Lessons learned with the use of demographic methods and multiple sources of data to evaluate the completeness and data quality from birth registration in Latin America[1]

Everton E. C. Lima[2]
Bernardo Lanza Queiroz[3]

Abstract

We perform an analysis of the main sources of data used to estimate fertility schedules in developing countries, giving special attention to the Latin American countries. In addition to the brief history of distinguished data sources, taking Brazil as a case study, we present some indirect demographic methods, commonly used to assess the quality of data and indirectly estimate fertility. From all the methods used, the Synthetic Relational Gompertz model appeared to give the most robust estimates of fertility, independent of the data source considered. Nevertheless, we conclude that different demographic data sources and methods have assumption that generates different estimates of fertility. We argue that this might be an obstacle to progress in the path to high quality birth records in the region.

Preliminary draft
Comments are welcome

[1]We would like to thank Tomas Sobotka and Krystof Zeman for all the help, assistance and tips with the paper editing.

[2] Assistant Professor at the College of Philosophy and Human Sciences (IFCH) at University of Campinas (UNICAMP) and research scientist at the Population Studies Center (NEPO).

[3]Associate Professor at Universidade Federal de Minas Gerais.

Introduction

When vital statistics in less developed and developing countries are addressed in international meetings, there is always a common concern among demographers, about data quality and the consensus on the need to improve vital statistics. This concern about the quality of vital statistics is understandable, since many countries in the globe and especially in Latin America still present a considerable degree of data problems (Faijer, 1994).

Despite the attempts to improve the quality of vital statistics records in Latin America, no considerable improvements have been noticed over the last years. Even more, it is very likely that the situation has worsened, due to recent years of economic insecurity in the region, which could have directly affected the improvements of these statistics.

However, due to the need to design, to evaluate and to implement political and social programs, there is a growing interest in having appropriate and good vital statistics, even in a context of limited financial resources and within a framework of decentralized governance. Nevertheless, due to the absence of trustworthy civil registration, many Latin American countries are facing problems to allocate recourses for the design of their social policies, as well as they are producing "out-of-date" social indicators, generally based on census or survey data, and population projections. Obviously, in this way, the impact of the programs may not be optimal, and the risks of fall into the already limited resources are also greater (Faijer, 1994).

Hence, the estimates of fertility and mortality (officially published and available by reports from statistical offices of these countries) of many countries in the region end up being derived only from other researches using data sources such as the population census, household surveys or Demographic Health Surveys (DHS). This can be a limitation to the development of reliable demographic estimates for the region, since even these alternative sources have presented numerous problems linked to low coverage of population and vital events, age heaping, age misreporting, etc. (,Guzman et al. 2006).

In spite of that, some countries in the region have developed considerably their civil statistics over the years, but there is still a high level of heterogeneity regarding data quality across and within countries, and in some cases we can even identify reverse trends in quality, making difficult a harmonization and unification of demographic estimates and sources.

Numerous demographic and statistical methods have been employed and developed to assess the quality of the demographic information and to provide vital statistic estimates for the region. Most of these methods have been applied to evaluate the completeness of births and deaths, using population censuses as mainly data source.

Within this context, given special attention to birth registration and fertility estimates, in the following sections we will make an historical overview of the main data sources used to access births information and to provide fertility schedules in Latin America. In a second part of the paper, taking Brazil a case study, we will stress some indirect demographic methods that have been applied to estimate fertility in the region, using different data sources as example. The main idea is to highlight how these methods perform to estimate fertility, stressing its advantages and setbacks.

1. Historical progress in demographic data quality in Latin America

1.1. Censuses

Historically the main source of vital statistic information in the countries of the region has been the population censuses, not only because its concern with the knowledge of population size and distribution, but also to provide information of internal migration, certain aspects of international migration, and also to provide a means to estimate fertility and infant mortality in the region (Guzman et al. 2006).

In the eighteenth and nineteenth centuries, the first population censuses were conducted in Latin America (Sánchez-Albornoz, 1977). However, only during the first half of the twentieth century several countries started to organize regular censuses, every ten year, such as Mexico, Chile and Panama, or more frequently in some cases like Honduras that organized six censuses between 1925 and 1950 (Guzman et al. 2006).

From the 1950s on, most of the countries in the region started to introduce regular decennial census. Although, due to political, financial or other reasons, some countries did not respect this time interval and allowed more than ten years to elapse between two censuses. In the early 1970s, basically all the countries of Latin America and the Spanish-speaking Caribbean conducted censuses, though several did not take part in the census series of the following decade and, above all, that of the 1990s decade ,primarily because of the economic crisis in the region.

Since 2000, most of the countries in the region have held a census, and results now become available much more quickly, thanks to progress in data processing and dissemination via the Internet (Jaspers and Poulard, 2002). Despite progress in the collection and dissemination of demographic data, there are still some problems and the census under-reporting still exceeds 3% in many countries (Guzman et al. 2006). In addition to that, this progress is uneven, with signs of improvement in countries with past high levels of under-reporting (Bolivia, Ecuador, Guatemala), but also a deterioration in other countries where under-reporting was historically low (Argentina and Chile).

For example, Chile's last census in 2012, we see that many problems and setbacks with the collection and data process have occurred. According to Bravo et al. (2013) the main problems were high level of omission (approximately 1.7 million people, i.e. under-reporting of 9.6%); problems related to identification of the numbers; under-reporting of foreign population(only 120 thousand were counted while official statistics indicated 220 thousand in-migrants); the question concerning "living place", which create problems with the counting of absentee residents versus the vacant homes; and the technological process that has generated considerable low response rates. We notice that censuses in the region have not progressed linearly, and many setbacks and reversing quality trends are sometimes identified.

The censuses' questions concerning reproductive information are typically asked to all women of childbearing age (12, or 15 and older, sometimes with an upper age limit of 49). However, due to the complexity of the census exercise, and the need to attempt to enumerate every individual, it is not feasible to ask detailed questions on fertility (Moultrie et al., 2013).

In practice, the census questions aim to obtain summary measures about lifetime fertility and fertility in a narrowly defined period of time before the census, from which demographers seek to estimate current fertility rates. Thus two key questions provided by censuses are the number of children ever born and the number of live births in the last twelve months previous to the inquiry. These two are the main information, which allow demographers to estimates age-specific fertility rates (ASRF) and total fertility rates (TFR), as well as average parities or cohort fertility (Moultrie et al., 2013). Although, due to the abridged nature of the questions asked, the scope for internal validation and cross-checking of the answers given is limited.

1.2. Household surveys

The household surveys are the most widespread tool for collection of social, economic and demographic information on large group's population. They are important tools to provide intercensal data about socioeconomic characteristics of the population of a country. Unlike the population censuses, which investigate certain demographic, social and economic characteristics of the entire population, the basic household surveys are applied to a sample of households that are selected based on certain criteria. Usually the sampling of household surveys are drawn from Geographic Information Systems (GIS), representing the area to be covered and which is normally updated based on most recent census available (Arieira, 1995).

Virtually all Latin American countries have had some activity in the area household surveys. The current household survey system in Latin American countries dates back to the early 1960s, when many countries have initiated programs of household surveys' designed to measure employment, unemployment and some labor force characteristics (CEPAL, 1983). Thus, the origins of household surveys in Latin American countries can be traced to a seminar that was held in Mexico City in 1965, when a study by the US Bureau of the Census was presented.

Until the late 1960s, they only cover some socioeconomic groups and major metropolitan areas of the Latin American region. Over time, however, the research on income/expenditure came to be regarded as an important source of information to measure costs and living standards, by comparing the structure and composition of income and consumption of households. This demand has also triggered the surveys to increase their coverage and the socioeconomic information was expanded.

As living conditions require more detailed information and variables that are not strictly related to the labor market and/or economic activity, these surveys started to include more detailed issues such as housing conditions, demographic trends (migration, fertility, infant mortality), etc.In the case of Brazil, the National Household Surveys (PNADs), for example, since 1972 included questions related to month and year of birth of the last child born alive, and the number of children ever born, within and outside the home (IBGE, 2015; Wong, 1988).

