminder 2009 dataset.

The country-specific conversion rate using regression with 0 intercept from UN data for 20 years between 1950-1969 is 0.152. This is illustrated in Figure 10. (We have taken 20 years as a subjective decision, trying to limit distance in time, while keeping enough observations from which to estimate.)

Figure 10

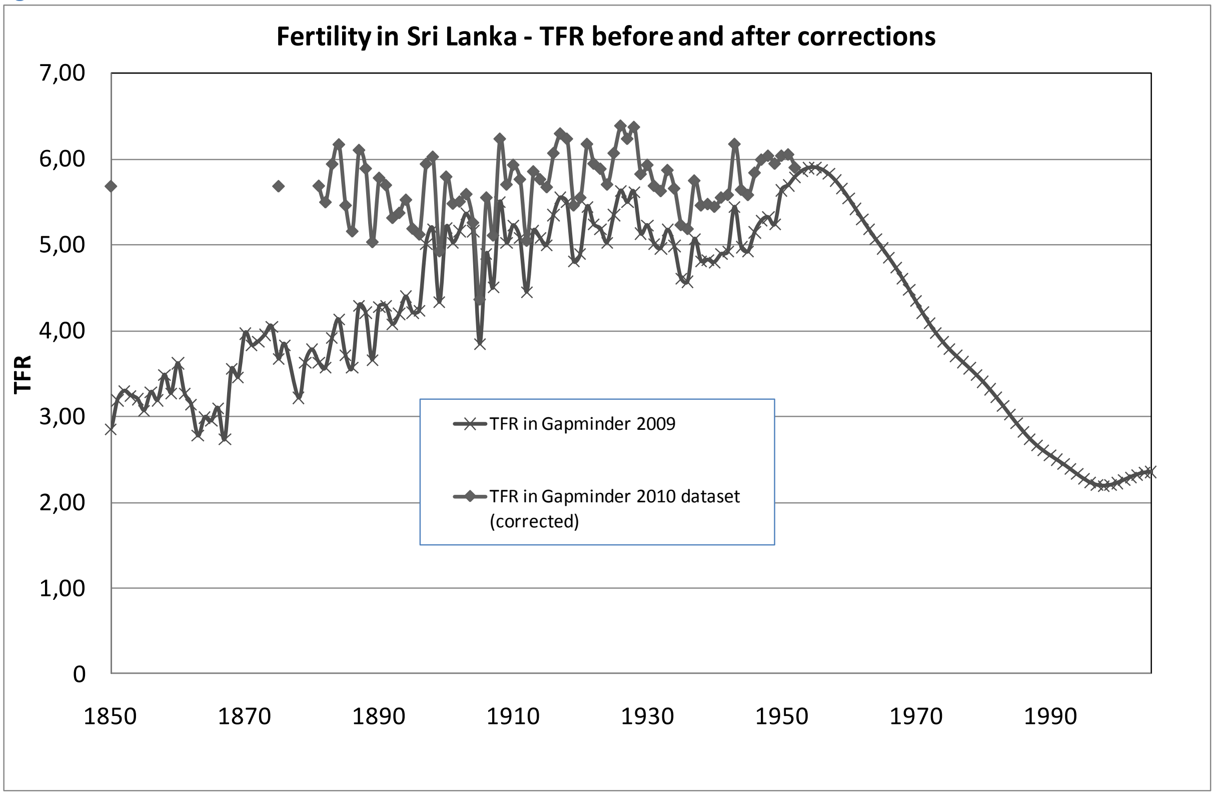


We apply the country-specific conversion rate (0.152 instead of 0.134) between 1906 and 1950[6]. This moves the TFR curve 13% higher for this period.

Finally, a small correction is needed between 1950-52, when the UN data show crude birth rates that differ from the official statistics. Looking at the trend, we think that this must be a mistake. So for these three years we replace the UN data with official statistics, also converted with the country-specific rate. This makes the transition between our data and the UN data quite smooth (Figure 11).

[6] In the Gapminder 2009 dataset, the "old" conversion rate was also used for years before 1905, but the calculation of the first adjustment above was such that it already corrected the problem. (The correction factor was calculated as the rate of the previous TFR and the TFR calculated from the child-woman ratio.) By this we have already corrected the TFRs in years before 1905, in this second step it is enough to adjust the years after that.

Figure 11



On the whole the TFR curve shows a moderately high pre-transition fertility level, fluctuation around 5.8 between 1800 and 1950 and a marked decline after 1950. This is in line with the picture that we saw in the fertility transition literature of Sri Lanka.

6.3 Singapore

6.3.1 Literature review - Singapore

Singapore started its spectacular development as a trading port under British rule, from 1819, when it was a part of the Straigt Settlements. The British East India Company started to develop the island, which has previously been inhibited by small Malay fishing communities. The Island attracted a large number of immigrants of Chinese, Indian and Malay origin. Singapore became an independent state in 1965. It has become a thriving metropolis, one of the most dynamic centers of Asia. The economic growth was spectacular, education has reached European levels. (Fawcett and Khoo, 1980; Hirschman 2001)

If there is but one country in the world where the state really attempted to influence fertility, it was Singapore. There are several reasons why the country is well suited to government intervention into fertility decisions: (1) it is a compact city-state, with high population density (Fewcett and Khoo, 1980), (2) due to its culture, state intervention into family matters is more readily accepted (Caldwell 2006). (3) and the government is very powerful, it controls several aspects of life – like housing and education – which give more room to positive and negative incentives then what is usual in Europe (Thang 2005).

Nevertheless, it is still subject to debate how effective and how successful this intervention was.

Some authors believe that family planning programs had an important role in reducing fertility, while some others are expressing more doubt (Hirchman, 2001). Singapore went through a rapid economic and social modernization process in parallel with the introduction of the family planning program, and it is difficult to sort out the two effects econometrically. Demographers usually believe that government programs rarely have primary importance, and they are more likely to strengthen an existing trend, then to reverse a trend. For example in the case of China it has been shown that in urbanized areas fertility decline was well on the way at the time when the first government measures to limit fertility were introduced (Coale and Chen 1987). The data indicate that the situation in Singapore was similar, and modernization was more important than the family planning program.

The major milestones of policy change were the following (source: Wong and Brenada 2003):

1966 Anti-natalist phase starts (“Two is enough”)
1983 Eugenics (selective policy, supporting fertility of graduate mothers)
1987 New population policy (pro-natalist) – (“Have three, if you can afford it”)
2001 Baby bonus scheme

Without explaining the details of the programs, we would like to note, how powerful these measures were. For example: changed priority in housing, schooling, and lump sum payments to support the sterilization of woman. The government also used mass media and marketing tools, like posters to support the current objectives, inventing simple slogans, like “two is enough”.

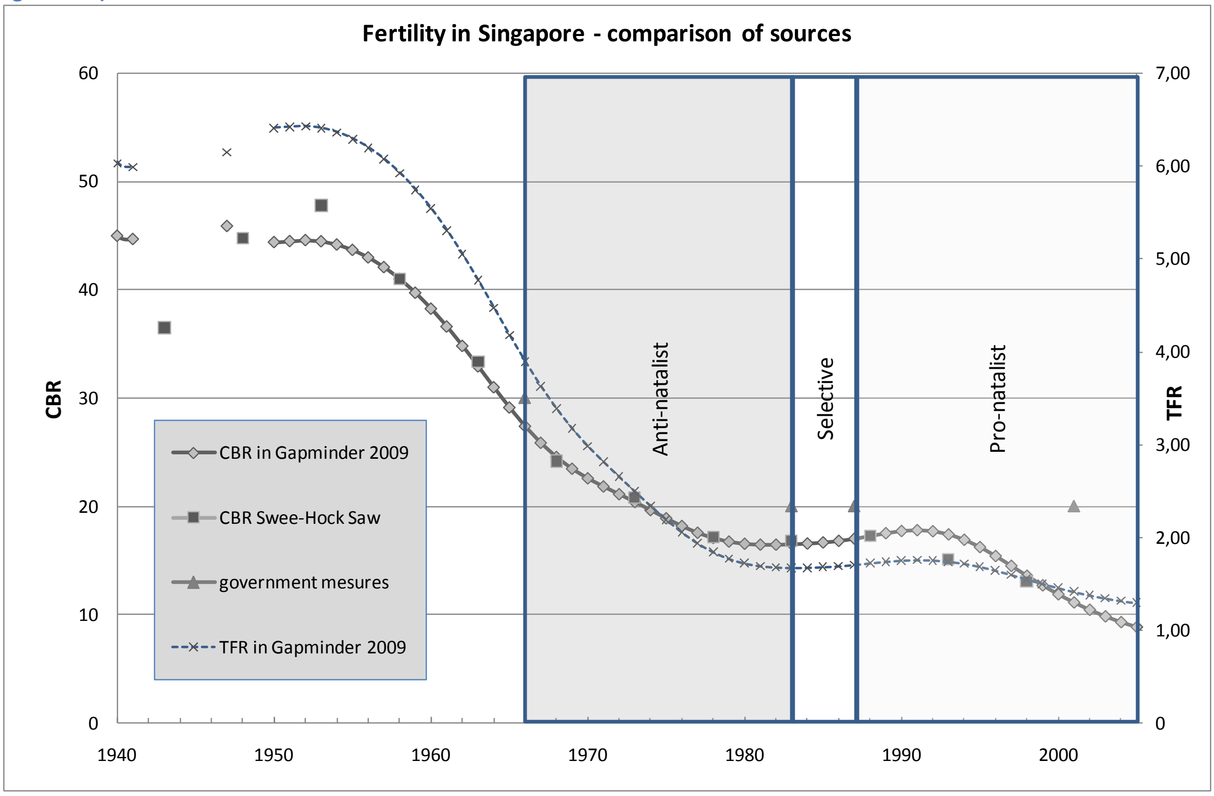
Neville (1978) decomposed the birth rate changes between censuses during the great fertility decline, and he has shown that while the decrease of marital fertility (which was the target of government intervention) was the most important determinant of the decline, “spontaneous” marriage delay also had a significant impact.

We have prepared a graph showing the main period of government policies together with fertility data (Figure 12). It is clear from the graph that the anti-natalist family planning program was introduced about a decade after the onset of fertility decline.

The anti-natalist policy "became a victim of its own success", because it soon became clear, that below-replacement fertility is also a problem (Graham, pg 62).

Moreover, the pro-natalist program was not successful in stopping the downward trend. Thang explains this with the social and economic changes which occurred as a result of modernization of Singapore: high level of education of woman, high level of female employment (Thang, 2005).

Figure 12 - data source: Wong and Brenada 2003, Thang 2005, together with Gapminder CBR and TFR data (TFR is on the right scale)

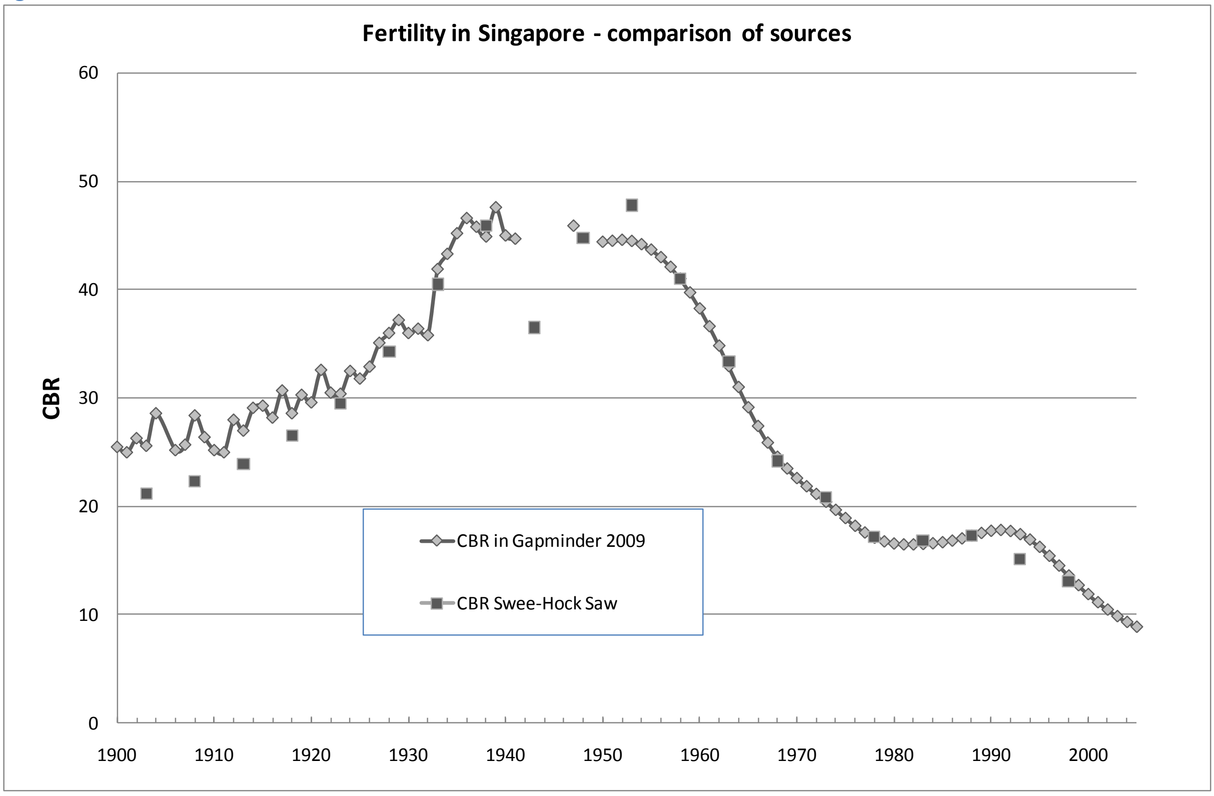


On the whole it seems that the effect of modernization on fertility was so strong, that even a well-designed and powerful government intervention was unable to change the downward trend of fertility.

6.3.2 Fertility data in Singapore

In the Gapminder 2009 set the data were u-shaped, with low fertility in the beginning of 20th century, long increase, and finally decrease. This is unusual. We may turn to a recent book about the demography of Singapore to check this trend (Swee-Hock 2007). The data of Swee-Hock are plotted together with existing Gapminder data on Figure 13.

Figure 13



The author mentions two main problems which may distort the reported crude birth rates: (1) under registration of births, (2) extremely high ratio of males compared to females (Swee-Hock, 2007, Chapter 8, pp 152-153). Although the relative effects of these two are not quantified, Swee-Hock believes that there was no true fertility increase during the first half of the 20st century.

Although Swee-Hock did not himself propose a correction, we may use his data about the sex ratio, to come up with a correction factor. The idea is that our conversion from CBR to TFR assumed equal sex ratio. The TFR calculated this way becomes incorrect in case of Singapore, as the sex ratio is far from 1:1.

It can be shown mathematically, that if the sex ratio is *p*, then the adjustment factor is *af=(1+p)/2*. For example, if the sex ratio is 3:1 (p=3), then we have only half as many woman as normally (25% versus 50%), then the adjustment factor is 2. In other words, the same crude rate corresponds to twice as high TFR, assuming of course that the age distribution is equal.

We have calculated the correction factor for census years in Table 6, and we have applied linear interpolation between these dates in the corrected dataset (see Figure 14).