1.3. Demographic Health Surveys

Many standardized data are also taken from the Demographic and Health Surveys (DHS). These surveys, along with the national surveys, are the most important data sources for estimating fertility, infant mortality and nuptiality. The range of information collected has broadened in scope to include gender issues (surveys increasingly include both men and women), HIV-AIDS, maternal mortality, violence against women, nutrition,etc (DHS, 2016).

Three important facts should be highlighted, however, in relation to these national survey programs. First, not all countries are covered by these surveys, and not all have expressed the wish to extend their thematic coverage. This is the case for Chile, Argentina, Uruguay and several Caribbean countries (as we can see in Table 1). Second, not all surveys are strictly comparable. Certain variations in fertility statistics, for example, may be due to a problem of data accuracy (Guzman et al. 2006). Third, the geographical extents of such surveys are also very limited.

Table 1: List of World Value Surveys and Demographic Health Surveys conducted in the region of Latin America

Country	WFS surveys	DHS surveys
Bolivia		1989 1994 1998 2003 2008
Brazil		1986 1991 1996 2006
Colombia	1976	1986 1990 1995 2000 2005 2010
Costa Rica	1976	
Dominican Republic	1975 1980	1986 1991 1996 1999 2002 2007 2013
Ecuador	1979-80	1987
El Salvador	1978	1985
Guatemala	1978 1983	1987 1995 1998-99
Guyana	1975	2004 2005 2009
Honduras		2005-06 2011-12
Haiti	1977	1994-95 2000 2005 2006 2012 2013
Mexico	1976-77	1987
Nicaragua		1997-98 2001
Panama	1975-76	
Paraguay	1979	1990
Peru	1977-78	1986 1992 1996 2000 2004-06 2007-08 2009 2010 2011 2012 2014
Trinidad and Tobago	1977	1987
Venezuela	1977	

Note: WHS: World Health Survey; DHS: Demographic and Health Surveys; Taken from Guzman et al. (2006) and adaptation from DHS (http://dhsprogram.com/).

Despite their limitations, the full birth history, provided by these surveys, has become the dominant source of estimates of fertility levels and trends for countries lacking complete birth registration (Avery et al. 2013). Hence, these reproductive health researches usually present very detailed information about the date of birth of each child for all women, and very useful information for studying fertility levels, trends and compare cohorts.

Thus, DHS historically has played an important role in certain countries as Colombia (DANE, 2014), Peru and Haiti, for example. Hence, due to lack of trustworthy information from census or civil registers, their official fertility estimates rely mainly on DHS information.

1.4. Vital registration systems

Civil registration is the first information collected as part of an ongoing vital registration system. In Latin American countries, the quality of these statistics covering recent decades is variable. Though coverage has improved in some countries, in others, where the registers of births and deaths are not exhaustive, the data collected are still very incomplete (Guzman et al. 2006). Generally, birth registers are of better quality than death registers, but they do not appear to be improving very much in countries where quality is poor, for example in Ecuador, Brazil, Dominican Republic and Venezuela, and much of this is associated with bad socioeconomic conditions (Lima and Queiroz, 2014; PLoS 2010; AbouZhar and Boerma, 2005). In these countries, the registration of births is generally incomplete, because parents often lack incentives to register births; or because babies who die shortly after birth may not be registered either as a birth or as a death; and late registration of births (for example, when the child attains school-going age) occur very often. It may mean that there is a delay of several years before all the survivors of the cohort born in a given year have their births registered (Moultrie et al., 2013).

However, the link between under registration of births and population social economics characteristics, in certain cases, is not very clear. Guatemala, for example, shows a relative high quality for births registers especially, despite the country's extreme poverty (Guzman et al. 2006). For all these reasons, the fertility estimates based on vital records are for many countries unreliable, especially those concerning mortality. On the other hand in some countries like Chile the proportion of late registered births declined steadily from 10-15% before 1980 to 5% in late 1980s and 1990s and further to around 1% in 2001–2004 (Zeman and Castro 2014).

However, experience has shown that achieving an improvement in the coverage and quality of vital statistics is not a simple task, because of the multiple factors is involved. According to Faijer(1994) the problems of vital statistics originate mainly from the operation of two types of factors:

1) Related to statement of death or birth by stakeholders (e.g. parents and relatives); and,
2) Those related to the registration system itself. The former can be considered "environmental", while this last is from administrative-institutional nature.

The "environmental" factors are a reflection of the social and economic development and cultural conditions in a country. They are caused by the lack of incentives to declare a birth (for example, the receiving of family allowances, health insurance, free education, etc.) or death (e.g. pensions or life insurance, burial authorization), and sometimes is difficult to convince responders to declare just to obey a legal requirement.

In addition, there are factors related to aspects of administrative and institutional nature, which might also affect negatively the decision to register. For example, difficulties of access the places of registration, the time spent in making arrangements and their potential costs. In this last case, it will be made by the stakeholders the assessment of registration costs, regarding the possible benefits that might result from that record (Faijer, 1994).

There are also the factors of administrative and institutional matter. As the vital statistics involve a large system within the state apparatus, with national coverage, ranging from too many administrative places, so there are many instances where errors can occur.

Hence, these two main factors can compromise registers in many Latin American countries. Thus concerning births registration in the region, the number of countries with good quality (over 90%) on birth registration increased significantly between 1965 and 2000 (from 20% to 25%). On the other hand, to the same extent, we see a growing number of countries where we have ***no information*** on the degree of birth registration (25%) in the last period, and the number of countries with up to 79% coverage of births has tripled during the same period.

Table 2: Classification of Latin American countries according to the degree of coverage of birth statistics

Rating	Period					
	1960-65	1975-80	1980-85	1985-90	1990-95	1995-00
Good (> de 90%)	45%	55%	55%	57.9%	54.5%	55%
Satisfactory (80 and 89%)	25%	25%	10%	5.3%	9.1%	0%
Regular (70 and 79%)	5%	5%	5%	5.3%	9.1%	15%
Deficient (< 70%)	5%	10%	5%	10.5%	9.1%	5%
No information	20%	5%	25%	21.1%	18.2%	25%
Total	20	20	20	19	22	20

Source: Bay, G. and Orellana. H. "La calidad de las estadísticas vitales en la América Latina". Taller de expertos en el uso de estadísticasvitales: alcances y limitaciones. LC/R. 2141. Santiago de chile, diciembre 2007.

1.5. Other sources of fertility data – The Human Fertility Database (HFD) and Human Fertility Collection (HFC)

The Human Fertility Database (HFD) and Human Fertility Collection (HFC) is not a data source in strict senses, but a compilation of data with goal to gather and provide as much as possible fertility data to a broad public.

The HFD is a joint project of the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany and the Vienna Institute of Demography (VID) in Vienna, Austria. The HFD has as main goal provide access to detailed high-quality data on cohort and period fertility to a broad audience of users, in order to facilitate research on changes and inter-country differences in fertility, monitoring, analyzing, comparing, and forecasting fertility as well as for studying causes and consequences of fertility change in the industrialized world (HFD, 2016).

The dataset is entirely based on official vital statistics and aims to provide important fertility measures such as age-, cohort- and birth-order-specific fertility rates(whenever possible), as well as crude, cumulative and total fertility rates, tempo-adjusted total fertility rates, mean ages at birth, standard deviation in mean ages at birth, parity progression ratios, and also cohort and period fertility tables for national populations or areas (HFD, 2013). As additional feature, the HFD provides the input data – birth counts and estimates of female population exposure obtained from officially recognized sources – from which these measures and fertility tables are computed(HFD, 2013).Besides, the dataset also offers high level of detail information; uniformity of methods; and data design and the display of order-specific fertility indicators, which should encourage fertility analyses, forecasts and further innovation in methodological research on fertility. However, the only Latin American country that fulfills strict data-quality requirements is Chile, which (since 2013) has launched fertility data from 1992 to 2005 in the HFD web page.