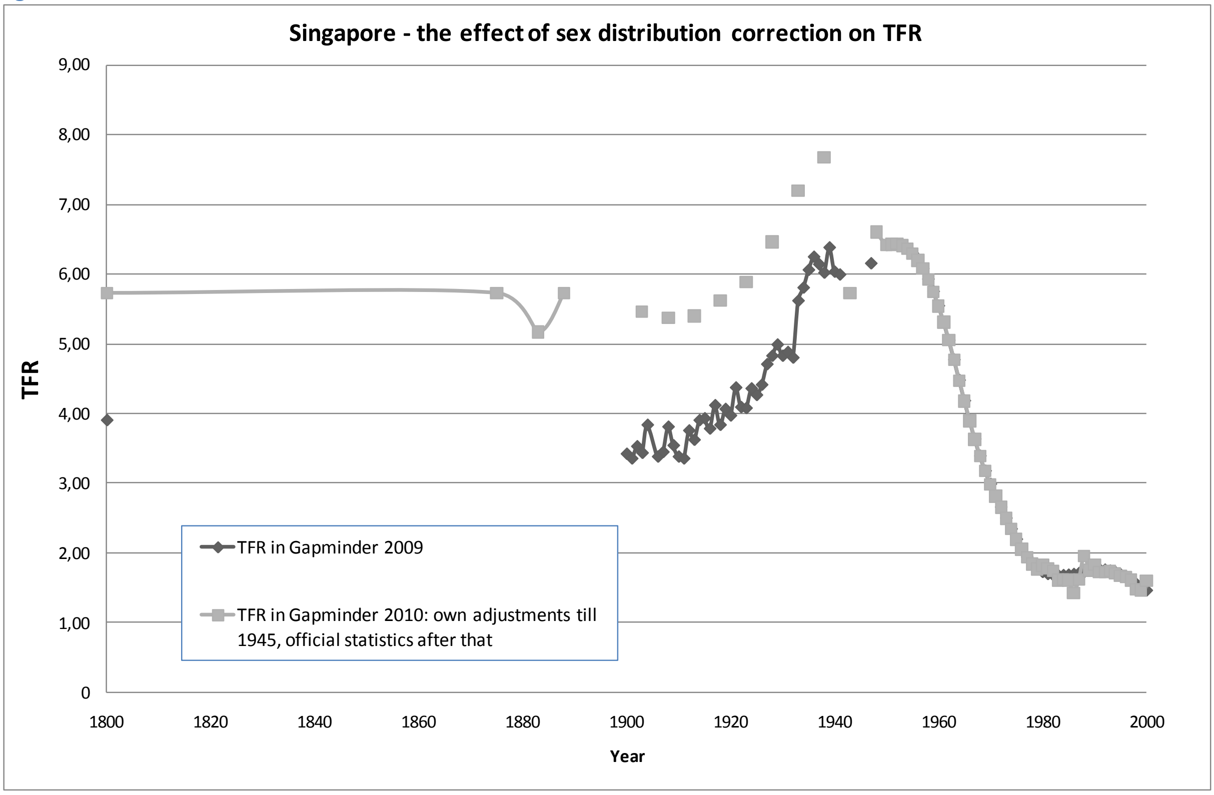
Table 6

Sex ratio in Singapore - calculation of the correction factor

	Sex ratio	Woman / woman assumed	Correction factor
1871	3.189	48%	2.095
1881	3.088	49%	2.044
1891	3.209	48%	2.105
1901	2.938	51%	1.969
1911	2.453	58%	1.727
1921	2.044	66%	1.522
1931	1.713	74%	1.357
1947	1.217	90%	1.109
1957	1.117	94%	1.059
1970	1.047	98%	1.024

Simply speaking, in Singapore around 1900 the fertility was high, but as there were only few woman, crude rates were still low, because the number of children compared to the population was low.

Figure 14



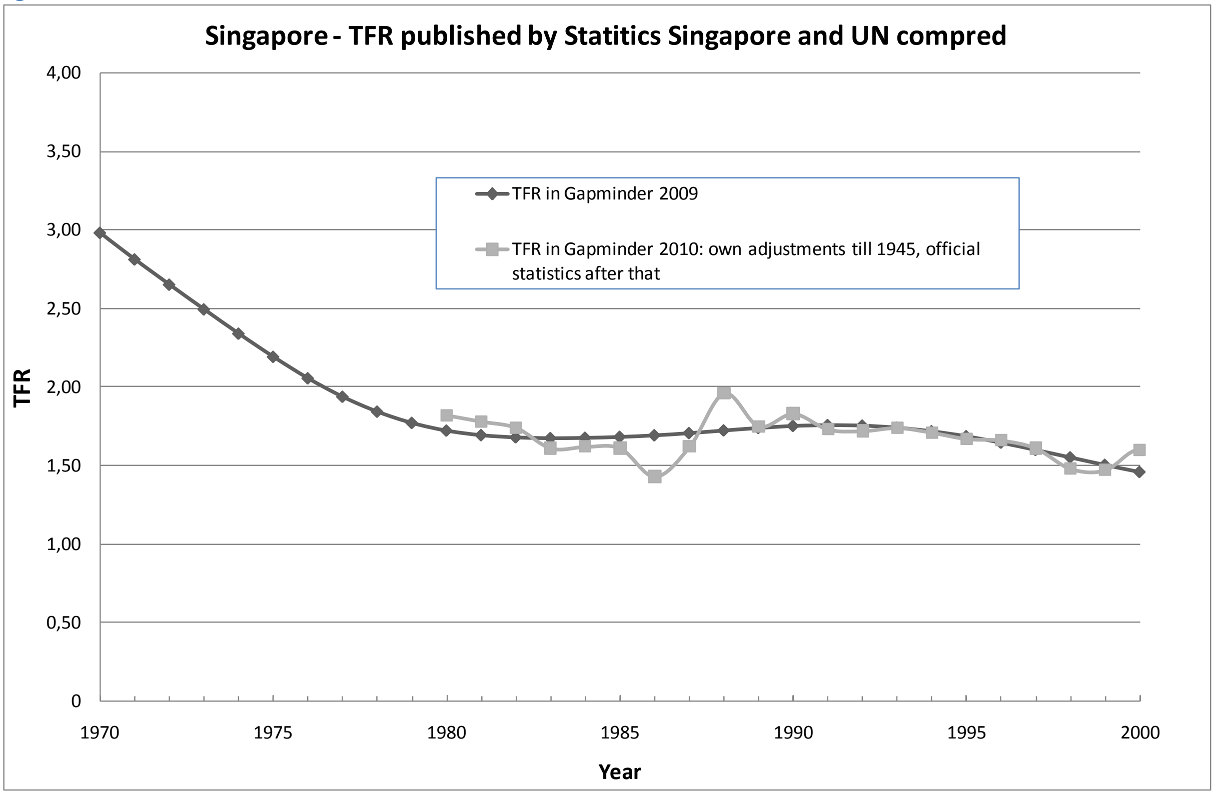
Please note that besides applying the correction factor, we have started to use the CBR data of Swee-Hock (2007) instead of Michell (1998). Given the fact that the two datasets differ (see Figure 14), we are assuming that Swee-Hock is correct, as he is an expert on Singapore's fertility. Mitchell probably included a smaller adjustment for omission, but the book is silent about how it was done.

After the adjustment the U-shaped pattern disappears, although we still see a quick rise in fertility before the decline. A pre transition increase would be in line with the general literature, and it is logical, given high crude birth rates (47-48) combined with still uneven sex ratio (1,7-1,5). But as my Singapore sources do not use TFR before 1945, we could not find anything in the literature that confirms this rise directly.

The large drop in fertility during World War II (1941-45) is confirmed by the literature, as the economic and social conditions were severe during the Japanese occupation (Swee-Hock , 2007).

Regarding the trend of decline, our post-1950 data agree quite well with the new sources. It can be seen that the UN applied some smoothing but the values remain close to recent official statistics (Figure 15).

Figure 15



We have replaced the UN data with official statistics from 1980, because even these small fluctuations may be interesting for some of the readers.

6.3.3 Summary

The purpose of the current study was to investigate, if pre-transition fertility was genuinely so low in Sri Lanka and Singapore, or there is a problem with the data.

In case of Sri Lanka the newly reviewed sources indicate, that in the early colonial period due to low levels of literacy and some resistance the vital statistics had large omissions. The extent of these omissions may be estimated from the child-woman ratios observed on census dates. It is possible to estimate TFR based on this, and comparing it with Mitchell's data, we may calculate a correction factor. After applying the correction factor to our data, the strange pattern disappears: we see that between 1850 and 1950 the adjusted TFR fluctuated at a high level, around 5,8.

A second problem with Sri Lanka was that as UN applied different conversion rate from 1950 then we did before that date, this caused a “jump” at 1950. We have now calculated a new conversion rate, specific to Sri Lanka, and applied that instead of the general one.

In case of Singapore we have concluded that while our crude rate data are correct, the conversion to TFR is incorrect, because it failed to take into consideration the unbalanced sex ratio. Due to the high rate of male labor immigration the sex ratio of the population was around 3:1 during the first part of 20st century. On the other hand our CBR-TFR conversion assumed equal sex ratio. With 3:1 sex ratio, we have half as many women in the population, so we have to double the TFR estimate. When we have done this, we arrived at a TFR curve which is more in line with theory, the strange trend disappears.

In other respects the literature reviewed confirms our data.

This investigation can be also regarded as a case study, testing our data and the methods applied when the large dataset was compiled. It is not a surprise for me that the most vulnerable point is the CBR-TFR conversion. In strict demographic terms, the TFR cannot be determined when we do not know the age-specific fertility rates exactly. The “empirical conversion” (Bogue, 1993) that was applied practically means that instead of TFR we are reporting a rescaled CBR, so it still has all the problems that a crude rate may have: an unusual age or sex distribution will distort it.

Although this problem comes from the nature of the exercise, the errors found in case of Singapore and Sri Lanka indicate, that it would be worthwhile to review conversion rates, and make them more specific to what we know about the country. An easy way to go about this would be to calculate country-specific rate from UN data instead of the general conversion rate in all cases where there is a “jump” around 1950.

7 Fertility in India: A review of estimates and theories

7.1 Literature review

7.1.1 India and the population debate

According to UN projections India will be the most populous country in the world by 2050, overtaking China. (India 1995: 934 Million, 2050 1529 Million; China: 1995 1221M 2050 1478M source: Bongaarts, 2000)

Throughout the 1950s, 1960s and 1970s it was feared that without intervention, some developing countries may be overpopulated, and one of the main examples was India. This “neomalthusian view” in the population debate stressed, that the too fast population growth may cause impoverishment in countries like India, and solution was expected from government measures to control population.

It is not surprising that India has received great attention from demographers. Perhaps the best known among them is Kingsley Davis, a Princeton demographer who wrote several articles and a book on the population of India during the 1940s and 1950s. The neomalthusian view clearly reflects in the writings of Davis about India: “One should not expect,(...) that economic changes will transform the fertility customs of the Indians soon enough to avoid future population problems”(Davis, 1943 pg 78). “Such rapid growth in a country already poor and densely settled cannot fail to be a handicap in the drive toward modernisation, industrialisation, and an increased standard of living” (Davis, 1946 pg 253). Indeed, at the time when Davis wrote his articles, mortality decline in India was well on its way, but there was no marked tendency for fertility decline, so he had a good reason to worry about population growths.

Davis (1944) has shown that between 1872 and 1920 population growth was moderate and fluctuating: every second decade was characterised by minimal population growth, and the long term growth remained moderate. This fluctuation ended after 1920, and after that the growth rate was accelerating to 1-1,2% per year, which was not extraordinarily high, but multiplied by the large existing population the growth in absolute terms was spectacular, for example 83 million more inhabitants between 1921 and 1941. The fluctuations in the growth rate were primarily due to mortality. The sudden increases in the death rate were due to famines, epidemics or wars. For example to major waves of increase were due famines in 1876-1878 and 1898-1900, and a mortality increase in 1918 was due to an influenza epidemic. From the 1920s there is a clear decline in mortality, while the decrease of fertility was delayed and much slower. So the large increase in the population of India was due to the classic scenario of demographic transition.

Acting on the neomalthusian fears, during the cold war mood of the 1950s, it was felt that the Western countries should maintain their influence in Asia by helping these countries to implement “Family Planning Programmes”. India was among the first countries to implement a “family control programme” in 1951. Since then, India has spent 6 Bn USD on its family planning program, and devoted 13% of the total health budget to it (Srinivarsan 1997; McCarthy 1997). The government programme leaned heavily towards sterilisation, providing various incentives both in cash and in kind. In the late 1980s the number of tubectomies annually reached 4,1 Million. Paradoxically, in spite of this, the yearly growth of population in India has accelerated from 1,3% (1940s) to 2,1% (1990s) (McCarthy 1997).

However, this high growth is now mainly due to population momentum. The ratio of young people is high, and fertility is still above replacement, so population growth will continue for some more time. On the other hand fertility decline has already started from about the 1970s (Coale 1988; Yadava and Yadava 1999), and the speed of this decline exceeds the expectations of forecasters (Bongaarts, 2000).

In the late twentieth century there has been turning point in the population debate: neomalthiusian ideas came under strong criticism. The basic idea that too high population density may limit economic growth has been rejected by Boserup, who argued that faster population growth often leads to technological innovation in agriculture (Boserup, 1983). Some others argue that the family planning programs were a “fatal misconception” (Connelly 2008). In India, as well as in many other Asian countries, fertility decline started in some districts well before the family planning program (e.g. around 1940 in the Ludhia district according to Das Gupta, 1995). Generally now we believe that the economic determinants of fertility decline are so strong that only a relatively small government intervention is possible, and even that often had huge humanitarian costs. (Gilbert 2005; Connelly 2008)

7.1.2 The determinants of fertility in India

The fertility decline is undoubtedly on its way in India. Many other things are uncertain, however. Demographers are just estimating the level of fertility, it is difficult to know, which source is reliable, because they sometimes differ so widely. Already in the colonial period the process of collecting vital statistics has had its pitfalls, and there is some argument that since then it has not improved, rather deteriorated (Dyson and Das Gupta 2001). This also affects the conclusions regarding when fertility decline started, and how fast it is. For a given decade, a confusing large number of fertility estimates exist, which significantly differ[7] (Srinivasan 1995).

It is also subject to debate, what is the relative importance of socio-economic development and government measures in fertility decline. In the next two sections shall briefly examine the main arguments regarding the factors influencing fertility in India, before and during fertility transition.

7.1.3 The determinants and stages of fertility decline in India

Studies often compare the determinants of fertility in India with that of historic Europe. India is large and very diverse culturally, so these simple comparisons always include some simplification.

Traditionally in India the age of marriage was very young compared to Europe girls often got married before puberty. By this the main control mechanism of historic Europe, the “marriage squeeze” (fertility control through later marriage) was not available in India. However, this has been largely counter-balanced by a number of social norms, which reduced fertility: long breastfeeding, prohibition of intercourse while nursing, return of a pregnant woman to her parental home to give birth, and the “grandmother” effect, which discouraged childbearing of grandmothers, and prohibition of widow remarriage (Coale 1998, Srinivasan 1995).

It may be argued that the marriage squeeze was also present in India, although in a very different way. While girls were universally married, among males the rate of celibacy was high. To some extent, he distorted sex ratio of the population made this necessary. Due to son preference,

[7] For example Bhat et al (1984) devoted a full book to estimation of vital rates in India during two recent decades, for the period of 1961-1981, but even then they remain to be cautious in the conclusion.

infanticide or neglect has lead to higher mortality among female children. This means that male celibacy was to some extent a necessity. Das Gupta argues that families have consciously used celibacy of their sons to adjust to the economic circumstances, for example if a family had more surviving male hairs, not all of them were allowed to marry, to avoid the sub-division of land. This is evidenced by fact that the ratio of never married man was much higher among landowners (12-22%) than landless (4-5%) (Das Gupta 1995, based on data from Punjab only).

In some Indian casts in North India (like the Rajputs) it was a social custom to give very large dowry with marrying daughters, and this made daughters undesirable, infanticide, neglect or ill-treatment of female children become frequent as a result. In the summary chapter of the book "Asian Population History" Caldwell (2004) stresses infanticide (and neglect) are the main mechanisms of lowering fertility in Asia – as opposed to late marriage in Europe. However, Dyson summarising a book about Indian Historical demography concludes that the main mechanism of pre-transition fertility control in India was probably widow remarriage, because under conditions of high mortality a large part of fecundity was lost this way.