However, the Human Fertility Collection (HFC) was made as an extension of the Human Fertility Database (HFD). The HFC is less stringent and more flexibility related to the source as well as the collection of data. In this sense, it is able to integrate a wide variety of fertility data, whether regional and/or national as well as various subpopulations. The quality requirements for the data selection are also lower, allowing the geographic expansion of the collection to the least developed parts of the globe (HFC, 2016).

Additionally, the HFC collects estimates from various data sources, not necessarily official ones, including research data and reconstruction data for individual researchers or research teams. Thus, Human Fertility Collection provides many alternatives to official indicators fertility and is not limited to continuous series (HFC, 2016).This flexibility allow the HFC contains fertility data for countries and years that cannot be included in the HFD, including estimates of order-specific fertility for countries where birth order registration is restricted to marital birth order only. HFC features age-specific fertility rates, cumulative rates, total fertility rates and mean age at birth .Regarding Latin American countries, HFC displays up to now fertility data from Argentina, Brazil, Chile, Costa Rica, Cuba, Mexico and Uruguay. Further data collection enlargement is in progress under the scope of the Latin American Fertility Database.

In the following section, we will stress the fertility schedules from Brazil, as example, giving a central attention to different indirect methods and data sources, and the fertility schedule generated by those methods. Hence, for being the most populated country in the region, and to present a huge heterogeneity regarding socioeconomic developments, Brazil represents a good case study.

2. Brazil, a case study

2.1. Methods to access completeness of births

In the 1960s, the need to obtain estimates of demographic indicators in countries with poor data quality has stimulated demographers to develop indirect methods for estimation of fertility and mortality rates. In the last five decades, many the Latin American demographers have made constant use of these methods and continue to rely on them, although the fertility and mortality rates have already reached low levels in some countries (Cavenaghi and Alves, 2016). In this context, one of the main aspects that compromise fertility data in the region is the enumeration or under-registration of births (often present in different data sources) and this continues to impose challenges for many countries.

In Brazil, as in other Latin American countries, there is a long tradition in collecting information on fertility and mortality in population censuses, even before the year 1960. In the 1940 census, it was included the question on total live births and surviving to women with 15 years of age or older (and to women with 10+ years old from 1991). In 1970, it was included the information on children born in the previous year 12 months from the research. Since then, the use of such data has increased, especially in combination with indirect techniques, based on data from "children ever born" information, which allow us to estimate average parity (P), and "the number births in the 12 months previous to the census" that produces cumulated current fertility (F), and the relationship between these two measures has been used to estimate the correct level of Total Fertility Rates (TFR) (Cavenaghi and Alves, 2016).

In the next section, we will discuss main method used in the region to access and correct completeness of births, namely the P/F of Brass (Brass, 1964), an extension of it, the Synthetic Relational Gompertz model (see Moutrie et al. 2013) and the own-children method[4]. In addition to that, we will illustrate how these methods perform in the contexts of fertility decline and at sub-national level, and compare how its estimates perform with distinguished data sources and other official and unofficial estimates.

2.2. Data and methods

We make use of data from population censuses from 1970 to 2010. We tabulate the information on the number of children ever born classified by five-year age group of mother, taken from each census, and the number of births during the year preceding each censuses, classified by five-year age group of mother, and the number of women in each five-year age group from each inquiry. These tabulationsare made for the urban Rio Grande do Norte (RN) State, a region that has experienced rapid changes in mortality and fertility (IDEMA, 2002; Fossa e Bezerra, 2002), and also has historically shown lower quality of vital registration (IBGE, 2003; Paes, 2006; Lima and Queiroz, 2014).

[4] This method by Lima (2013) to estimate series of fertility schedules for the country as whole from 1966 to 2010. These estimates can be accessed at http://www.fertilitydata.org/cgi-bin/data.php.

Further tabulations are also made for the country as a whole, but only for inquiry years 2000 and 2010. Thus, we also make use of births registers publicly available by Ministry of Health (DATASUS) and other official estimates provided by the National Statistical Office (IBGE) and Demographic Health Survey (PNDS). This time, we compare the results of different data sources and methods combined in a scenario of fertility decline to below replacement level.

2.2.1. P/F of Brass (One census method)

According to Brass (1964) data on current fertility in demographic censuses (or household surveys) are generally underestimated for all age groups, and some empirical evidences showed that this underestimation has no difference by the age of women, a fact which led to adjust factor for TFR as comparing cumulated fertility (F) from a period with the average parity distribution (P) measurement of a cohort fertility (United Nations, 1983). Brass defined P as the average parity (cumulated lifetime fertility) of a cohort of women up to a given age, and F to be closely related to the cumulated current (period) fertility up to that same age. The P/F ratio expresses these two quantities in relation to each other in the form of a ratio for each age group (Brass, 1964).

However, the comparison between these two measures needs to take to account that comparison of cohort and period fertility has to deal with the probable shifting of the data on recent fertility brought about by the question being based on the age of the mother at the time of the inquiry rather than her age at the time of her most recent birth(United Nations, 1983; Moultrie et al., 2013). Second, while the accumulation of period fertility to any given age will reflect the fertility experience of all women up until that age, the average parities typically calculated reflect women aged at the midpoint of 5-year age group.

In addition to that, the method relies on the fact that if fertility schedule has been constant for an extended period of time, cohort and period measures of fertility will be identical. In other words, under conditions of population stability, with constant fertility and mortality rates, the cumulated fertility of a cohort of women up to any given age will be the same as the cumulated fertility up to that same age in any given period, and if the data are free of error, the P/F ratio would equal 1 in every age group (United Nations, 1983).

There is also the necessity of constant mortality that relies in the fact that there are no substantial mortality differentials by the fertility of mother, in other words, the surviving women do not have significantly different levels of childbearing from deceased women, and the cumulated fertility of a cohort of women up to any given age is the same as the average parity in that cohort (Brass, 1964; United Nations, 1983). However, according to Moultrie et al. (2013), this last assumption is not very important as even if there are differentials in the fertility of living and deceased women, because in most populations the magnitude of female mortality in the reproductive ages is very small and the effect of differential survival will therefore be small.

Although, when fertility has been falling, cumulated life time fertility (P) would be greater than cumulated current fertility (F). In this case (in the absence of errors in the data) the P/F ratio would depart from unity systematically with increasing age of mother. We would expect the P/F ratio to be fairly close to unity at the youngest ages, because even by women's mid-twenties one would not expect significant deviation of cumulated period fertility from cumulated lifetime cohort fertility, as most of the births to women in that cohort would have happened fairly recently. In this case, the P/F ratio derived from women aged 20-24 at the time of a survey is held to be the most reliable indicator of the quality of the fertility data collected. Conveniently, the supposition is that the average parities of younger women are usually fairly accurately reported, at least relatively to those of older women (Moultrie et al., 2013; United Nations, 1983).

However, in cases of strong fertility decline associated with variation oflevel and pattern of current fertility might affect the method, presenting an error of 5% to the adjusted TFR (Moultrie and Dorrington, 2008). In other to avoid such error, other methods have been implemented, commonly known as extensions to Brass to P/F ratio, but based on estimation of age-specific fertility from the increment of cohort parities between two inquiries (United Nations, 1983).

2.2.2. *The Synthetic Relational Gompertz (SRG) model (Two censuses method)*

The Synthetic Relational Gompertz (SRG) model is an extension of the relational Gompertz[5] method for the estimation of age-specific and total fertility and makes use of two sets of parity data, collected at different points in time, in combination with estimates of current fertility in the intervening period based on reports of recent births classified by age (Moultrie et al., 2013). The relational Gompertz method seeks to solve the errors commonly found in fertility data associated with too few or too many births being reported in the reference period, and the under-reporting of lifetime fertility and errors of age reporting among older women. The extension of this method allows changes in fertility to be taken into account and it is designed to be applied to censuses or surveys conducted either 5 or 10 years apart (Moultrie et al., 2013; United Nations, 1983).