While the relative importance of social controls[8] (lactation, no remarriage, sexual abstinence after birth) is debated, it is very clear, that there existed a traditional fertility regime in India, which resulted in moderate fertility levels due to social controls. By moderate level we mean TFR usually being in the 4-6 region, well below the biological maximum.

7.1.4 Case study evidence for India

Two studies have recently been published for smaller territories in India: Berar and Ludhiana (Dyson 1989, Dyson and Das Gupta, 2001). Detailed, yearly historic vital statistics records are available for these regions, and have been shown to be reliable. This enables the authors to reconstruct the demographic development of these regions.

Regarding fertility, these studies confirm that fertility before the demographic transition (before 1940) was moderately high, in the 4-6 range, and relatively stable. Mortality fluctuated a lot at the same time, growing above 100 per thousand in some years. Life expectancy was 20-30 years, sinking to 15-20 in the worst years (Dyson 1989). Dyson explains this so that in the traditional regime the question was: how to maintain the population in spite of the high mortality. The correlation between mortality and fertility was positive: fertility was lower in times of epidemics.

The traditional demographic regime lasted until about the mid 1940s. In the 1950s and 1960s a pre-decline fertility increase started, probably due to the breakdown of the customs regarding widow remarriage, in Berar the pre-decline increase in TFR was 19%.

[8] By social control I mean that fertility control was indirect, as it was a result of adherence to some social norm, and not the result of an individual, conscious decision. Classic article about this subject is: Lestaege (1980) "On the social control of human reproduction."

7.1.5 Stages of demographic transition in India

Researchers outline the timing of demographic transition in India as follows.

1. Before 1920 India was in the first stage of demographic transition, when mortality is still high and fluctuating (Yadava and Yadava 1999). Davis(1944) assumes that before the first census (1872) the population was close to stationary, as the increase in the good periods were wiped out by the next mortality wave, this implies that the death rates were constantly around the level of 44-47, which is the level measured for the bad decades of 1891-1901 and 1911-1921 respectively.
2. Between 1920 and 1970 there has been a clear decline in mortality, but fertility decline has only started in some special areas, the general level of fertility remained high.
3. After 1970, fertility decline became clearly visible, and it gradually penetrated the lower educated and lower status groups (Bhat 2002; Rele 1987; Yadava and Yadava 1999).

7.1.6 Diffusion was not balanced

Davis argues that the diffusion of the western culture and technology was much easier in case of the practices which effected mortality, than in case of the ones which effected fertility. The colonial rule has made it possible to introduce famine control and epidemic disease control, by external government measures. On the other hand a change in fertility would have required a more fundamental in internal socio-economic development process, which has been delayed, due to "rigid cast system and very otherworldly religion" (Davis pg 265). Due to the semi-colonial economy of India, the development before WWII was not balanced. This has created an unstable demographic situation (Davis 1944).

7.1.7 Economic explanations

Crimmins et al (1984) compares the progress of fertility decline in Karnataka (a state in India) with that of Taiwan, using the Easterlin-Crimmins model as a theoretical framework. They conclude that the pace of decline in the Indian state is "considerably slower than that of Taiwan, probably because of its slower pace of social and economic development". Using the model they argue, that especially in the lower status households in India, the desired number of surviving children is still above the supply (natural fertility minus child mortality), so there is no motivation for fertility control. On the other hand the costs of fertility regulation are initially high. Based on the fertility survey for Karnataka in 1975, they make the following estimates:

Table 7 Quantification of the factors in the Easterlin-Crimmins model for Karnataka in 1975

Natural fertility	6,0
Survival rate	0,764
Potential family size	4,58
Desired family size	4,2
Motivation for control	0,38
Percent controlling	25,6%
Children ever born	5,8
Children surviving	4,4
Birth averted	0,2
Unwanted children	0,2

The forces to increase and to decrease fertility may temporarily be in balance during the early stage of transition. Natural fertility may increase due to better health, education and nutrition, and as a result, even if fertility control starts, its effects may not be immediately visible. However, the authors reject the popular view that India performs badly in demographic terms. The demographic transition gradually progresses in India, it is clearly visible in higher status groups and cities, and gradually penetrates the rural areas as well (Crimmins et al 1984).

The role of education in the fertility transition is analysed in detail by Mari Bhat (2002), and the conclusion is that the desire to substitute child quality to quantity was probably a very important factor. A very large part of the fertility decline was achieved among illiterate woman. Decomposing the TFR change between 1981 and 1991, the author finds that 49% of the fertility change was achieved among illiterate woman 31% among educated woman, and 20% was due to change in educational composition, and parallel with this more and more illiterate woman started to send their children to school.

7.2 Fertility data and sources

It may be a bit surprising, how much uncertainty exists regarding fertility data in India, even regarding recent decades.

7.2.1 Problems with the 2009 estimates

The figures in the Gapminder 2009 dataset were based on three main sources:

- 1880-1908: a study by Saito et al (2005) about two provinces, Punjab and Madras
- 1911-1945: CBR published in Mitchell (1998), converted to TFR using empirical conversion
- UN dataset from 1950

In the Gapminder 2010 dataset we have included a revised series for India, after reviewing a number of additional sources, as it is explained in the current section.

The Gapminder estimate in the 2009 dataset was based on own estimates using Saito (2005) and Mitchell (1998) which originates from official statistics. The estimates of Gapminder 2009 became too low, because too high weight was given in our estimates to Madras (we were using the series in Saito 2005, and calculating a very rough whole-India estimate based on that). Madras is one of three big urban areas of India around 1900 (like Calcutta and Bombay), and as Davis (1946) noted, urban fertility was much lower in this period in India, then rural fertility. On the whole India was still predominantly rural, so the inclusion of Madras with high weight resulted in a too low estimate in the 2009 dataset. However in the Gapminder 2010 dataset we have found new sources for India as a whole, which makes this estimate unnecessary.

The data by Mitchell reflected official statistics, which suffered from large omissions. Davis(1944) calculated estimated fertility rates for the 1881-1941 period, using “Reverse Survival method”. This is probably more reliable than reported fertility rates, because in the initial period vital statistics may have had omissions, but a rate based on age distribution in census is less sensitive for this. Indeed, he estimated CBRs to be about 10 per thousand higher, than what was reported based on vital statistics.

The third source (the UN dataset) is free from such problems it is largely in line recent research papers, although it apparently includes some smoothing, we think it is not necessary to change it.

7.2.2 Newly found data and estimates

Estimates by Davis using reverse survival, 1881-1941

The estimates of Davis (1944, 1951) and converted TFR using the general conversion ratio of 0.134 is shown in table 8. In the first half of project when collection TFR for all countries in the world we used different conversion ratios and it lies between 0.12 and 0.14. So for the purpose of comparison we have calculated both cases where we have 0.12 and 0.14 for conversion ratios.

Table 8 The estimates of Davis and the CBR figures from vital statistics compared

Calendar year	Estimated CBR	From vital statistics (reported CBR)	TFR (CBR*0.134)	Assuming a range of conversion rates (*0.12 ~ *0.14)
1881 – 1891	49	–	6.57	5.88~6.86
1891 –1901	46	34	6.16	5.52~6.44
1901 – 1911	49	37	6.57	5.88~6.86
1911 – 1921	48	37	6.43	5.76~6.72
1921 – 1931	46	33	6.16	5.52~6.44
1931 – 1941	45	34	6.03	5.4~6.3

Estimates by Preston and Bhat

To overcome the weakness of vital registration, from the late 1960s India uses a Sample Registration System to collect fertility data. The article of Preston and Bhat (1984) focuses on cross-checking three different sources and estimates for the 1960-1981 period (table 9). Firstly, they take the data from the Sample Registration System (column two). Secondly, the take the data from two individual surveys (column three). Finally, they make their own calculation from official census data using the Brass method. According to their revised estimate TFR was 5.60 in 1971-72 and 4.69 in 1978.

Table 9 Comparison of fertility data from three different sources by Preston and Bhat

Year	Reported (Sample Registration System) (CBR)	Two nation-wide surveys (TFR)	Best estimate using census data (Preston and Bhat 1984) (TFR)
1971-72	36,75	5,45	5,60
1978	33,3	4,30	4,69
Decline	9,4%	21,1%	16,3%

It is surprising that the three sources give very different estimates for the rate of decline, ranging from 9% till 21%. The fall in the TFR was faster than the fall in the CBR because of both the age pattern of fertility change and growth-simulative changes in age distribution. Some of the difference among the rates of decline can be explained by changing age patterns of fertility. Fertility declines were largest at the older ages, which receive much heavier weights in the TFR than in the CBR. But this does not explain everything.

The main conclusion of the authors is that it is difficult to determine the correct level of recent (1978) child mortality, and the Sample Registration System probably overestimates that, reporting too high fertility for 1978. Their figure (TFR 4,69 in 1978) is more in line with the two major surveys and the census age patterns. So the official data underestimated the pace of decline.

Summary by Srinivasan

The book of Srinivasan (1995) contains two chapters which summarize data from various sources regarding fertility in India. We have used the tables prepared by Srinivasan to compare the sources with Gapminder figures.

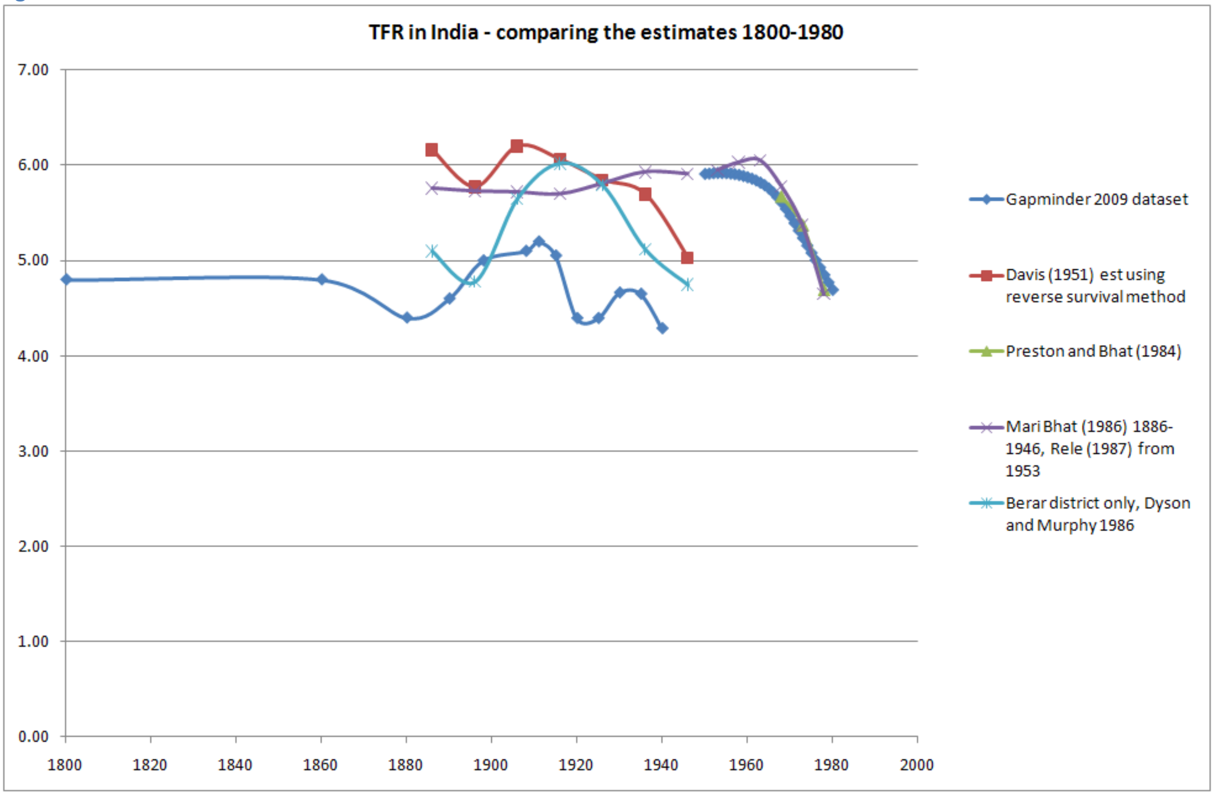
The data from the various sources mentioned above are shown in the following table, putting the years side-by-side. For the estimates which refer to 10 years, we are using the mid-point of the decade, eg. 1886 instead of 1881-1890.

Table 10 TFR data for India from various sources

	Total Fertility Rate (TFR)					
	A	**B**	**F**	**C**	**D1**	**D2**
	Gapminder 2009 dataset	Davis (1951) est. using reverse survival method	Preston and Bhat (1984)	Mari Bhat (1986) 1886-1946, Rele (1987) from 1953	Ludhiana district only, Dyson (1989)	Berar district only, Dyson and Murphy (1986)
1800	4.80					
1860	4.80					
1880	4.40					
1886		6.16		5.76	4.85	5.10
1890	4.60					
1896		5.77		5.73	5.46	4.78
1898	5.00					
1906		6.20		5.72	5.15	5.65
1908	5.10					
1911	5.20					
1915	5.05					
1916		6.06		5.70	5.69	6.02
1920	4.40					
1925	4.40					
1926		5.85		5.81	5.21	5.79
1930	4.66					
1935	4.65					
1936		5.70		5.93	5.31	5.12
1940	4.29					
1946		5.03		5.91	4.53	4.75
1950	5.91					
1951	5.91					
1952	5.91					
1953				5.95		
1954	5.92					
1955	5.92					
1956	5.91				4.54	4.75
1968	5.60		5.67	5.78		
1973	5.23		5.37	5.37		
1978	4.85		4.69	4.65		

This is best summarized using a graph.

Figure 16



Going backwards in time, we see that from now until 1950, the current Gapminder data (based on UN) are confirmed by Preston and Bhat (1984) and Rele(1987).

Before 1950 the estimates widely differ, but it is clear that the Gapminder 2009 TFR data are too low between 1880 and 1950.

In the Gapminder 2010 dataset we have included the values of Mari Bhat (1989) which joins well with the UN series. This way we have reliable data between 2000 and 1880.

To sum this up, the Gapminder 2010 data for India are now mainly based he work of Mari Bhat (1989) which is explained in some more detail in section 7.2.3. We have made some additional changes regarding the pre-transition fertility estimates before 1880, as explained in section 7.2.5.

7.2.3 More details regarding the preferred estimate

The detailed study of Mari Bhat (1989) is available in the book “Indian Population History”, edited by Tim Dyson. The author uses the “variable r” method to reconstruct vital rates from population age distribution. The method applied to “all India” census age distributions. According to the author, his estimates are better than that of Davis, because his assumptions regarding age-specific mortality are more realistic, and this affects all other estimates, including fertility. Davis used the model west life table, derived from the European experience, shifting the curve towards higher mortality. According to Bhat, this gives unrealistically high infant and child mortality. In contrast, Bhat suggests that high mortality countries have a different pattern; it is not enough to shift the European curve. High mortality countries have proportionally higher adult mortality. So instead of the Coale-Demeny model life tables he uses his own estimates for age-specific mortality for India, which in turn lead to different fertility estimates.

7.2.4 The territorial range

Our understanding is that the figures refer to the India according to its current borders: colonial India for the colonial period, including Pakistan.

Mari Bhat's work is based on Mukherjee's (1977) book, which is for the "entire Indian Union"[9]. Bhat himself writes that the estimates are "based on Mukherjee's compendium on Indian age distributions, which incorporate elaborate adjustments for massive territorial changes during 1881-1961" (Bhat 1989, pg 80). But we do not learn more about the nature of these adjustments. The fact that he compares his estimates directly with Davis, indicates that it refers to Colonial India.

Neither Bhat nor Srinivarsan include footnotes for the fertility estimates table regarding how they handled territorial changes. This makes it likely that they report for the territory which was India in the given year.

The estimate of Rele (1987) is clearly for the current territory of India, because this can be seen from the map included on pg 523 of the study.

According to Gapminder's definition the data should relate to the current territory of countries, so strictly speaking an adjustment would be needed for the data provided by Bhat (1881-1950). However this adjustment would require a separate study, which is out of scope for this paper. Therefore we have kept the estimates as they were published by Mari Bhat.

7.2.5 Estimate for pre-transition fertility

It is likely, that our previous estimate for pre-transition fertility was also too low, because it was also based on the Punjab-Madras study. Unfortunately, we could not find a specific expert estimate, although several sources agree that it was well below the biological maximum.

However, it is possible to use the recent levels of tribal fertility as a "proxy", and it is also possible to estimate the likely range for pre-transition fertility.

Tribal TFR in India

An interesting article by Maharatna (2000) reviews the tribal fertility in India. The definition of tribes is not included in the paper, but in India this term has a special meaning: it refers to indigenous people who are outside of the cast system, living in a classless society, and usually speaking a special language.[10] More recently the list of "Scheduled Tribes" in India is recognised in India's Constitution; a total of 645 district tribes. The term "Scheduled Tribes" refers to specific indigenous peoples whose status is acknowledged to some formal degree by national legislation.

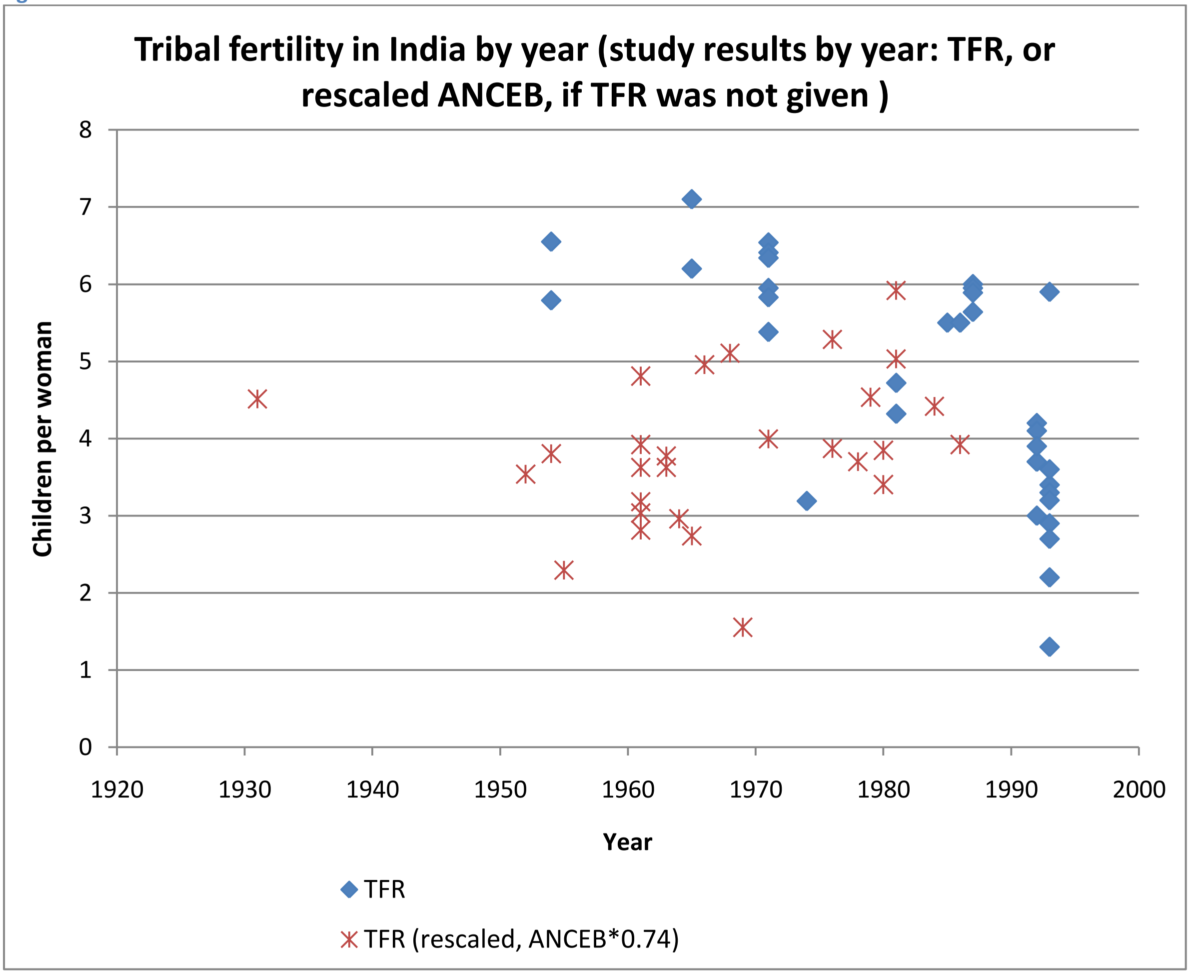
---

[9] According to the Cambridge economic history of India, Volume 2 - By Dharma Kumar, Tapan Raychaudhuri, Meghnad Desai, pg 504 footnote.

[10] "Unlike castes, which are part of a complex and interrelated local economic exchange system, tribes tend to form self-sufficient economic units. Often they practice swidden farming--clearing a field by slash-and-burn methods, planting it for a number of seasons, and then abandoning it for a lengthy fallow period--rather than the intensive farming typical of most of rural India. For most tribal people, land-use rights traditionally derive simply from tribal membership. Tribal society tends to be egalitarian, its leadership being based on ties of kinship and personality rather than on hereditary status. Tribes typically consist of segmentary lineages whose extended families provide the basis for social organization and control. Unlike caste religion, which recognizes the hegemony of Brahman priests, tribal religion recognizes no authority outside the tribe." (source: http://countrystudies.us/india/70.htm, accessed on 2010.08.01)

In the study of Maharatna (2000) a large table is shown, summarizing various studies which measure fertility in tribes. The data contain results from various locations and times, over the 20st century in India. We have prepared a graph, to show, what the most frequent TFR values measured for tribes are. This gives some indication of what could have been the level of natural fertility earlier on (although it is subject to debate, how accurately). Each dot on the graph is one observed population or study at a given time. Different markers are used for studies which report TFRs, and for studies which report average number of children ever born (ANCEB). The two figures differ due to mortality of woman in fertile ages. We have calculated a conversion rate from the more recent studies, where both values were given, and used this to re-scale the ANCEB values for the graph.

Figure 17



This suggests that the typical TFR among tribal population is around 5 – 5,5. Ignoring the re-scaled values a higher estimate would result, closer to 6. But in my view it would be a mistake to ignore those studies which only report ANCEB, as this may introduce a bigger (selection) bias then the rescaling, so my preferred estimate is 5 – 5,5.

Reasonable range

The following table shows what the likely range of pre-transition fertility in India is. The main assumptions are that population growth has been low (2,3 per thousand) and mortality remained in the 40-49 range. This gives a TFR range of 5,3-6,5.

Table 11

Reasonableness estimates for India 1800-1856

Estimation of population growths

	Population	Years	Growth	Growth per year	Life expectancy
1800	200000000				
1856	227000000	56	1.1350	1.0023	

What-if analysis for a range of death rates

conversion rate: calculated from Mari Bhat (1986) for India, using data for 1886-1946

Growth	2.3	2.3	2.3	2.3
Death rate	40	43	46	49
Birth rate	42.3	45.3	48.3	51.3
Conversion	0.126	0.126	0.126	0.126
TFR	5.3298	5.7078	6.0858	6.4638

conversion rate: general

Growth	2.3	2.3	2.3	2.3
Death rate	40	43	46	49
Birth rate	42.3	45.3	48.3	51.3
Conversion	0.134	0.134	0.134	0.134
TFR	5.6682	6.0702	6.4722	6.8742

Based on this, currently the 5,8 – 6,1 range seems to be a reasonable estimate for the 1800-1880 period. The point estimate is TFR=5,95.

7.2.6 Summary

The Gapminder 2009 dataset included too low values for India before 1950, as the sources used were not the best ones. This has been overcome by using the estimates of Mari Bhat (1986) for the 1880-1950 period, and including a new estimate (TFR=5,95) before that. This means that our data are still be estimates, because the exact figures are not known, due to the lack of reliable statistics. The revised data are much more in line with the literature about historic fertility in India, and the various other estimates agree with that of Bhat (1986) and UN in terms of the major tendency.

8 Fertility in China: A review of estimates and theories

8.1 China in the 19th century

General consensus about Chinese demography in the 19th century is that a) the level of mortality in China did not exceed those of Europe before the industrial revolution, b) the living standard for China was not lower than that of Europe, c) TFR was lower than previously thought, either because of infanticide or widespread of sexual restraint (customs) combined with breast feeding. All of these factors would account for long intervals between births. Due to the fact that China was a special social environment, organized of communities in which people are tightly controlled, we may assume that migration was low or negligible. This would attract us to calculate fertility level by using population growth and mortality data. Once we get the CBR data we will get the TFR by the empirical conversion we as have done before. Entwisle (1981) critically reviewed the use of CBR in research papers and concluded there is no reason to disregard research results which are based on CBRs, and this suggests that the conversion from CBR to TFR is a reasonable compromise. However, the "empirical conversion" (Bogue, 1993) that we apply practically means that instead of TFR we are reporting a rescaled CBR, so it still has all the problems that a crude rate may have: an unusual age or sex distribution will distort it.

Lavely and Wong (1998) have assembled three series of population figures for China Proper. The government enumeration of population estimates of China proper are 350 million for 1808 and 430 million for 1850. Agnus Maddison[11]'s webpage[12] provides us with the historical statistics data in which we can find Chinese population in the 19th century. According to this source the estimates are 381 million for 1820 and 412 million for 1850. Although these are just estimates, broadly speaking they give the same result regarding long term population growth. . As Lavely and Wong argued, there is an agreement among researchers regarding the basic trajectory of slowing population growth in the 19th century in spite of the uncertainties regarding the data they are using.[13] We are going to use the data provided on Maddison's webpage (revised in 2010) for population.

Imperial Lineage data's quality deteriorated after 1840 (due to dynastic decline)[14] caution must be taken for the data after 1840. Mortality must have risen from the beginning of Opium War in 1840, in the Taiping crisis (1851-64) due to the civil war and in subsequent uprising and famines, and in Sino-Japanese war in 1894. In this time period population were decreasing. However we have no mortality data for this period. Even Lavely, Wong (1998) have not provided the life expectancy at birth data for the latter 19th century China.

8.2 Use of model life table to see the CDR (crude death rate) level

Based on Lavely and Wong (1998)[15] we may assume that in the first half (at least before 1840) of the 19th century, life expectancy at birth was around 30 to 40 This way we have an estimate for life expectancy at birth but crude death rates for the 19th century are not available. The relationship between life expectancy and crude death rate depends on the population shape. Since we don't

[11] Maddison is Gapminder's data source of population for China
[12] http://www.ggdc.net/maddison/
[13] Lavely, Wong (1998) p.718
[14] Lee, Feng, Campbell (1994) p.396
[15] Lavely, Wong (1998) p.721 Table 3A.

know how the population pyramid for China in the 19th century looks like even if we have life expectancy data we cannot calculate the crude death rate.

However we may build on our knowledge that the 19th century China was not yet industrialized and it had a pre-transition demographic regime. We may look at other pre-transition societies for which we know the population shape. We may assume that the population shape for the early 19th century China would be close to one of these stated below (China for 1950, Japan for 1950, South Korea for 1950, Hong Kong for 1950, Eastern Asia for 1950 or South-East Asia for 1950). If we would assume the population shape and using the model life tables provided by UN (1982), we would be able to calculate the crude death rates as well as age specific death rates. Which type of model life table should we use? UN has provided us with 5 patterns for both sexes separately (Latin America pattern, Chilean pattern, South Asian pattern, Far Eastern pattern and General pattern). Since we don't know which type, we would pick every pattern to see the possible range of values. The life tables are provided for life expectancy at birth ranges from 35 to 75 years old. Since we have already known the level of life expectancy at birth for the 19th century China (before 1840). We had used the life table with the life expectancy at birth of 35 years old. The result will be shown below.

Table 12 CDR (Crude death rate: both sexes combined) for E_0=35

Life table / If Pop shape was	General	South Asian	Far-east	Latin America	Chilean
China (1950)	28.8215	28.4266	37.4397	27.9492	28.9625
Japan (1950)	28.8831	28.5565	37.3943	27.9962	29.0056
Hong Kong (1950)	27.7444	27.7202	36.7901	27.5982	28.2900
South Korea (1950)	28.0943	28.6803	34.8177	27.8548	29.0469
East Asia (1950)	28.7638	28.4297	37.2862	27.9259	28.9437
Southeast Asia(1950)	28.1401	28.2624	35.6210	27.6537	28.5945

We would assume that CDR level was between 27 and 38. We have population data and the bounds of range in which CDRs would lie. Therefore we can calculate the estimates for CBR (crude birth rate) level for those years. As can be seen in the table, using Far-east life table will reveal the highest CDR. Actually Far-east life table looks most close to Campbell and Lee's enumeration[16] in terms of life expectancies at birth, at the age of 10 and at the age of 20. So we would put more emphasis on figures in Far-east scenario. In our view it is likely, that the true value was close to the upper bound (CDR=38), given by the China 1950 population shape and the Far-east model life table: However we may note that even the highest estimate was below 40. Some uncertainty remains however, as we are using 1950 population shape, as that is the earliest date for which data were available in the UN database.

8.3 Calculation for the estimates of fertility level

We would be able to calculate CBR (crude birth rate) if the society is rather closed to migration by the following formula:

CBR=Population growth (/1000 population) – CDR.

CDR level must have been higher in the latter half of the 19th century than that before 1840.

[16] Bengtsson (2004) p.296 Table 10.1

After 1850 it would be unrealistic to assume that CDR level did not change when population was going down. So we would assume 1.1 times higher CDR level for the calculation of higher bound for CBR (that is 38*1.1.This 10% is arbitrary but in the World War Ⅱ Japanese mortality rate was said to increase by 20% and that 20% is too much so 10% could be reasonable regarded as higher bound) but for the lower bound for CBR we would use 27, based on the lower bound calculated above in Table 12, because we are also interested in a lower bound for TFR (assuming lowest CDR leads lowest estimation for CBR). The latter half of the 19th century would have rather broad (unreliable) estimates for CBR data.

In addition to that, we all know that the sex ratio has been skewed in China. We would like to see how much this effect affects the results. Sex ratio of male to female for children age 0 – 4 is 1.2 according to the census of China for 2000. If 1.2 held true for the 19th century China, the discrepancy between 1.2 and 1.05 which is the human sex ratio commonly assumed, would be thought mainly due to infanticide. The correction will be divided by (105+100*105/120)/(105+100)=0.939.

The result is shown below (Table 13). The figures with bold letters are more trustworthy than other figures (As was said before the calculation of CBR based on CDR of 38 would be more trustworthy. Besides CBR data after 1840 is based on assuming 10% increase in CDR in the chaos period of the latter half of the 19th century). TFR#1 is using general conversion that was used in the previous project, TFR#2 is using conversion rate calculated using Chinese CDR and TFR data for 1950-2009, and TFR#3 is using conversion rate calculated using Chinese CDR and TFR data for 1950-1978 excluding post one child policy (1979) period.