In such circumstances, the change in the average parity of the cohort can be estimated, as the survivors of a cohort of women at the first inquiry can be identified at the second. By cumulating the sequence of parity increments for different cohorts, during the period between the inquiries, we estimate average parities for a hypothetical cohort experiencing the fertility implied by the observed parity increments (Moultrie et al., 2013).

The period fertility rates that are compared with these synthetic cohort estimates should ideally refer to the entire period between the two inquiries that asked about lifetime fertility. This comparison gives the true estimate of TFR for the period in question (Moultrie et al., 2013).

[5] The relational Gompertz method is a refinement of the Brass P/F ratio method that seeks to estimate age-specific and total fertility by determining the shape of the fertility schedule from data on recent births reported in censuses or surveys while determining its level from the reported average parities of younger women.

2.2.3. The Own-Children Method (OCM) and the reconstruction of fertility history

The Own-Children Method (OCM) is a census or survey-based reverse-survival technique designed to estimate age-specific fertility for the period preceding the enumeration (Retherford and Cho, 1978). In the absence of detailed data on reproduction, this reverse-survival method, also called reverse-projection, uses current age structure and assumptions about mortality in a given population to reconstruct age structure of this population in the period preceding the survey.

The estimation follows an explicit order. First, the enumerated children are matched to their mothers living in the same household (Retherford and Cho, 1978; Cho et al., 1986). These matched (i.e., 'own') children, classified by their age and by mother's age, are "reverse-survived" to estimate the number of births by age of mother in previous years. Reverse survival is similarly used to estimate the number of women by age in previous years. After adjustments for under-enumeration (mainly due to undercount and age misreporting) and unmatched (i.e., "non-own") children, age-specific birth rates are calculated by dividing the number of reverse-survived births by the number of reverse-survived women (Retherford and Cho, 1978).

The allocation of children (from zero to 14 years-of-age) to the respective mother (from 15 to 64 years-of-age) in the census data is based on the information on their relationship to the head of the household, to other women in the household, and the mothers' information on the number of children ever born and children surviving. Once the data were matched, we have tabulated age of children by age of their mothers. In some cases, it was not possible to match children with their mothers. This occurs either because the information about the relationship to the head of the household is insufficient, or because some children do not live in the same household as their mother or because the mother is deceased. These children are so-called 'other than own,' or, simply 'not-own' children in the household. They are proportionally distributed by age of mother using the age distribution of the mothers with identified own children. Besides this addition of 'not-own' children, age misreporting and age under-enumeration can also bias the data. Finally, survivorship function is applied to generate the number of births by age of mother for the years prior to the survey (Retherford and Cho, 1978).

The OCM has important advantages over other methods that reconstruct age-specific fertility rates from census or survey data (Retherford and Cho, 1978). The method can simply be applied to an existing census or survey, and it does not require additional data, except for life tables or mortality estimates used in computing reverse survival. If these are not available, survivorship may also be estimated directly from the census or survey by linking estimates of child mortality — based on commonly asked questions regarding total children ever born and children surviving — to appropriate model life tables. Furthermore, the OCM is not very sensitive to life table estimation errors under mortality levels currently prevailing in most parts of the world (Retherford and Cho, 1978; Cho et al., 1986).

The results of these estimates are available on the website of the Human Fertility Collection (HFC). In the HFC web page there are series of fertility rates for Brazil as a whole, estimated for the years 1966 to 2010 (Lima, 2013).

3. Results

a) Comparison between traditional Brass and Synthetic Relational Gompertz models in fertility change scenario

Figure 1 shows the performance of the P/F adjustment versus the Synthetic Relational Gompertz. The analysis here is a scenario of falling fertility. Thus, as an example, it was taken the Brazilian urban sub-population of Rio Grande do Norte (RN) State. The choice was not at random, whereas this region has experienced a huge fertility decline between the years 1970 and 2010, with an observed TFR of 5.11 in 1970 and 1.53 in 2010, representing a decrease of almost 70% in 40 years period.

In order to test the performance of the methods, we have created two scenarios, one with level correlation in the age-specific fertility rates given by P/F ratio from the age group 20-24 years old, and other one without correction. All the estimates use same census data.

From 1970-80,with exception of the group of 15-19 years of age, we observed a sharp decline in fertility, especially after the age of 25. The P/F adjustment has raised the level of fertility to a TFR of 7.18 (increase of approximately 41% in the ASFR) in 1970 and to 4.72 (15% in the ASFR) in 1980. For the SRG, however, we see no significant changes in the pattern and level of fertility, before and after the Brass correction was applied. Therefore, the SRG estimates give a TFR ranging from 5.48 to 5.46 (with and without correction, respectively). These values are very similar, showing a good consistency of the method in a scenario of strong fertility change.

In order to decide which scenario is most likely to occur and so to validate our estimates we have two possibilities:

1) TFR 5.11 (1970) ⟹ 5.48 (Approximately 1975) ⟹ 4.09 (1980)
2) TFR 7.18 ⟹ 5.46 ⟹4.72

Taking to account that the fertility is falling, the common sense leads us to believe that the scenario two is more plausible here.

In the second period between 1980 and 1991, the decline in fertility was more pronounced and fast than in the previous one (about 40% fertility decline in the period), and in almost all ages (again with exception the group of 15-19 years old that shows slight increase in its specific rates). In 1991, the adjustment P/F was 16%, raising the TFR to 2.83.

The TFR estimated among the period by the SRG has not changed its value (TFR of 3.48).The same can be said about the pattern of the fertility schedules, which practically overlapped. Moreover, once again we have two scenarios of fertility change to choose.

1) TFR 4.09(1980)⟹ 3.48(Jan 1986) ⟹2.45 (1991)
2) TFR 4.72 ⟹ 3.48 ⟹ 2.83

Since we did not observe changes in both pattern and fertility level estimated by SRG, we take this estimate (middle year between censuses) as a benchmark for comparison. The first scenario, when we analyze the first five-year period the fertility fell almost 15%, and 30% from 1986 to 1991. The second scenario shows a fertility drop of 26% (between 1980 and 1986) and 17% (between 1986 and 1991). Once again it is more plausible to believe that the biggest drop occurred from 1980-86 (scenario 2), while we expected that the strongest fertility falls take place in the past (when the fertility levels were higher).

From 1991 to 2000, the decline in fertility was less pronounced (16-19% in the period, considering the observed or corrected data by the Brass method). The SRG estimates continued to have very similar values, with a TFR ranging from 2.51-2.53 (in approximately August 1996). The fertility pattern was also very similar between estimates. This fact might point out to the robustness of the method.

This time, the choices scenarios are:

1) TFR 2.45 (1991) ⟹ 2.51 (1996) ⟹2.06 (2000)
2) TFR 2.83 ⟹ 2.53 ⟹ 2.30

Both possibilities are plausible, but it is more likely to believe that fertility fell from 1991 to 1996 and it was not kept stable. Therefore, we disregard the first scenario, without P/F correction.

For the last period of analyses (2000 to 2010), the decline in fertility was extended to adolescent and young adult ages (under 30 years). The P/F adjustment indicates an increase in the sub-enumeration of births of 20%. This is something that goes against of what is expected, i.e. a continued improvement in data collection over the years in the country. However, the effect of the sharp drop in teenage fertility in the country between 2000 and 2010 may be affecting the estimates of P/F. Moreover, this effect seems not to affect the relational Synthetic Gompertz model, which continues to show very similar fertility schedules in both scenarios.