Table 13. TFR calculation using three different conversion methods

year	population		CBR		TFR#1		TFR#2		TFR#3	
	N	growth	high	low	high	low	high	low	high	low
1800	320,000	3.8	**41.8**	30.8	**5.59**	4.12	**5.93**	4.37	**6.22**	4.58
1808	350,000	2.6	**40.6**	29.6	**5.44**	3.96	**5.76**	4.20	**6.05**	4.41
1820	381,000	2.8	**40.8**	29.8	**5.47**	3.99	**5.79**	4.23	**6.08**	4.44
1830	409,000	0.3	**38.3**	27.3	**5.13**	3.66	**5.44**	3.88	**5.71**	4.07
1840	412,000	0.0	**38.0**	27.0	**5.09**	3.62	**5.40**	3.83	**5.66**	4.02
1850	412,000	-3.5	38.3	23.5	5.13	3.15	5.44	3.34	5.71	3.50
1860	377,000	-1.9	39.9	25.1	5.35	3.36	5.67	3.56	5.95	3.74
1870	358,000	1.0	42.8	28.0	5.74	3.75	6.08	3.98	6.38	4.17
1880	368,000	1.2	43.0	28.2	5.76	3.78	6.11	4.00	6.41	4.20
1890	380,000	2.0	43.8	29.0	5.87	3.89	6.22	4.12	6.53	4.32
1900	400,000	0.2	42.0	27.2	5.63	3.65	5.97	3.86	6.26	4.05

By comparing the TFR without sex ratio correction for 1800 between our result and TMFR of 6.3 for TMFR by Lee and Feng (1999)[17] our result is consistent to the empirical evidence considering TMFR is bit higher the TFR (not everyone marries and percentage of having children outside marriage is lower than that of married couples).

The result with the sex ratio correction will be shown below.

[17] Lee, Feng (1999) p.48 Table 3. TMFR of 6.3 is based on 3000 sample drawn from Lianing in the time period of 1774-1873

Table 14. TFR estimates after sex ratio correction

year	TFR#1		TFR#2		TFR#3	
	high	low	high	low	high	low
1800	**5.96**	4.39	**6.31**	4.65	**6.62**	4.88
1808	**5.79**	4.22	**6.14**	4.47	**6.44**	4.69
1820	**5.82**	4.25	**6.17**	4.51	**6.47**	4.73
1830	**5.47**	3.90	**5.79**	4.13	**6.08**	4.33
1840	**5.42**	3.85	**5.75**	4.08	**6.03**	4.28
1850	5.47	3.35	5.79	3.55	6.08	3.73
1860	5.69	3.58	6.03	3.80	6.33	3.98
1870	6.11	4.00	6.47	4.23	6.79	4.44
1880	6.14	4.02	6.50	4.26	6.82	4.47
1890	6.25	4.14	6.62	4.39	6.95	4.60
1900	5.99	3.88	6.35	4.11	6.67	4.32

To sum this all up, while the estimates show a broad range of possible values, our best estimate is shown by the TFR#3 column in Table 14, and we think that the likely values are close to the upper bound.

8.4 Conclusion

We have estimated TFR for the 19th century China. Based on the calculations we assume that TFR for China was around 6.0 all over the 19th century. The preferred estimate is shown in the TFR#3 column in Table 14. In addition to that the lowest bound was estimated to be around 4.0. However, due to the dynastic decline in the latter half of the 19th century the data are very unreliable. Although we have calculated an estimated TFR for this period as well, the estimated range became too wide.

On the whole, we could not find additional sources which would enable more accurate estimates, and the reasonableness calculations performed have not rejected the data that we have included in the Gapminder 2009 set. So we kept the dataset unchanged. However, it will be worthwhile to review the dataset in a few years time, because the historical demography of China develops quickly.

9 Smaller country reviews (data cleaning)

The Gapminder 2009 dataset has been reviewed in several steps as it has been explained in chapter 5, and the corrections are now included in the 2010 data. In the previous chapters we have shown the work done in cases where a more detailed country review was felt to be necessary. The current chapter explains the “data cleaning”, the work done in case of countries where we only identified a smaller problem with the 2009 dataset. The countries for this chapter have been selected using a systematic review of the whole 2009 dataset. The Gapminder 2009 dataset sometimes showed a strange or unusual trend, or there was a gap in the data when we switched from one source to another. We have prioritised all countries, considering the size of the problem perceived, and the potential availability of additional information. Finally, twenty-two countries were selected for data cleaning.

The work done is shown in a tabular form in Table 15.

There is one row per country. The second column describes what the problem was with the data for the given country. The third column explains the work done in the 2010 revision. The last column refers to a detailed literature list that can be found at the end of this paper.

The dataset now includes the outcome of this investigation. The revised dataset looks the same way as before. Only one new column was added, which contains y2009 or y2010, the year when the data was complied. We have added the new sources to the source list, and referred to in the row where it was used. We have adjusted the data quality indicators. The new data can be found using a filter for this field.

Table 15

Country	Concern regarding the Gapminder 2009 data	Work done	additional literature or source
Canada	maybe some more data exist	Additional literature was found. McInnis published If (Princeton) indexes by region (Quebec had much higher fertility then the rest of Canada), from what I have calculated a weighted average for 1861. The 1891 figure was published for the whole Canada (pg 409 in the text), so I did not need to calculate weighted average. After finding these If values, I have used the usual conversion from Princeton index to TFR. I have reviewed the 1800 estimate in light of these new data, extrapolating backwards, assuming the same fertility as in 1861. I have also used CBRs from McInnis for 1871 - 1881. Between 1900 and 1921 I replaced the Chesnais data with the more recent values in McInnis, pg 547. The crude rates were converted using country-specific conversion rate to TFR. For 1931 the two sources agree, I kept Chesnais.	McInnis M. 1997
Taiwan	CBR-TFR conversion not country specific	Reviewed by a classmate from Taiwan. Generally OK. Additional source was found for 1956-1976, and data for this period slightly changed. I calculated new conversion rate from recent values (1956-1971, as much higher than the general one: 0.145. The earlier TFR values are recalculated accordingly.	Statistical Office
Japan	OK, but a bit more data for 19th century?	New source was found, containing Princeton indices 1700-1935. I use this now to add some data points for 19th century. Data are given for intervals: I have used the midpoint. The general picture is unchanged: surprisingly low pre-transition fertility in Japan, similar to England. I have deleted the 1820 and 1855 estimates, used in the previous dataset, because that was only based on two villages.	Tomobe K. 2001.

Table 15

Country	Concern regarding the Gapminder 2009 data	Work done	additional literature or source
Poland	Some more data? (weak before 1920)	With the help of Statistical Office of Poland, I found a new source with yearly CBR data from 1816-1910. Although the data are only for Central (and 1816-1823 Western) Poland, I used these without any change, because I do not know the weights. This gives a good indication of a gradual decline - fits well with other sources.	Wojtun 1983
Mexico	1895-1899 too low TFR from Chesnais, I seem to remember that it is some war, probably OK	Additional literature was found. Pre-1900 lines were deleted, as probably unreliable. The literature contains expert estimate for pre-transition fertility in Mexico for 1900 (TFR 6.8) the previous estimate for 1800 has also been replaced with this figure.	estimate for 1900 by McCaa 1997
Slovak Republic	jump between 1935 and 1950 Sprocha-UN	Received very nice time series from Slovak stat office, and updated the data. New data added for the 1945-1950 period, showing a gradual fertility rise, which indicates that both Sprocha and UN were correct.	Statistical tables received from Statistical Office of the Slovak Republic

Table 15

Country	Concern regarding the Gapminder 2009 data	Work done	additional literature or source
Panama	fluctuation 1938-1941 (wartime) otherwise OK	Several new sources were found. Trussel(1980) for 1970s, which still only estimates fertility based on a survey. Palloni (1990) publishes additional data for 1900, which agrees with Collver. I added and revised estimate (slightly upwards) for 1800. I have replaced Chesnais with a more region specific source, Collver 1965 book, pg 155. The figures in Collver are his estimates considering under-reporting.	Collver 1965 Trussell and Hill 1980 Palloni 1990
Trinidad and Tobago	May be OK, but slightly strong fluctuations.	New data for 1814 in Haines, chapter by Engerman. The author suggests that fertility was low in the Caribbean among slaves (lower then US slaves), because of low living standards and unhealthy conditions especially on sugar cane plantations. I use the estimate for slaves for the whole population, because the prop of slaves was 8:1 in 1830, see pg 496.	Engerman 1997 Higman 1995 pg 308
Guatemala	1930 outlier: 7.6	Difficult to get further info. The UN values for 1950s (CBR above 50) are confirmed by another study by Brea (Pop. Dinamics in Latin Am.), and similarly high values seen in neighbours. Still 56.8 CBR (from Chesnais) appears too high. Deleted Chesnais and used Palloni instead.	Brea 2003

Table 15

Country	Concern regarding the Gapminder 2009 data	Work done	additional literature or source
Fiji	UN and Mitchell data contradict, jump in 1950, too few data	Bayliss-Smith, pg 152: the vital registration system was incomplete. But his data start only in 1960. The table of Edge (1937) confirm the current data for 1935 Fiji. (Fertility is probably understated, but I do not have enough data to change it now) (Fertility is probably understated, but I do not have enough data to change now)	1. Bayliss-Smith T 1988. 2. Edge 1937.
Tunisia	low values 1947-48, why? Chesnais<->UN	The UN statistical yearbook is the same as Chesnais for 1946-1948. The UN document "The state of demographic transition…" includes a one page review of Tunesia, which confirms very high CBR around 1950 (they write that CBR was "around 50") and marriage was early. This makes it probable that the 1946-1948 values were under-recorded, but as this is the official statistics, I prefer to keep it. The 1800 estimate appears to be realistic. No change.	UN 2001. The state of demographic transition in Afrika
Korea, Dem. Rep.	strange drop in the UN data, Rele before 1950 is probably OK	North Korea is a mystery, but it is probable that the UN data are wrong in this case. Calwell estimates that the fertility decline started in the 1975-80 period, and before that TFR was 7.0 -- this contradicts with the UN. He also gives different estimate for 1995 TFR: 2.4. I found an alternative series in Kandlah(2003), and compared it with the 5-yearly averages of the UN series. This shows that until 1970 it is worth replacing the UN series with Kandlah.	1. Caldwell and Caldwell 1997 2. Kandlah 2003

Table 15

Country	Concern regarding the Gapminder 2009 data	Work done	additional literature or source
Cyprus	jump betw. actual and estimate, estimate may be too high	Could not find additional data - but estimate referring to Greece is reasonable.	
Korea, Rep.	strange drop in the UN data, Rele before 1950 is probably OK	The UN series show a big fertility drop, to avg. 5.26 level in 1951-1955 while in other sources do not conform this fully, although evidently there is a drop, maybe not as deep. Eg. Kandlah shows 5.4 (1951-55) Rele shows 5.79. The onset of decline is dated 1960-65 by Caldwell. I have added one more row for 1948 with Rele's estimate, and deleted the three lowest lines (1950-52) from the UN, to correct for the too deep dive, keeping all other data unchanged. The new sources differ a bit from the old ones (Rele, UN), but after this correction not s significantly.	1. Caldwell and Caldwell 1997 2. Kandlah 2003
Croatia	gap in the data 1924-1950	Three more values 1947-49 were added. The stat office said they do not have more.	
Egypt, Arab Rep.	outlier 1961, why lower then UN?	Chesnais data were checked against UN 1954 YB: roughly the same. UN from 1950 includes 6% adjustment upwards. I applied the same for Chesnais data (106% of published CBR). TFR-CBR conversion rate was similar to UN after 1950, so I left it unchanged.	

Table 15

Country	Concern regarding the Gapminder 2009 data	Work done	additional literature or source
Albania	very few data, guess in 1800, but much better is not expected	UN 1953 yearbook is about the same as what we already have (Chesnais). Hofsten's article is specifically for Albania's demography, but the first date is 1938, for which CBR agrees with ours. Hofsten concludes that there was a pre-transition fertility increase in Albania. I have modified the estimate for 1800 and 1925 to 4.6 from 5.6, assuming that low fertility was genuine in 1938, and not-under-reporting, as it was previously assumed.	Hofsten 1975 pp. 147-158
Malaysia	few data, but, limited avail. of further data	UN 1953 yearbook is about the same as what we already have (Chesnais). Out of three additional sources reviewed (Hirschmann, Davanzo, Kandlah) (journal articles) two indicate lower fertility for 1950 then UN, one indicates about the same. None of the new sources extend before 1945. Saw writes that the data do not go back before 1930. The book contains a time series after that, which is nearly the same as Chesnais. The current data are confirmed, no change was justified.	1. DaVanzo and Haaga 1982. 2. Hirschman 1980 3. Saw, Swee-Hock. 2007
Barbados	outlier 1915, but possible	Replaced the strange value with a 1916 value in Engerman, this solves the problem.	

Table 15

Country	Concern regarding the Gapminder 2009 data	Work done	additional literature or source
Slovenia	too few data, but OK	Turned to Stat office of Slovenia. There exists a publication about historic fertility in Slovenia, with yearly CBR series from 1857. The new data now replace the previous Princeton data. A country-specific conversion rate was determined from the UN data, and this rate is used to get TFR.	Rodnost v Sloveniji od 18. do 21 Stoletja (2006); Tabela 97 pg 274. Statistical Office of Slovenia, Ljubljana
Venezuela, RB	large pre-transition increase or omission in the data	Several sources (Collver, Stycos) agree that there was an increase of fertility in many Latin American countries during the 20st century. Collver (1965) calculated adjustment factors and published corrected CBRs, adjusting for under-reporting. The result still shows an increasing fertility after 1935 in a number of Latin American countries, including Venezuela. Collver also calculated standardised rates, which filter out changes in age distribution and sex distribution, and this gives a more flat curve: so the fertility increase was mainly due to favourable changes in the age composition. (this would not normally effect TFR only CBR, but the way we calculate it it does) But he also notes that behind this stability there was a balance of two forces: increasing marital fertility and marriage postponement. Included new values from Collver's 1965 book from 1900-1947. Using Collver meant an adjustment upwards, so probably earlier figures of Chesnais are also too low. I have deleted Chesnais figures.	1. Collver and Arevalo 1967 2. Stycos 1978 3. Collver 1965 4. Palloni 1990
Guyana	large pre-transition increase or omission in the data	The pre-transition increase is in line with the regional literature. The conversion rate was set to be the same as what the UN is using – this eliminates a big part of the jump in 1950, because the different conversion caused about 10% shift in the 2009 dataset. But even in CBRs Engerman shows an increase for Guyana between 1906 and 1950, from 36,7 to 43. This is now roughly in line with what we have. I took one value from Engerman (for 1906, pg 513), and used a specific conversion factor (0,155 instead of	Engerman 1997

Table 15

Country	Concern regarding the Gapminder 2009 data	Work done	additional literature or source
		0,134) and changed the 1800 estimate upwards.	
In addition:			
Bolivia Colombia Ecuador El Salvador Guatemala Honduras Panama Peru Venezuela		Found additional data in Palloni (1990) and Collver (1965) for these countries between 1880 and 1950. I have added these and modified the 1800 estimate accordingly, where it was necessary. I have reviewed all Latin American countries, comparing our data to Collver's series (eg. For Argentina and Cuba our series were confirmed)	

10 Reasonableness testing

A reasonably narrow interval for TFR can be given for countries where we know that there was no significant immigration and growth was very slow.

A very simple approach, using the basic idea behind demographic reconstruction methods is as follows:

Table 16 Death rate – TFR relationship using some simple assumptions (rates are per thousand population)

Growth rate (r)	1	1	1	1	1
Death rate (CDR)	39	41	43	45	47
Birth rate (CBR)	40	42	44	46	48
Conversion	0.134	0.134	0.134	0.134	0.134
TFR	5.36	5.628	5.896	6.164	6.432

The results suggest that assuming very slow growth (0,1% per year) the reasonable range of TFR is for pre-transition populations with high mortality (39-47 per thousand) is in the range of 5,5-6,5 (rounded up).