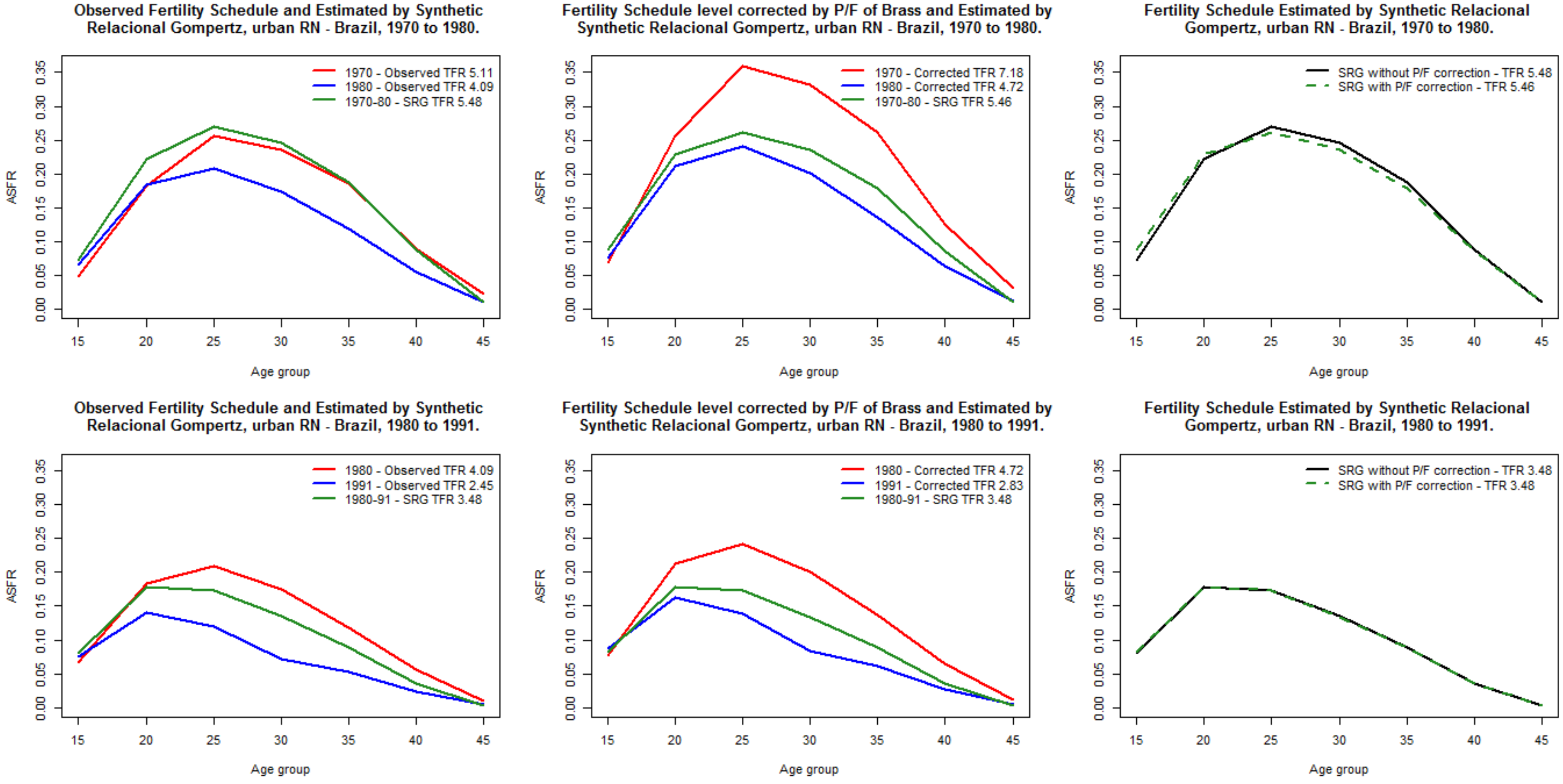
The scenarios for 2000–2010are :

1) TFR 2.06 (2000) ⟹ 1.74 (2005) ⟹ 1.53 (2010)
2) TFR 2.30 ⟹ 1.74 ⟹ 1.84

Differently from previous analyses, the second scenario with fertility correction in 2010 seems less possible to happen. The likelihood that fertility has increased between 2005 and 2010 is very low. Thus it is hard to believe that the Brass method is reliable in this case.

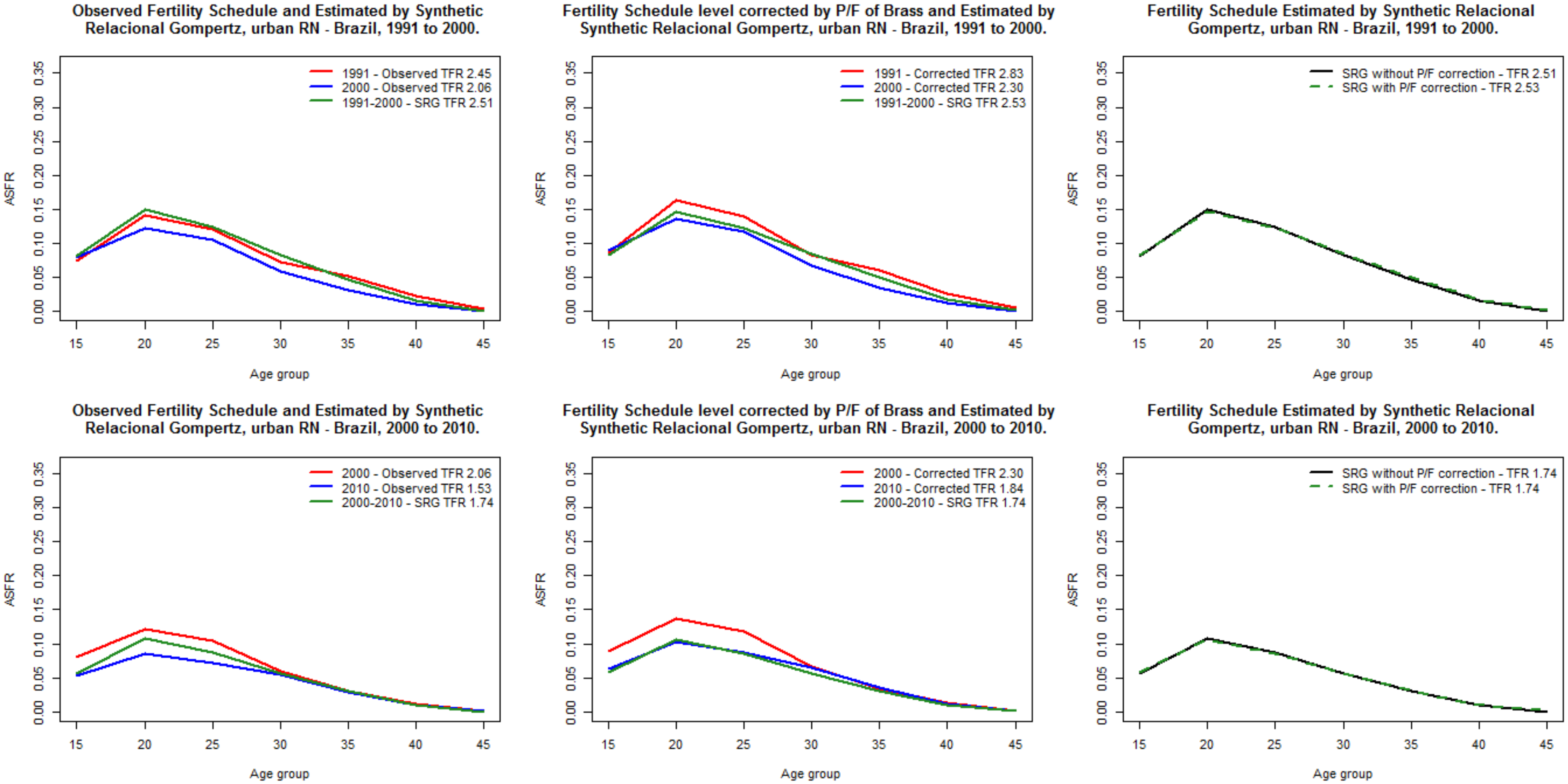
In the next point, Figure 2 and Table 3, we will keep comparing both methods, but this time using different data sources, i.e. information derived from the civil registration and censuses for the country as whole.

Figure 1: Comparison between Fertility Schedules estimated by traditional Brass method and by Synthetic Relational Gompertz (SRG) model. Rio Grande do Norte, 1970 to 1980 and 1980 to 1991.



Source: Brazilian Population censuses 1970 to 2010

Continuation Figure 1: Comparison between Fertility Schedules estimated by traditional Brass method and by Synthetic Relational Gompertz (SRG) model.Rio Grande do Norte, 1991 to 2000 and 2000 to 2010.



Source: Brazilian Population censuses 1970 to 2010.

b) Comparison between data sources, Brass and SRG models

In this section, our concern is to compare the performance of the methods through their application to different data sources during rapid changes in schedules fertility. The Table 3 and Figure 2 show the estimates when we take as source the civil registration or the census data for the years 2000 and 2010.

First, again, we notice that the SRG proves to be the most robust method to estimate fertility schedules. Regardless of the source used, the estimated TFR ranged from 1.84 to 1.85 and the fertility structure are also overlapped (see Figure 4that compares the different patterns and fertility curves), independent of the method in data source. These values are also quite similar to the estimates given by PNDS(or DHS) 2006, which considers the reproductive behavior of the 36 months (i.e. 2004-2006) prior to the interview, and the estimated Total Fertility Rate is 1.8 children per woman.

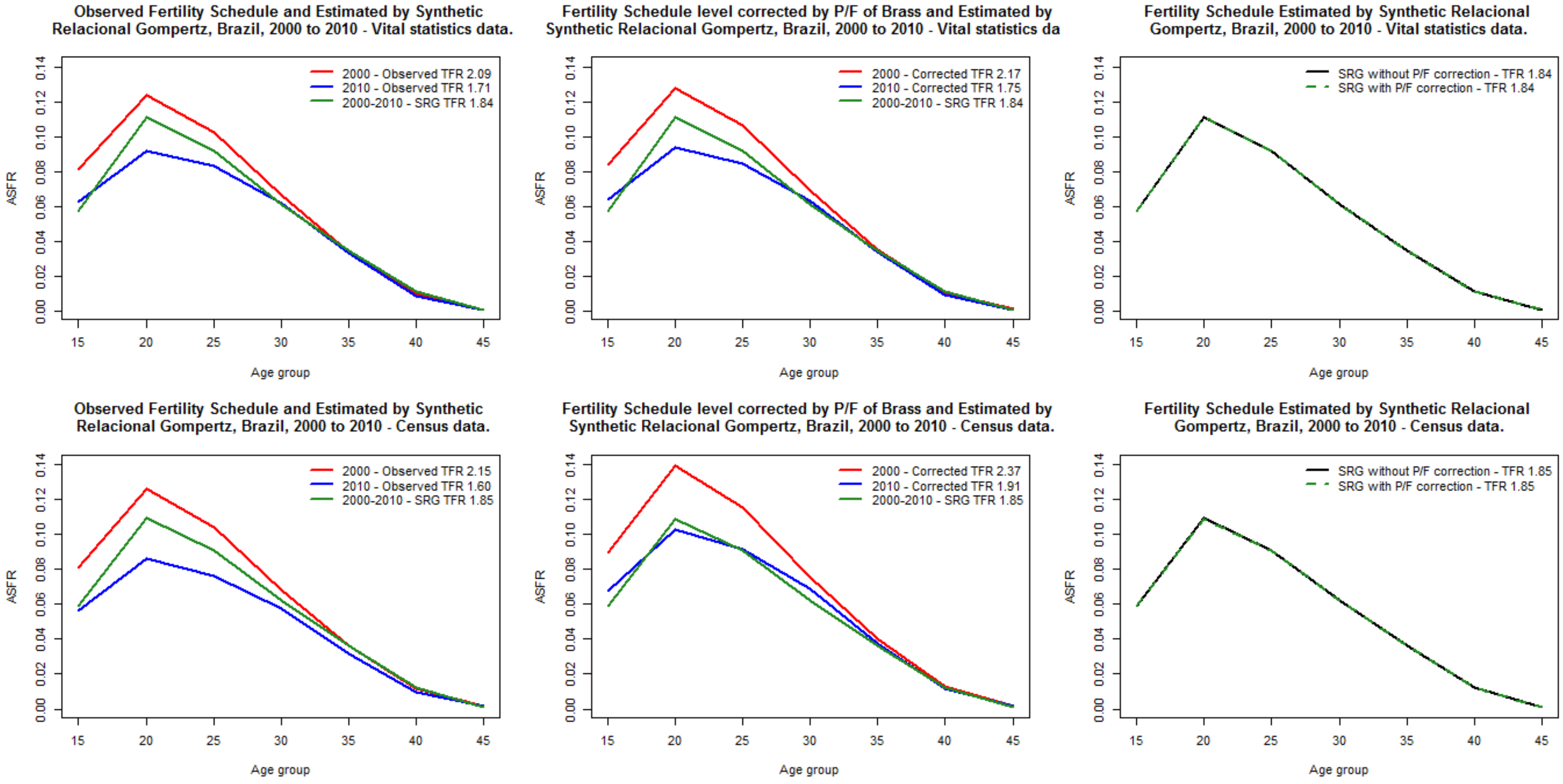
The fertility structures differ slightly between the sources. However, the adjustment by Brass method generates different estimates, according to the source used. The civil records present always lower fertility levels than census, and the P/F ratio also points to small corrections necessary (4% in 2000 and 2% in 2010),while considering civil registration (Cavenaghi & Alves, 2016). Considering the census data, however, the TFRs presented the highest level in 2000, but in 2010 the fertility declines to 1.60 (observed data). The Brass correction is also more pronounced for census data (indicating sub-enumeration of birth in the order of 19.3%), increasing the TFR to 1.91 in 2010 (Cavenaghi and Alves, 2016). This is something unexpected and unlikely to happen, and further research need to be conducted.

Table 3: Comparison between TFR according to different data sources and estimates, Brazil, 2000 to 2010.

Source of data	Year of the inquiry			
	2000	2005 (SRGestimates)	2010	Plausibility
Vital Registration	Observed data			
	2.09 ⇒	1.84 ⇒	1.71	Plausible
	P/F adjustment applied			
	2.17 ⇒	1.84 ⇒	1.75	Plausible
Census	Observed data			
	2.15 ⇒	1.85 ⇒	1.60	Plausible
	P/F adjustment applied			
	2.37 ⇒	1.85 ⇒	1.91	Less plausible

Sources: DATASUS 2000 and 2010 and Censuses – IBGE, 2000 and 2010.
Note: Correctionappliedto the ASRFs of 4% in 2000 and 2% in 2010, for vital registration data. The official TFRs estimates were 2.28 in 2000 and 1.82 in 2010 (DATASUS, 2004).
Correction in the ASRF of 11% in 2000 and 19.3% in 2010 for census data. The official TFRs estimates were 2.37 in 2000 and 1.87 in 2010 (IBGE, 2013).
TFR estimates of 1.80 according to PNDS (2006)

Figure 2: Comparison between Fertility Schedule estimated according to traditional Brass method and by Synthetic Relational Gompertz (SRG) model and different data sources. Brazil, 2000 to 2010.



Source: Brazilian censuses and vital registration 2000 to 2010.