This range of death rates (39-47) is reasonable for a pre-transition population in my view. For example we have looked at India and Mexico.

In India, for the decades with the worst epidemics, death rate was 44 (1891-1901) and 47 (1911-1921) (Davis 1944). In better (lower mortality) pre-transition years it was 41 (1881-1891). For Mexico, there are various estimates for the pre-transition death rate for the decades 1900-1909 and 1910-1919, but they range between 34 and 44,6 , while the life expectancy is estimated to be 25-28 years at the same time (Haines and Steckel 2000 pg 604-607). Note that the Gapminder estimate for life expectancy in Yemen is 33 years from 1800-1913.

The big question is, if we may exclude the possibility of higher then 47 death rate persisting over decades or not. If we may exclude that (we think so) then the estimates above 6,5 are probably too high for populations which are not growing substantially.

10.1 Empirical approach to find the long term maximum death rate

We have collected "worst practice" mortality rates from Mitchell (2003). What was the highest recorded mortality experience for a 5 year period?

Table 17

Some examples of high mortality periods from Mitchell (2003)
(Crude death rates)
Note: going through the pages of vital statistics, I identified the "worst"
five yearly period, in order to get closer to what is the highest
level of mortality which may be sustained

		CDR	Yearly CDR				
			1	2	3	4	5
AFRIKA							
Egypt	1929-29	26.44					
Mauritius	1900-04	36.54	36.5	40.4	33.8	40.1	31.9
South Africa Coloureds	1940-44	23.86	22.7	23.8	24.6	23.8	24.4
ASIA							
Singapore	1900-04	40.38	40	43	39.4	39	40.5
Sri Lanka	1905-09	30.92	27.7	35.1	30.7	30.1	31
Taiwan	1906-09	32.2		33.4	32.4	31.9	31.1
OCEANIA							
Hawaii	1870-74	56.32	64.6	60.2	53.2	53.4	50.2
SOUTH AMERIKA							
Chile	1850-54	35	(Collver's estimates)				
	1855-59	34.7					
Venezuela	1885-89	32.3					
	1890-94	28.2					
Guatemala	1905-09	35.4					
	1915-19	40.8					
Argentina	1860-64	31.7					
Trinidad and Tobago	1885-89	29	28.7	27.9	27.3	31	30.1
Barbados	1920-24	33.14	33.4	43.3	22.4	37.1	29.5
EUROPE							
Russia	1870-74	37.16	35	37.9	41.2	36.5	35.2
Hungary	1870-74	44.54	33.5	40.1	42.9	62.9	43.3
Finland	1805-9	38.4	21.2	21.9	29.2	60.5	59.2

With the exception of Hawaii, in all cases the five year average remains below 50, and 35-40 death rate is already very high, such high death rate rarely occurred in the periods which are covered by vital statistics. As some may argue that high mortality correlates with no vital registration, I briefly looked at England, based on parish registers using the work of Wrigley and Schoefield. In fact, the same is also true for the CDRs reported for England in the 17th century: CDR fluctuated between 20-35 most of the time, only rarely did it go above 40, and then only for a single year.

Naturally, this could be investigated much more thoroughly. The selection of countries here introduces a bias: it is likely, that the existence of records correlates with mortality level. It is possible, that the decease environment in Africa was more severe. But even so the above gives some indication regarding the reasonable level of long term death rates in pre-transition societies.

Natural fertility

Since Henry, demographers distinguish natural fertility and controlled fertility populations. Controlled fertility means that couples make a conscious decision to limit family size. To detect the existence of conscious control, demographers usually look at the signs of “stopping behaviour”: when fertility drops after the desired number of children is reached Wood (1990). Wood summarized the range of human fertility as shown in the following figure.

Figure 18

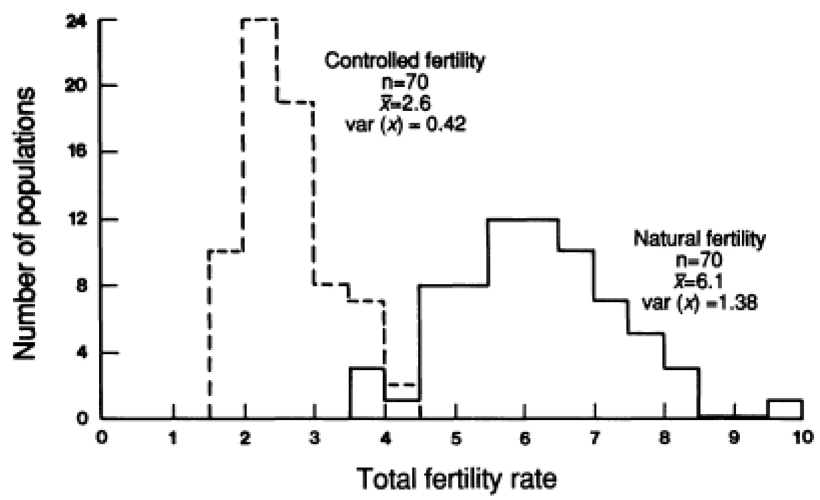


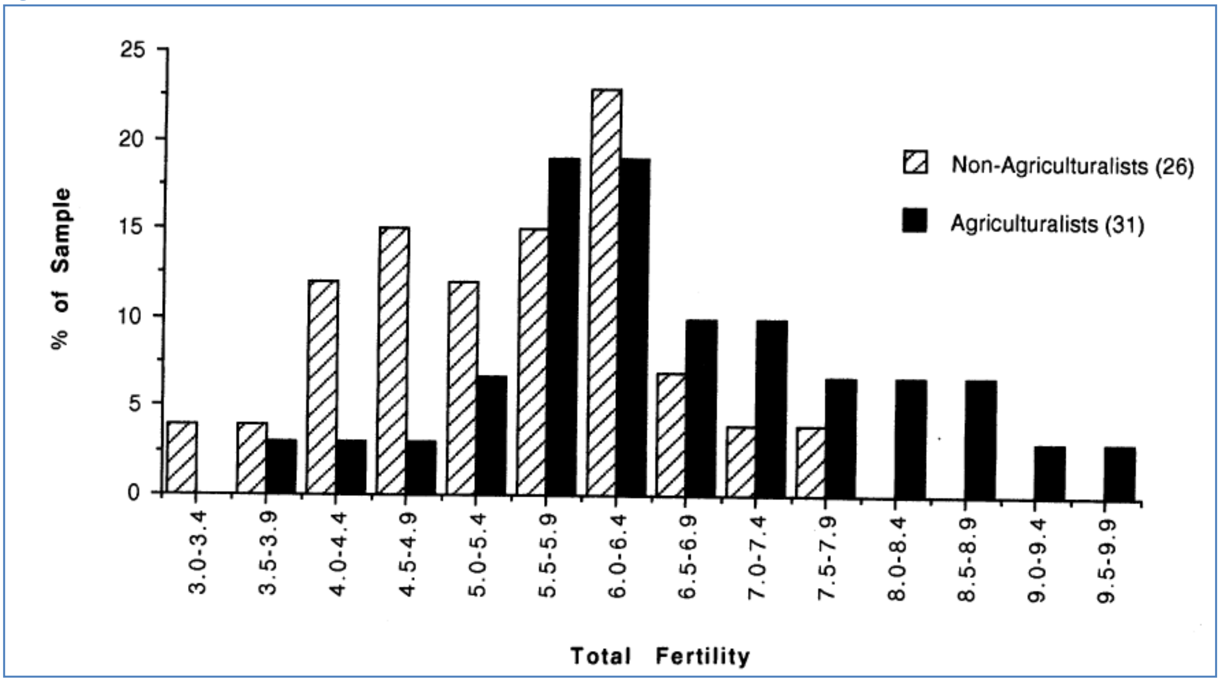
Figure 1 Distribution of total fertility rates in 70 natural-fertility populations *(solid histogram)* and 70 controlled-fertility populations *(broken histogram)*. Populations known to have high prevalences of pathological sterility are excluded. (For original data sources, see 28.)

It may be noted that the whole concept is debated: some anthropologists and demographers believe, that it is not correct to make such strong distinction, and even pre-transition fertility was consciously controlled (Hirschmann 1994). But even assuming n control, the possible range of fertility is quite broad (3,5-10, or more narrowly, 4,5-8), due to differences in marriage timing, fecundity, lactation end so on.

Fertility and subsistence

There is a growing body of evidence from anthropological studies, which report fertility levels for smaller communities or tribes. Bentley et al (1993) summarizes the literature about natural fertility populations, considering also the subsistence method. The following chart shows the percentage distribution of populations by the level of TFR. The result suggest that observations above 7 was rare, and the 5,5-6,5 range is a better estimate for long term natural fertility. In this study the authors show that the introduction of agriculturalists has significantly higher fertility than non-agriculturalist, but this result is debated.

Figure 19



Fertility and population growth

As Pennington (2001) pointed out, throughout of the history of mankind population growth was nearly zero, and substantial population growths are a recent phenomenon. How was this low population growth possible, if fertility was not controlled? Either mortality was exceptionally high, reducing life expectancy to nearly 20, or fertility was only moderately high – which is more likely. At the times when population density was low, illnesses could not spread so easily. The more likely solution to this puzzle is that fertility was only moderately high: either because it was controlled, at least through social customs, or because fecundity dropped due to illnesses.

10.2 Demographic reconstruction methods

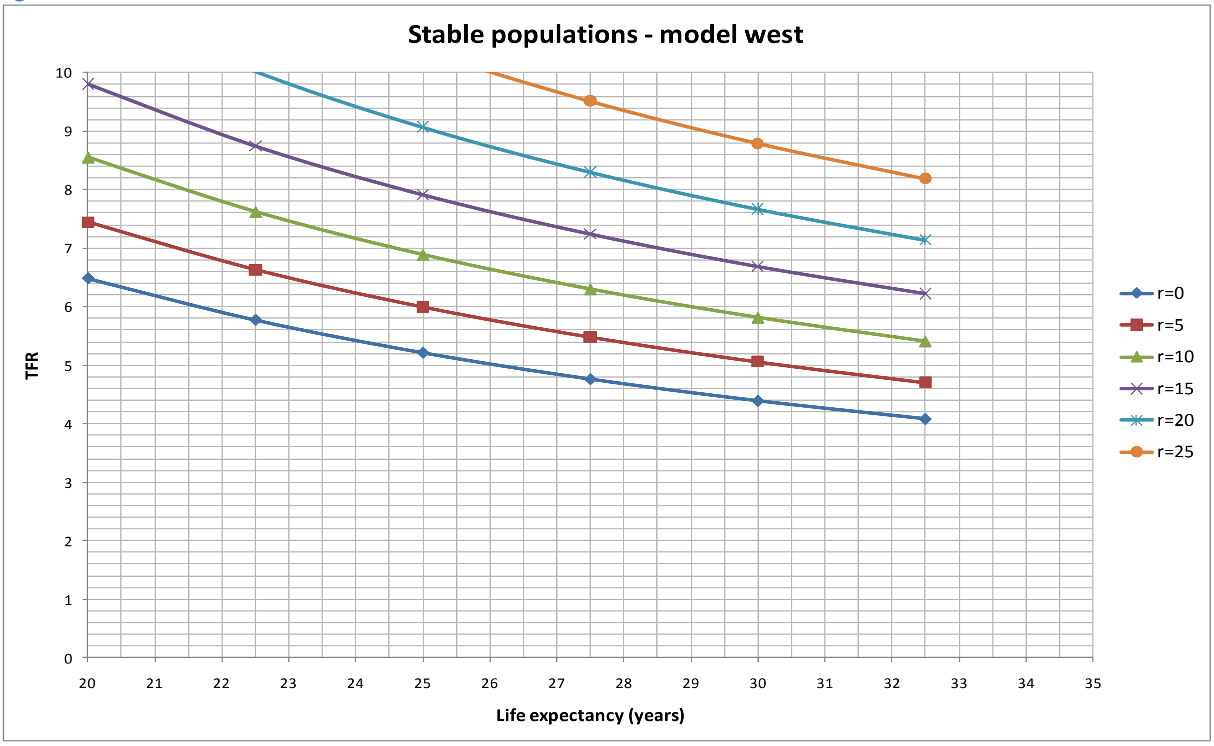
We have used Coale-Demeny (1983) model life tables to reconstruct fertility. The idea behind model life tables is that if we have a population with stable age-specific fertility and mortality rates, with zero migration, then after some time the age structure becomes stable. These are the so called stable populations.

Coale and Demény used a large number of historic life tables to derive four main model life tables, which represent the usual patterns of human mortality, they called the West, East, South and North model life tables. They have also calculated tables for stable populations, which link together age structure, fertility and mortality.

As we have discussed, ideally the population growth, mortality and fertility estimates of Gapminder should be in accordance. Otherwise it is likely that the fertility figure is not reasonable, because the other two estimates are less likely to be incorrect.

Within one model like the Coale-Demény Model West to one population growth figure and one mortality level only one fertility level belongs. In other words, if we have the population growth rate and the life expectancy, we can determine the TFR. It can be illustrated using Figure 1.

Figure 20



The value of r shows population growth per thousand, r=10 means 1% yearly growth rate.

We know that population growth was usually very low before the demographic transition, we may expect it to be between r=0 and r=5 in most 19th century populations. What is striking here that even at very high mortality levels (life expectancy e_0=20) fertility is only between 6,5 and 7,5 according to model West, and the other models are similar in this respect.

Model selection

We have used Model West, because this is the most frequent choice in the literature. The following table shows, that the choice of the model only has a minor effect on the results in this case.

Table 18 How does the choice of model effect the results in case of high mortality – low growth populations

Model assumptions		TFR values in model...		
		East	South	West
Zero growth r=0				
	Highest mortality level (e0 female=20)	6.51	6.73	6.48
	2nd highest mortality level (e0=22,5)	5.81	5.99	5.77
0.5% growth r=5				
	Highest mortality level (e0 female=20)	7.49	7.74	7.45
	2nd highest mortality level (e0=22,5)	6.69	6.89	6.64

Reasonableness testing using Coale-Demeny Model West

The reasonableness test has been done in the following steps.

1. Calculate long term population growth rate, using Gapminder data
2. Take the estimate of life expectancy from Gapminder dataset (for 1800)

3. Determine the TFR using model west life table
4. Determine a range for the TFR, assuming 5 years higher and 5 years lower life expectancy
5. Compare with existing estimate and evaluate the result

Calculation of population growth

We have calculated population growth for a long period, using the population indicator of Gapminder. For example the population estimate for Djibuti almost tripled over the 1820-1950 period, and this corresponds to 0,75% (r=7,5) yearly growth rate, using geometrical average, taking 130th route.