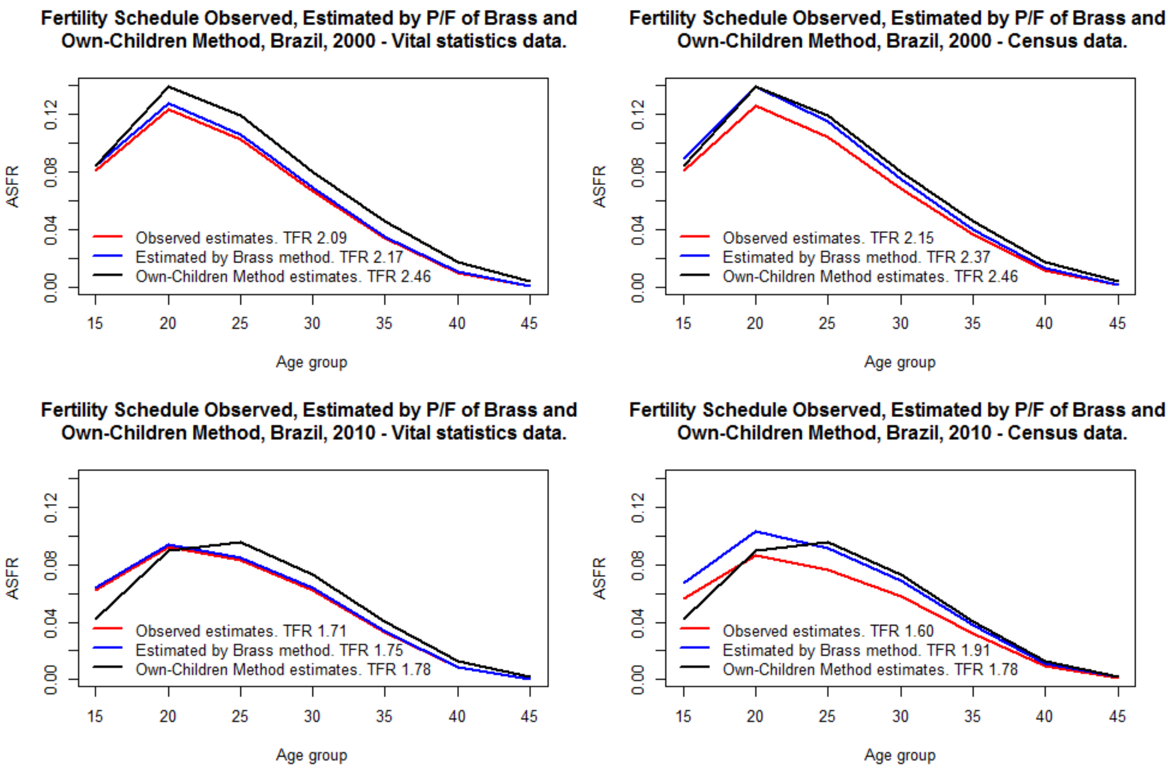
c) *Comparison between data sources, Brass and OCM*

Regardless of the source data, the comparison between Brass and OCM shows that in 2000 both methods produce relatively similar estimates (mainly considering census as data source), with only minor differences at the older ages of the reproductive period. The level of fertility, after correction, was also quite similar (using Brass adjustment TFR of 2.37 vs. 2.46 based on OCM estimates). This is expected since the OCM estimates and the rates (based on the information of births in the 12 months prior to the survey) are from the same data source.

Considering the estimates with civil registers as source, again we notice that fertility shows a relatively similar structure, but with tiny differences in levels. In both cases, the differences at the younger ages are relatively small, especially for the civil records.

However, despite the small differences in level, in 2010 the fertility structure differs considerably between methods, and the main differences are encountered at the young ages (15-19 years old) in the case of comparison with civil records. Although, taking to account the census, in virtually all ages there are differences. As we can see, the estimates for the OCM show a fertility peak at the age group 25-29, whereas estimates based on other sources still present the 20-24 age group as the highest reproductive level.

Figure 3: Comparison between Fertility Schedules, Observed, Estimated by P/F of Brass and Own-children method (OCM). Brazil, 2000 and 2010.

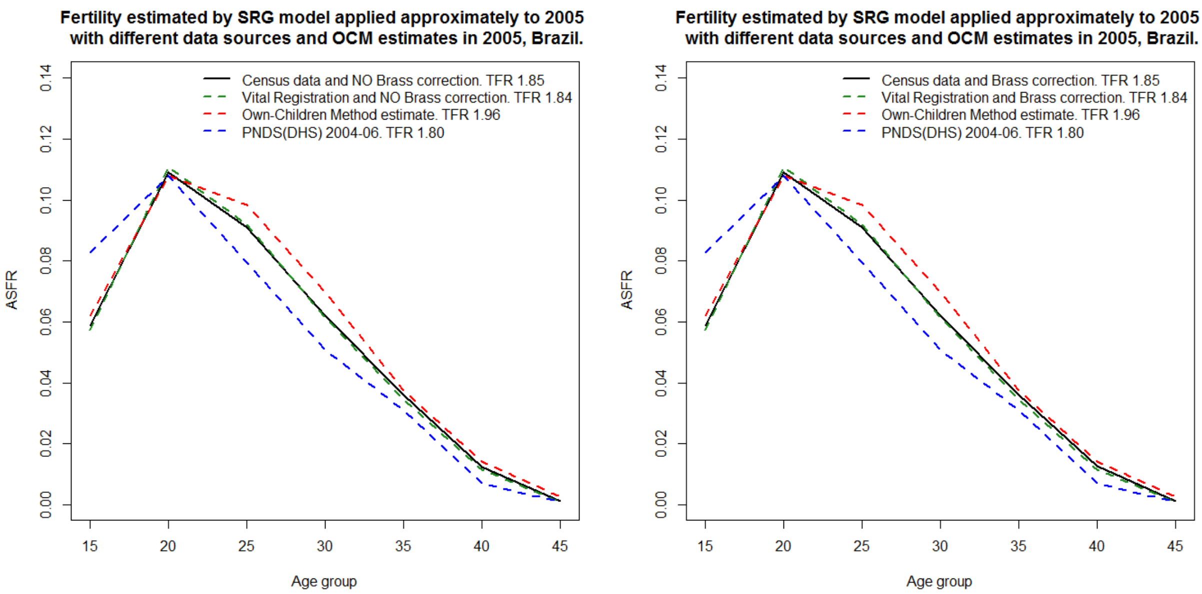


Source: Brazilian censuses and vital registration 2000 and 2010.

It is important to highlight that the OCM has several methodological limitations. The first limitation is related to distortions in the age pattern of fertility and the estimated fertility trend, caused by age misreporting. Age misreporting typically causes year-to-year fluctuations in own-children fertility estimates. Another drawback relates to the method itself. The OCM does not work with complete childbearing history, since the method excludes children who were not alive or present in the household at the time of the census. This also limits estimations of birth intervals. Finally, the constraint of allocation errors must be considered, concerning "own children" and "other than own children", even if this bias tends to be less serious than age misreporting (Cho et al., 1986). However, according to Avery et al. (2013), all these problems affect the Own-Children method only modestly.

d) *Comparison between data sources, Brass, SRG and OCM*

Figure 4: Comparison between Fertility Schedules estimatedby Own-children method (OCM), aSynthetic Relational Gompertz model, with and without Brass correction, and different data sources. Brazil, 2005.



Source: Brazilian censuses and vital registration 2000 to 2010.

Figure 4 shows the estimates of fertility schedules according to Demographic Health Survey (PNDS 2006), and indirectly estimated by a Synthetic Relational Gompertz model (applied to civil registration and censuses) and the own-children method. This time, our concern is to compare different sources and methods of fertility estimation combined.

In the first case, the current fertility rates, used in the SRG model, are not subjected to level adjustment by the P/F of Brass. Secondly, the fertility schedule is subject to level correction during the intercensal years, i.e.in 2000 and 2010. This procedure is applied to the vital registration and census data. In both cases, the generated rates represent the number of children per woman for a period approximately to August 2005. The PNDS correspond to a reproductive period of approximately 36 months before the survey, e.g. it produces estimates for the period 2004-06. The third and last estimate refers the ASFRs obtained by own-children method in the year 2005.