Table 19

Population	From year	Till year	Open	Close	time (years)	incr	r
Afghanistan	1820	1950	3,280,000	8,150,368	130	1.0070	7.0
Algeria	1820	1950	2,689,000	8,892,718	130	1.0092	9.2
Azerbaijan	1820	1950	879,960	2,885,332	130	1.0092	9.2
Djibouti	1820	1950	22,848	60,036	130	1.0075	7.5
Georgia	1820	1950	1,072,178	3,515,602	130	1.0092	9.2
Guinea-Bissau	1820	1950	218,167	573,268	130	1.0075	7.5
Iraq	1820	1950	1,093,000	5,163,443	130	1.0120	12.0
Jordan	1820	1950	217,000	561,254	130	1.0073	7.3
Kuwait	1820	1950	81,280	144,774	130	1.0045	4.5
Madagascar	1820	1950	1,683,000	4,620,437	130	1.0078	7.8
Micronesia, Fed. Sts.	1820	1950	16,416	30,715	130	1.0048	4.8
Morocco	1820	1950	2,689,000	9,343,384	130	1.0096	9.6
Nicaragua	1820	1950	186,000	1,097,916	130	1.0138	13.8
Oman	1820	1950	318,000	488,588	130	1.0033	3.3
Saudi Arabia	1820	1950	2,091,000	3,859,801	130	1.0047	4.7
Somalia	1820	1950	1,000,000	2,437,932	130	1.0069	6.9
Syria	1820	1950	1,337,000	3,495,000	130	1.0074	7.4
Tonga	1800	1950	18,658	45,744	150	1.0060	6.0
Ukraine	1820	1950	11,215,490	36,774,854	130	1.0092	9.2
Vanuatu	1820	1950	27,791	52,000	130	1.0048	4.8

10.3 Results

The countries which we reviewed are listed in the table below, together with the results of the calculations. Yemen was excluded, as we already excluded the estimates for Yemen.

Most of the values are outside of the reasonable range. This means that even if the life expectancy figure is 5 years lower than in the Gapminder dataset, the fertility estimate we have in the Gapminder 2009 dataset would produce a higher growth, than what we actually see. In other words the fertility estimate is not in line with our population and life expectancy estimates, but higher.

In some cases we could find additional data which help to resolve this problem, but in most cases the selected countries are such that they had no vital registration or reliable census before WWII.

In the “Action as a result” column we show what change has been implemented in the Gapminder 2010 dataset, compared to the Gapminder 2009 dataset, in case of the given country. In a number of cases we have deleted all estimates before 1950, because the reasonableness test indicates that the estimates were too high, and we had no exact way of determining, how much lower they should be. In some other cases an expert estimate was found which confirmed the estimated level of pre-transition fertility, and this is indicated with an “OK” in the action column, while the source of the estimate is referred to in the last column. Finally, in a few cases which are indicated as “corrected” additional literature has been found which has helped to refine the estimates.

Table 20 The result of reasonableness testing

Nr	Country	Input from other Gapminder indicators: Estimate for pre-transition LE	Input from other Gapminder indicators: r	TFR in 1800, in Gapminder 2009	Estimates using "model west" stable populations: TFR mid	TFR high	TFR low	Gapminder 2009 data within range?	Action as a result	Note regarding additional literature / solution
40	Afghanistan	28	7.0	7.7	5.7	7	4.4	No	deleted the estimates	Islam
50	Algeria	29	9.2	7.3	5.9	6.9	4.9	Nearly	deleted the estimates	Islam
37	Azerbaijan	30	9.2	8.1	5.7	6.7	4.7	No	Ok	Lutz et al (1994) estimate for 1900 repeated for 1800. The estimate was kept, as it is based on an expert estimate of pre-transition fertility, not solely on back-projecting 1950 UN data. Although it is way out of the range, the mortality could be lower / population growth higher then shown here – as this may be classified as a frontier population.
38	Djibouti	30	7.5	7.8	5.4	6.5	4.3	No	deleted the estimates	Islam
39	Georgia	30	9.2	7.8	5.7	6.7	4.7	No	Ok	Lutz et al (1994) estimate for 1900 repeated for 1800. The estimate was kept, as it is based on an expert estimate of pre-transition fertility, not solely on back-projecting 1950 UN data. Although it is way out of the range, the mortality could be lower / population growth higher then shown here – as this may be classified as a frontier population.

Table 20 The result of reasonableness testing

Nr	Country	Input from other Gapminder indicators: Estimate for pre-transition LE	Input from other Gapminder indicators: r	TFR in 1800, in Gapminder 2009	Estimates using "model west" stable populations: TFR mid	TFR high	TFR low	Gapminder 2009 data within range?	Action as a result	Note regarding additional literature / solution
45	Guinea-Bissau	29	7.5	7.4	5.6	6.7	4.5	No	deleted the estimates	we only have UN, other study was not found
49	Iraq	31	12.0	7.3	5.9	7	4.8	Nearly	deleted the estimates	Islam
46	Jordan	32	7.3	7.4	5.1	6.1	4.1	No	deleted the estimates	Islam
53	Kuwait	26	4.5	7.2	5.7	7	4.4	Nearly	deleted the estimates	Islam
48	Madagascar	30	7.8	7.3	5.5	6.6	4.4	No	deleted the estimates	Garenne 2008
57	Micronesia, Fed. Sts.	25	4.8	7.2	5.9	7.3	4.5	Yes	corrected	New estimate, based on the article TFR 6.02 for 1800 and 1925; Pirie 2000
59	Morocco	33	9.6	7.2	5.4	6.3	4.5	No	delete	Islam
56	Nicaragua	25	13.8	7.2	7.6	9.4	5.8	Yes	ok	We know very little about historic fertility levels. In Collver (although it has some good data for other countries) I could not find much data before 1950. For 1900 he estimates CBR=46, I am now using this as a new estimate, also extrapolating backwards.
54	Oman	32	3.3	7.2	4.6	5.2	4	No	deleted the estimates	Islam

Table 20 The result of reasonableness testing

Nr	Country	Input from other Gapminder indicators		TFR in 1800, in Gapminder 2009	Estimates using "model west" stable populations				Action as a result	Note regarding additional literature / solution
		Estimate for pre-transition LE	r		TFR mid	TFR high	TFR low	Gapminder 2009 data within range?		
55	Saudi Arabia	32	4.7	7.2	4.7	5.3	4.1	No	deleted the estimates	Islam
51	Somalia	29	6.9	7.3	5.7	6.8	4.6	No	deleted the estimates	Islam
52	Syrian Arab Republic	31	7.4	7.3	5.3	6.2	4.4	No	deleted the estimates	Islam
47	Tonga	25	6.0	7.3	6.2	7.7	4.7	Yes	corrected to 6.5	Expert estimate: TFR 6.5 around 1950 Khalidi 1995 pg 28
43	Ukraine	37	9.2	7.5	4.8	5.4	4.2	No	ok	Lutz et al (1994) estimate for 1900 repeated for 1800. The estimate was kept, as it is based on an expert estimate of pre-transition fertility, not solely on back-projecting 1950 UN data. Although it is way out of the range, the mortality could be lower / population growth higher then shown here – as this may be classified as a frontier population.
42	Vanuatu	25	4.8	7.7	5.9	7.3	4.5	Yes	corrected	Correct to 6.6 based on pre-transition level estimate by Lucas and Ware, (1981) pg 306

11 Literature

Banens M, Bassino J, and Egretaud E. (1998). ”Estimating population and labour force in Vietnam under French rule (1900-1954).” Institute of economic research Hitotubashi university (Discussion Paper No. D98-7) from table 22.

Barclay GW, Coale AJ Stoto MA, Trussell J T. (1976). “A Reassessment of the Demography of Traditional Rural China”. Population Index, Vol. 42, No. 4 (Oct., 1976), pp. 606-635

Bayliss-Smith T (1988). “Islands, islanders, and the world: the colonial and post-colonial experience: the colonial and post-colonial experience of Eastern Fiji.” Cambridge University Press, NY.

Bengtsson, et al. (2004). “Life under Pressure: Mortality and Living Standards in Europe and Asia, 1700-1900.” Cambridge, MA: MIT Press

Bentley G, Goldberg T, Jasienska G. (1993). ”The Fertility of Agricultural and Non-Agricultural Traditional Societies Population Studies”, Vol. 47, No. 2 (Jul., 1993), pp. 269-281

Bhat P N Mari (1989). “Mortality and Fertility in India 1881-1961” in : Dyson T. India's Historical Demography: Studies in Famine, Disease and Society. Curzon London

Bhat P N Mari and A.J.F. Zavier. (2003). “Fertility decline and gender bias in northern India”, Demography 40(4): 637-657

Bhat P N Mari, Preston S H, Dyson T, National Research Council (U.S.). Committee on Population and Demography Panel on India. (1984).” Vital rates in India, 1961-1981” National Academy Press, Washington, D.C.

Bhat, P N Mari (2002). “Returning a Favor: Reciprocity between Female Education and Fertility in India”, World Development Vol.30, No10 (October 2002)

Bocquet-Appel JP, Rajan I S, Bacro J, Lajaunie C. (2002). “The Onset of India's Fertility Transition”. European Journal of Population , Volume 18, Number 3, pp. 211-232

Bogue, Donald J. et al (1993): “Empirical interrelationships among standard fertility measures. Readings in population research methodology”. Volume 3. Fertility research, edited by Donald J. Bogue, Eduardo E. Arriaga, Douglas L. Anderton, George W. Rumsey. Chicago, Illinois, Social Development Center, 1993. :11-51 - 11-53. Published for the United Nations Population Fund [UNFPA]

Bongaarts, J. and Bulatao, R. A. (eds.) (2000). “Beyond Six Billion. Forecasting the World’s Population” National Academic Press, Chapter 3-6.

Boserup, E. (1983) “The Impact of Scarcity and Plenty on Development”. Journal of Interdisciplinary History, vol. 14, no. 2, 383-407

Brea J. (2003). “Population Dynamics in Latin America”. Population Bullettin vol 58 no 1 Population Reference Bureau.

Caldwell J C and Caldwell B K. (1997). “Asia's Demographic Transition”. Asian Development Review 15(1) pp. 52-87

Caldwell J C. (2006). “Demographic Transition Theory”; Chapter 10.: Regional Paths to Fertility Transition, Dordrecht, The Netherlands: Springer. pp.418. ISBN 1-4020-4373-2.

Caldwell J. (2004). “What do we know about Asian Population History”. In: Asian Population History ed.: Liu T, Lee J, Reher D S, Saito O, Feng W. Oxford.

Caldwell JC and Srinivasan K. (1984). “New Data on Nuptiality and Fertility in China”. Population and Development Review, Vol. 10, No. 1 (Mar., 1984), pp. 71-79.

Caldwell, J. et al (1989): “Is marriage delay a multiphasic response to pressures for fertility decline? The case of Sri Lanka”; Journal of marriage and family 51 (2)

Caldwell, John C. (2006) : Demographic Transition Theory; Chapter 10.: Regional Paths to Fertility Transition, Dordrecht, The Netherlands: Springer. pp. 418. ISBN 1-4020-4373-2.

Calot G. (1998). “Two centuries of Swiss demographic history. Graphic album of the 1860–2050 periods”.Table F_05_01. Swiss Federal Statistical Office 1998.

Castles L. (1975). “Sources for the Population History of Northern Sumatra.” Masyarakat Indonesia Tahun Ke 2/2 pp 189-208 (Data from one region in Northern Sumatera(Batak)) Based on baptism records data was in the following book:東南アジア人口民族誌 "坪内良博", 1986

Census of the Philippines Islands: (1903) Volume 1 to 4 Director: Gen. J. P. Sanger, USA. United States Bureau of the Census, Washington, 1905

Chesnais, Jean-Claude (1992): “The Demographic Transition : Stages, Patterns, and Economic Implications. A Longitudinal Study of Sixty-Seven Countries Covering the Period 1720-1984”; Oxford : Clarendon Press , 1992

Coale A, Demeny P, Vaughan B. (1983). “Regional model life tables and stable populations”. ISBN 0-12-177080-X 2. ed. New York : Academic P., cop. 1983 English viii, 496 p.

Coale A. (1988). Book review article of “Regulating Reproduction in India's Population: Efforts, Results, and Recommendations by K. Srinivasan” Source: Economic Development and Cultural Change, Vol. 46, No. 2 (Jan., 1998), pp. 440-444

Coale, A.J and Chen Sheng Li (1987): “Basic data on fertility in the provinces of China, 1940-82” (Papers of the East-West Population Institute No. 104)

Collver A, Arevalo JV. 1967. Current Trends and Differentials in Fertility as Revealed by Official Data. The Milbank Memorial Fund Quarterly, Vol. 46, No. 3

Collver A. (1965) “Birth Rates in Latin America. New Estimates of Historical Trends and Fluctuations” (Berkeley: Institute of International Studies, 1965), pp. 36-37."

Connelly M J. (2008). “Fatal misconception: the struggle to control world population”. Belknap Press of Harvard University Press. ISBN-10: 0674024230 544 pp.

Corruccini R S, Handler J S. (1983). “Plantation Slave Life in Barbados: A Physical Anthropological Analysis”, Journal of Interdisciplinary History, Vol. 14, No. 1 (Summer, 1983), pp. 65-90

Crimmins E M, Easterlin R A, Jejeebhoy K, Srinivasan K. (1984). “New Perspectives on the Demographic Transition: A Theoretical and Empirical Analysis of an Indian State, 1951-1975”. Economic Development and Cultural Change, Vol. 32, No. 2, pp. 227-253

Das Gupta M (1995). “Fertility Decline in Punjab, India: Parallels with Historical Europe”. Population Studies, Vol. 49, No. 3, pp. 481-500

Das Gupta M, Shuzhuo L. (1999). “Gender Bias in China, South Korea and India 1920-1990: Effects of War, Famine and Fertility Decline”. Development and Change, Vol. 30 No. 3 pp. 619-652

DaVanzo J, Haaga J. (1982). “Anatomy of fertility decline: Peninsular Malaysia, 1950-1976”. Population Studies 36: 373-393.

Davis K (1944). “Demographic Fact and Policy in India”. The Milbank Memorial Fund Quarterly, Vol. 22, No. 3, pp. 256-278

Davis K. (1943). “The Population of India”. Far Eastern Survey, Vol. 12, No. 8, pp. 76-79

Davis K. (1946). “Human Fertility in India”. The American Journal of Sociology, Vol. 52, No. 3, pp. 243-254

Davis K. (1951). “The Population of India and Pakistan”. Princeton: Princeton University Press

Dyson T, Murphy, M. (1985). “The Onset of Fertility Transition”. Population and Development Review, Vol. 11, No. 3 (Sep., 1985), pp. 399-440

Dyson T. (1989). “India's Historical Demography: Studies in Famine, Disease and Society”. Curzon London

Dyson T. and Das Gupta M. (2001). “Demographic trends in Ludhiana district, Punjab, 1881-1981: an exploration of vital registration data in colonial India”. In: Asian Population History by Ts'ui-jung Liu and James Lee and David Sven Reher and Osamu Saito, Oxford UP New York 451p.

Dyson, Tim and Murphy, Mike (1985): “The Onset of Fertility Transition”. Population and Development Review, Vol. 11, No. 3 (Sep., 1985), pp. 399-440

Edge G. (1937). “The Demography of British Colonial Possessions: A Note on the Assembly and Interpretation of Data”. Journal of the Royal Statistical Society, Vol. 100, No. 2 (1937), pp. 181-231"

Eltigani E. (2005). “Fertility transition in Arab countries: a re-evaluation”. Journal of Population Research Vol. 22, No. 2, pp 163-183

Engerman S L. (1997). “Population History of the Caribbean”. In: Haines M R, Steckel R H.: The Population History of North America. 1997. Cambridge U P

Entwisle, Barbara (1981). “CBR versus TFR in Cross-National Fertility Research”. Demography, Vol. 18, No. 4, pp. 635-643

Eswaran M. (1996). Book review of “Regulating Reproduction in India's Population: Efforts, Results, and Recommendations by K. Srinivasan” Source: Pacific Affairs, Vol. 69, No. 3 (Autumn, 1996), pp. 435-436

Fargues (1989). “The Decline of Arab Fertility”. Population: An English Selection, Vol. 44, No. 1 (Sep., 1989), pp. 147-175

Fawcett J.T. and Khoo S. (1980): “Singapore: Rapid Fertility Transition in a Compact Society“ Population and Development Review, Vol. 6, No. 4 (Dec., 1980), pp. 549-579

Fernando D.F. (1981): “Problems in development of vital registration systems: the situation in Sri Lanka”. In: International Population Conference, Manila, 1981: solicited papers, vol. 3. Liege, Ordina Editions, [1981]. :331-42. Papers of the 19th General Conference of the International Union for the Scientific Study of Population.