On the one hand, we did not observe expressive differences between the estimates generated by Synthetic Relational Gompertz model. This is true regardless of the data source (whether it is census or vital statistics), and whether or not previous correction are applied to fertility levels by the P/F ratio. This may be the indication of the method robustness. In addition to that, the average parity estimated by the synthetic cohort seems to capture well the reproductive changes during intercensal period, as well as the relational model of Gompertz confines well the pattern inherent in these rates.

On the other hand, even with similarity in fertility level, the DHS shows a large difference in terms of structure. This last still indicates a high level of teenage fertility from 2004-06, if compared with the SRG model, though the fertility among the other ages is clearly underestimated. Furthermore, the own-children method also appears to generate a small over-estimation of fertility rates between the ages of 25 to 30 years old, if we compare to the SRG estimates. This is also reflecting in the total fertility rate of1.96, which is very close to the level of population replacement, while the other estimates are bit below this level.

4. Guidelines for the future

In this work we have tried to give a small overview of the main data sources, imperatively used in Latin America as a tool to provide births information. In addition to that, we have applied a small exercise, using three main indirect methods: (1) P/F ratio of Brass, (2) Synthetic Relational Gompertz model and (3) the Own-children method, in order to see how there are performing, while applied to different sources and demographic circumstances, using Brazil a case study.

Our analyses show how sensible are the births information and registers (provided by the different data sources) as well we the fertility schedules, estimated by indirect methods, that we have analyzed in Brazil. The different data present distinguished results, given a variability of TFR varying from 40% to 2%, depending of the method and data considered. Nevertheless, not only the method, but also the fertility shape differs among methods and data.

In this context, we advocate for further steps, in order to improvise the births information and estimates of fertility in the region:

1) To invest and to promote more accurate information about: age of mother, place of residence of the mother and births information (Tacla, 2009). This will be the first desirable stage. Second, once the first is achieved, to invest in socioeconomic characteristics of the father or mother in civil registration.
2) Concerning the methods, we judge the combination of SRG model and Brass P/F ratio as good tool to evaluate completeness of births and to estimate “right” the fertility shape, especially in a scenario of fertility change;
3) To promote other alternative data sources, providers of fertility information. Hereby, we highlight the (future) project of Latin American

Fertility Database in HFC and the Latin American census data in CFE database http://www.cfe-database.org/database/

References

AbouZhar C, Boerma T. Health information systems: the foundations of public health. Bull World Health Organ 2005; 83:578-83.

Arieira, M. L., (1995) Pesquisas Domiciliares Na América Latina e no Caribe: Uma Avaliação. In: Schwartzman, S. Encontro Nacional de Produtores e Usuários de Informações Sociais, Econômicas e Territoriais.

Avery, C. et al. 2013. "The 'Own Children' fertility estimation procedure: A reappraisal." Population Studies 67(2): 171-83.

Bay, G. and Orellana. H. (2007) Bay, G. y Orellana, H. 2007. La calidad de las estadísticas vitales en América Latina. (Versión preliminar para discusión). Documento preparado para el taller de expertos en el uso de estadísticas vitales: alcances y limitaciones. Santiago, Chile, diciembre, 2007.

Bravo, D.; Larrañga, O.; Ruiz, Magda & Zamorano, F. (2013) Informe final comisión externa revisora del censo 2012. 41p.

CAVENAGHI, Suzana M. and ALVES, José Eustáquio Diniz. Qualidade das informações sobre fecundidade no Censo Demográfico de 2010. Rev. bras. estud. popul. [online]. 2016, vol.33, n.1, pp.189-205.

CEPAL (1983) Las Encuestas de Bogares eu Amenca Latina, Cuademos de la CEPAL, no. 44, Santiago de Chile: Publicacion de las Naciones {}nidas

Cho, L. J., R. D. Retherford, and M. K. Choe. 1986. The own-children method of fertility estimation. Honolulu: University of Hawaii Press, 188 p.

Demographic Health Survey – DHS (2006). DHS-Overview. Available at: http://dhsprogram.com/What-We-Do/Survey-Types/DHS.cfm

Faijer (1994). La Calidad De Las Estadisticas Vitales En America Latina. Seminario Internacional De Calidad Estadistica. DANE-Colombia.

Fossa, M. G. &Bezerra, G. M. A. (2002). "Reposição da força de trabalho no Rio Grande do Norte (1991 a 2001)". Revista Electronica de Geografia y Ciencias Sociales, Universidad de Barcelona, vol. VI, nº 119 (69), 2002.

Guzmán, J. M. et al. 2006. The Demography of Latin America and the Caribbean since 1950. Population-E, 61(5-6), 519-576.

HFD (2016). Human Fertility Database. Max Planck Institute for Demographic Research and Vienna Institute of Demography. Available at www.humanfertility.org

HFC (2015). Human Fertility Collection. Max Planck Institute for Demographic Research and Vienna Institute of Demography. Available at www.fertilitydata.org

IBGE (2003). Estatísticas do Registro Civil 2003. IBGE, 273 p.

IBGE (2015). PNAD: Um Registro Histórico da Pesquisa Nacional por Amostra de Domicílios, 1967-2015. 178p.

IDEMA, (2002). Perfil do Estado Rio Grande do Norte. Natal/RN.

Jaspers D., Poulard S., 2002, "El software Redatam para divulgación y análisis de datos censales", Notas de Población 29(75) pp.321-341.

Lima, E. E. C. (2013). Reconstructing Brazilian Fertility Rates by Age From 1966 To 2010 – Technical Note. Human Fertility Collection documentation. Available at: http://www.fertilitydata.org/cgi-bin/country.php?code=bra

Lima, E. E. C., & Queiroz, B. L. (2014). Evolution of the deaths registry system in Brazil: associations with changes in the mortality profile, under-registration of death counts, and ill-defined causes of death. Cadernos de Saúde Pública, 30(8), 1721-1730.

Moultrie, T. A., & Dorrington, R. (2008). Sources of error and bias in methods of fertility estimation contingent on the P/F ratio in a time of declining fertility and rising mortality. Demographic Research, 19, 1635.

Moultrie, T., Dorrington, R., Hill, A., Hill, K., Timaeus, L., &Zaba, B. (2013). Tools for demographic estimation. International Union for the Scientific Study of Population (IUSSP).

PAES, N. A. (2006). Avaliação das estatísticas vitais Para estudos de causas de morte dos adultos para os Estados e capitais do Brasil, 2000. In: Trabalho apresentado no XIV Encontro Nacional de Estudos Populacionais, ABEP. MG – Brasil, de 20-24 de Setembro de 2006.

PLoS Medicine Editors. Can we count on global health estimates? PLoS Med 2010; 7:e1001002.

Retherford, R.D., and Cho, L.J. (1978). Age-parity-specific birth rates and birth probabilities from census or survey data on own children. Population Studies 32(3): 567–581.

Sánchez-Albornoz N., 1977, La población de América Latina. Desde los tiempos precolombinos al año 2000, Madrid, Alianza Universidad, 321 p.

Tacla, O. (2009) Seminario- Taller "Los censos de 2010 y la salud. Cobertura De Las Estadísticas Vitales APartir De La Información Recogida En CensosO Derivada De EllosExperiencias Y Lecciones Aprendidas Em La Región. Noviembre de 2009. Access via the link: http://www.cepal.org/celade/noticias/paginas/5/37635/odettetacla.pdf.

United Nations (1983). Manual X: Indirect Techniques for Demographic Estimation. United Nations.

Wong, L. R. (1988). Avaliação Das Informações Sobre Fecundidade Da Pnad-84. O Seminário de Avaliarão das PNADs da Década de 80, Nova Friburgo, 13-15dejunho de 1988.

Zeman, K. andCastro, R. (2014). Age-specific fertility rates calculated from official birth counts and population exposures, 1990-2011. Human Fertility Collection documentation. Available at: http://www.fertilitydata.org/cgi-bin/country.php?code=chl