Festy P. (1979). “La fécondité des pays occidentaux de 1870 à 1970”. Travaux et Documents No. 85. Paris: INED – PUF.

Freedman R, Ming-Cheng Chang, Te-Hsiung Sun (1994): “Taiwan's Transition from High Fertility to Below-Replacement Levels”. Studies in Family Planning, Vol. 25, No. 6 (Nov. - Dec., 1994), pp. 317-331

Garenne, M. Michel. (2008). “Fertility Changes in Sub-Saharan Africa”. DHS Comparative Reports No. 18. Calverton, Maryland, USA: Macro International Inc.

Gentile M. (2005) “Population Geography Perspectives on the Central Asian Republics”. Swedish Institute for Futures Studies. Stockholm 2005, ISBN 91-89655-74-5

Gilbert G. (2005). “World Population: A Reference Handbook”. 2d ed. Santa Barbara, Calif., ABC-CLIO, 2006. 298p.

Graham, E. (2007): “Son Preference, Female Deficit and Singapore's Fertility Transition"; Watering the Neighbour's Garden: The Growing Demographic Deficit in Asia; I Attane, C Z Guilmoto (ed); CICRED; 89-106

Graham, E. Teo, P and Yeoh, B. (2002) “Reproducing the Asian Family Across the Generations: ‘Tradition’, Gender and Expectations in Singapore”. Asian-Pacific Population Journal 17(2): pp 61-86

Haines M R, Steckel R H.(1997). “The Population History of North America”. Cambridge U P

Heuser R L. (1976). “Fertility tables for birth cohorts by colour”. U. S. Department of Health, Education, and Welfare. DHEW Publication No. (HRA)76-1152.

Higman B W. (1995). “Slave populations of the British Caribbean, 1807-1834”. University of the West Indies, 781 pages (reprint of the 1984 edition) birth rates: pg 308-310

Hirschman C. (1980). “Demographic Trends in Peninsular Malaysia, 1947-75” Population and Development Review, Vol. 6, No. 1 (Mar., 1980), pp. 103-125

Hirschman, C. (2001) “Fertility transition in Southeast Asia” In: Smelser and Baltes (eds) International Encyclopedia of the Social and Behavioral Sciences Vol. 8 pp 5597-5602 Oxford, Elsevier.

Hofsten E. (1975). “Demographic Transition and Economic Development in Albania.” European Demographic Information Bulletin, Vol. 6, No. 3 (1975), pp. 147-158 "

Hollerbach P E, Diaz-Briquets S. (1983). “Fertility determinants in Cuba”. National Academy Press, Washington. (tableau 44)

Indepth Network (2004). “Indepth Model Life Tables: for Sub-Saharan Africa : Indepth Network Ghana.” Ashgate Publishing

Inoue (1997). “中国とインド, 人口学からの視点” (China and India, from demographic point of view). The institute of humanities and social sciences, Nihon university. Institution’s research report (研究紀要) Vol 53.

Kandlah V. (2003). “Fertility levels and trends and their implications for policies and programmes”. In: Asian Population Studies Series No. 158. (Fifth Asian and Pacific Population Conference: Selected Papers)

Keyfitz N, Flieger W. (1968). “World Population: Analysis of Vital Data”. Chicago : University of Chicago Press

Khalidi N A. (1995). “Population, fertility, mortality and migration in Tonga”. Pacific Health Dialogue vol 2 no 1.

Klein H S. (1986). “Familia y fertilidad en Amatenago, Chiapas, 1785- 1816”. Historia Mexicana Vol. 36, No. 2 (Oct. - Dec., 1986), pp. 273-286

Kollega T I. (1996). “Hungary in the 20st century - Magyarország a XX. Században”. Babits, Budapest.

Kučera M, Srb V. (1959). “Vývoj obyvatelstva českých zemí v XIX. Století”. In: Statistika a demografie I., ČSAV, Praha.

Kumo K, Morinaga T, Shida Y. (2007). ”Long-Term Population Statistics for Russia, 1867-2002 RRC” Working Paper Series No. 2, December 2007

Langford, Christopher (2001): “Trends and fluctuations in fertility in Sri Lanka during the first half of the twentieth century”. In: Liu, Ts'ui-jung, (ed.) Asian population history. Oxford University Press, Oxford, pp. 176-188. ISBN 9780198294436

Lavely, Wong (1998). “Revising the Malthusian Narrative: The Comparative Study of Population Dynamics in Late Imperial China”. The Journal of Asian Studies, Vol. 57, No. 3 pp. 714-748

Lee, Feng, Campbell (1994). “Infant and Child Mortality among the Qing Nobility: Implications for Two Types of Positive Check”. Population Studies, Vol. 48, No. 3, pp. 395-411

Lee, Feng. (1999). “Malthusian Models and Chinese Realities: The Chinese Demographic System 1700-2000”. Population and Development Review, Vol. 25, No. 1, pp. 33-65

Liu T, Lee J Reher D S, Saito O Feng W. (2001). “Asian Population History”, Oxford: Oxford University Press

Lucas D, Ware H. (1981) “Fertility and Family Planning in the South Pacific Studies” in Family Planning, Vol. 12, No. 8/9 (Aug. - Sep., 1981), pp. 303-315"

Ludwig, Armin K. (1985). “Brazil: A handbook of Historical Statistics”. ISBN 0816181888

Lund University Macroeconomic and Demographic Database

Lutz W, Scherbov S, Volkov, A G. (1994).”Demographic trends and patterns in the Soviet Union before 1991”.Routledge, New York.

Maharatna A. (2000). “Fertility, mortality and gender bias among tribal population: an Indian perspective”, Social Science & Medicine Vol. 50 No. 10. pp. 1333-1351

McCaa R. (1997): “The Peopling of Mexico from Origins to Revolution”. in: Haines M R, Steckel R H. The Population History of North America. 1997. Cambridge U P

McCarthy J E. (1997). Book review of “Regulating Reproduction in India's Population: Efforts, Results, and Recommendations by K. Srinivasan” Source: The Journal of Environment & Development, Vol. 6, No. 1, 96-98 (1997) and also : www.math.wustl.edu/~mccarthy/public_papers/india_population.pdf

McInnis M. (1997). “The population of Canada in the 19th century”. in: Haines M R, Steckel R H. The Population History of North America. 1997. Cambridge U P

McNay K, Arokiasamy P, Cassen R H. (2003). “Why Are Uneducated Women in India Using Contraception?”. Population Studies, Vol. 57, No. 1 , pp. 21-40

Mitchell, B.R. (1998): “International Historical Statistics” Macmillan, Volumes I-III. UK 1998

Mukherjee S B. (1977). “The Age Distribution of the Indian Population. A Reconstruction for the States and Territories, 1881-1961”. University of Hawaii Press.

Nag, M. (1980). “How Modernization Can Also Increase Fertility” Current Anthropology, Vol. 21, No. 5

National Statistical Offices (of the given country)

Neville, Warwick (1978): “The Birth Rate in Singapore”; Population Studies, Vol. 32, No. 1 (Mar., 1978), pp. 113-133

Noriko T. (2001). “Birth regime of Modern Japan: Event history analysis in 18th and 19th Century Two Northeastern Villages" KESDP(Keio Economic Society Discussion Paper series) No.00-9 (in Japanese)

Palloni A. (1990). “Fertility and Mortality Decline in Latin America”. Annals of the American Academy of Political and Social Science, Vol. 510, World Population: Approaching the Year 2000, pp. 126-144"

Pantelides E A. (2002). In: Expert Group Meeting on Completing the Fertility Transition, New York, 11-14 March 2002, [compiled by] United Nations. Department of Economic and Social Affairs. Population Division. New York, New York, United Nations, Department of Economic and Social Affairs, Population Division, 2004. :333-342.

Pennington, R. (2001). “Hunter-gatherer demography”. In C.Panter-Brick, Robert Layton and Peter Rowley-Conwy, eds., Hunter-gatherers: An interdisciplinary

Pirie P. (2000). “Untangling the Myths and Realities of Fertility and Mortality in the Pacific Islands”. Asia-Pacific Population Journal, Vol 15 No 2

Population Statistics of Japan (2008), Table 4.3

Population Trends (2009) – Statistics Singapore, ISSN: 1793-2424 (http://www.singstat.gov.sg/pubn/demo.html#popntrend)

Preston S H, Bhat M. (1984). “New Evidence on Fertility and Mortality Trends in India”. Population and Development Review, Vol. 10, No. 3, pp. 481-503

Princeton European Fertility Project. (1996). “If indexes were shown in the dataset, and converted using 12,44 as a conversion key”. (Sardon, 1996)

Puvanarajan, P. (2002): “Estate Women’s Fertility in Sri Lanka: Some Aberrant Perspectives in the Causal Links” Asia-Pacific Population Journal, September

Rashad H. (2000). “Demographic transition in Arab countries: A new perspective”. Journal of Population Research Vol 17 No 1 pp 83-101.

Rédei M. (1960). “A születések és halálozások alakulása” KJK 1960

Rele J R. (1987). “Fertility Levels and Trends in India, 1951-81”. Population and Development Review, Vol. 13, No. 3, pp. 513-530

Rele R J. (1988). “70 years of fertility change in Korea: New estimates from 1916 to 1985”. Asia-pacific Population Journal. 1988 Jun; 3(2):29-54.

Retherford R D, Rele J R. (1989). “A Decomposition of Recent Fertility Changes in South Asia”. Population and Development Review, Vol. 15, No. 4 , pp. 739-747

Rodnost v Sloveniji od 18. do 21 Stoletja (2006); Tabela 97 pg 274. Statistical Office of Slovenia, Ljubljana

Rothenbacher F. (2002). “The European Population 1850-1945”. CD-Rom. Basingstoke, Palgrave.

Saito, Osmu (2006): “Pre-Transition Fertility in Asia : A Comparative-historical Approach” Journal of International Economic Studies No.20, 1–17

Saito, Singh, Takahama, Yamamoto (1998). “南インドの人口動態 1891-1948” (Vital statistics of Southern India 1891-1948). The Journal of Population Studies Vol. 23. (published by Population Association of Japan). (this document is in Japanese).

Sardon J-P. (1991). “Generation replacement in Europe since 1900”. Population: An English selection 3(1991): 15-32. Table 2.

Sardon J-P. (1996). “Coale's Indices, Comparative Indices, Mean Generation, Total Fertility Rate and Components”, Population: An English Selection, Vol. 8 (1996), pp. 252-257

Saw, Swee-Hock. (2007). “The population of Peninsular Malaysia”. Institute of Southeast Asian Studies. Singapore University Press. 342 p.

Schoen R. (2003). “Insights from parity status life tables for the 20the century U.S." Working Paper 03-08, Population Research Institute, The Pennsylvania State University.

Seers D. (1957). “A Fertility Survey in the Maltese Islands”. Population Studies, Vol. 10, No. 3 (Mar., 1957), pp. 211-228

Smith P C, Ng S. (1982). “The components of Population Change in Nineteenth-century Southeast Asia: Village Data from the Philippines”. Population Studies, 36:2, pp. 237-255. Comment: Data from one region, Nagla prefecture. Based on Baptism record. Data was in the following book: 東南アジア人口民族誌 "坪内良博", 1986

Sobotka, Tomas (2009) Dataset kindly shared with Gapminder. Data taken from his file are indicated in column "Remark", but as in other cases, the source reference used by Sobotka is copied and indicated as "Source"

Sprocha B, - TIŠLIAR P. (2008). “Plodnost a celková reprodukcia obyvatelstva Slovenska v rokoch 1919 – 1937”. Stimul: Bratislava

Srinivasan, K., (1995) “Regulating reproduction in India's population : efforts, results, and recommendations”. New Delhi : Thousand Oaks, Calif. : Sage Publications. 329 p.

Stycos JM. (1978). “Recent Trends in Latin American Fertility”. Population Studies, Vol. 32, No. 3 (Nov., 1978), pp. 407-425

Swee-Hock, Saw (2007): “The Population of Singapore” (2nd edition). Institute of Southeast Asian Studies (ISBN 978 981 230 739-2)

Thang, L.L. (2005): “Private matters, public concern: Procreation issues in Singapore”. The Japaneese Journal of Population, Vol 3, No1. 2005, pp 76-101

Tomobe K. (2001). “The level of fertility in Tokugawa and Meiji Japan, c. 1800s-1930s”. (pg 138) In: Asian Population History by Ts'ui-jung Liu and James Lee and David Sven Reher and Osamu Saito, Oxford UP New York 451p.

Trussel T J, Hill K. (1980). “Fertility and Mortality Estimation from the Panama Retrospective Demographic Survey, 1976” Population Studies, Vol. 34, No. 3 (Nov., 1980), pp. 551-563"

UN (1986), ”Socio-economic development and fertility decline in Sri Lanka”. United Nations. Department of International Economic and Social Affairs; New York, United Nations, 1986. 112 p. (ST/ESA/SER.R/54) Background paper prepared for the Project on Socio-Economic Development and Fertility Decline

UN (1954). “Demographic Yearbook”. Statistical Office of the United Nations 1954. 738 p.

UN (2001). “The state of demographic transition in Afrika. Economic Commission for Africa”. Food Security and Sustainable Development Division (FSSDD)

UN World Population Prospects 2008 dataset (contains smoothened data end estimations)

United Nations (1982). “Model Life Tables for Developing Countries” (United Nations publication, Sales No. E.81.XIII.7). http://www.un.org/esa/population/publications/Model_Life_Tables/Model_Life_Tables.htm

Unknown (????). www.chs.nihon-u.ac.jp/institute/human/kiyou/53/H-053-003.pdf (this document is in Japanese)

Walters S L. (2008). “Fertility, Mortality and Marriage in Northwest Tanzania, 1920-1970: a Demographic Study Using Parish Registers”. Cambridge University PhD Thesis. (http://www.geog.cam.ac.uk/people/walters/)

Watkins, Suzan Cotts (1987): “The Fertility Transition: Europe and the Third World Compared”: Sociological Forum, Vol. 2, No. 4, Special Issue: Demography as an Interdiscipline (Autumn, 1987), pp. 645-673

Vishnevskij (ed). (2006). “Demographic modernisation of Russia: 1900-2000”. Демографическая модернизация России: 1900-2000. Под редакцией Анатолия Вишневского. Серия "Новая история" Москва, Новое издательство, 2006, 601 страница.

Wojtun, Bronislaw. (1983). “Zmiany rodności i umieralności ludności Polski Centralnej w XIX i XX w.” PWN, 153 pages. Tabela 2. Page 147-149

Wong, T; Brenada J. (2003): “Fertility and the Family: An Overview of Pro-natalist Population Policies in Singapore”. Asian Meta Centre Research Paper Series 12. National University of Singapore

Wood J W. (1990). “Fertility in Anthropological Populations”. Annual Review of Anthropology, Vol. 19. pp. 211-242 Published by: Annual Reviews

Wright, N. (1970): “CEYLON: The Relationship of Demographic Factors and Marital Fertility to the Recent Fertility Decline”. Studies in Family Planning, Vol. 1, No. 59 (Nov., 1970), pp. 17-20

Wrigley E A. (1978): “Fertility strategy for the individual and the group”. In C. Tilly (ed.), Historical Studies of Changing Fertility: 135-154. Princeton, NJ: Princeton University Press.

Yadava K N S, Yadava S S. (1999). “Women's status and fertility in rural India”. The History of the Family, Volume 4, Issue 2, June 1999, Pages 209-228

Yapa, Lakshman and Siddhisena, Padmasiri (1998), “Locational specificities of fertility transition in Sri Lanka”, GeoJournal 45: 177-188