Assessment of subnational level birth registration data and estimation of fertility in South Africa

Tawanda Madamombe (Mr)

University of Cape Town

Thesis submitted to the Faculty of Commerce of the University of Cape Town in partial fulfilment of the requirements for the degree of Master of Philosophy in Demography

March 2021

PLAGIARISM DECLARATION

I, Tawanda Madamombe do hereby declare that this research is my original work, produced with supervisory assistance from Prof T A Moultrie. I acknowledge that copying someone else's work or any part of it, is wrong and hence the work of other authors has been duly cited or referenced. Neither the whole nor any part of this research has been submitted for another degree at this or any other university.

Signature: Signed by candidate Date: 14/03/2021

ABSTRACT

Birth registration data forms part of vital statistics. It is the right of the child to be registered immediately after birth and to acquire an identity and a nationality as stipulated by the Convention on the Rights of the Child (UNICEF 1989). The assessment of birth registration is important to help authorities in the processes of planning and decision making. This study investigates birth registration data at a district level with the aim to establish the data's usefulness in determining reliable completeness of births and estimating total fertility rate (TFR). The study further analyses the spatial relationships between respective districts to each other based on levels of birth completeness and total fertility rate. The Geographical Information Systems (GIS) technique of the Global Moran's Index spatial autocorrelation is used to examine the spatial distribution of completeness and TFR. The Indirect method of relational Gompertz model is used to calculate robust estimates of actual births that occurred in the twelve-month period before 2011 Census and 2016 Community Survey, respectively. Then, the Hauer and Schmertmann (2020) method of determining fertility was used to validate results from the Gompertz model.

The study establishes that there was an improvement in the promptness of birth registration between 2011 and 2016, highlighted by an 82% completeness in 2011 that increased to 85% in 2016 for births that were registered within the same year of occurrence. This is evidence that mothers are registering births at younger ages than before. The TFR decreased from 2.55 in 2011 to 2.28 in 2016. Apart from that, the study illustrated that districts with higher completeness levels tend to be in major urban agglomerations. However, no spatial relationship could be established meaning that the neighbouring districts do not follow any pattern when compared to each other. It was also noted that districts with low fertility are clustered near major cities. Although there are issues with data at lower levels of disaggregation such as districts, it has been shown that the use of robust methods produces results that help to give meaningful insights of birth registration data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CMC	Century-Month Code
CRVS	Civil registration and vital statistics
CS	Community Survey
CWR	Child to Woman Ratio
DHA	Department of Home Affairs
DHS	Demographic and Health Surveys
ESRI	Environmental Systems Research Institute
GIS	Geographical Information Systems
HDSS	Health and Demographic Surveillance System
MICS	Multiple Indicator Cluster Surveys
MYE	Mid-Year Estimates
PAHO	Pan American Health Organization
PES	Post-enumeration survey
SDG	Sustainable Development Goal
Stats SA	Statistics South Africa
TFR	Total Fertility Rate
UCT	University of Cape Town
UN DESA	United Nations Department of Economic and Social Affairs
UNICEF	United Nations Children's Fund
UNSD	United Nations Statistics Division
WHO	World Health Organisation

ACKNOWLEDGEMENTS

I would like to profusely thank my supervisor Prof T A Moultrie for his tireless support and guidance during my research. He was patient, insightful and always availed himself to enlighten me on various components of my research. I also extent my gratitude to Prof R Dorrington for his assistance with child mortality data as well as the use of the Thembisa Model. I would like to thank Thomas Slingsby, Nicholas Lindenberg at the UCT GIS Laboratory. Many thanks go to my close friends, colleagues and relatives who kept on encouraging me to pursue academic excellence. My love goes to Eve, Liam and Zane, whose love in return, keeps me going.

1. INTRODUCTION

1.1 Background and context

Since fertility is the key driver of population growth, there is need to understand the levels and trends of birth registration data. Notably the analysis of birth data, particularly at subnational level, helps to plan the provision of services such as health and education. PAHO (2017) notes that birth registration data by subnational areas can help authorities to identify high risk areas for diseases and this leads to solutions such as targeted health campaigns and raising awareness.

This study seeks to establish the accuracy, validity and reliability of birth registration at subnational level; that is, at provincial and district levels in South Africa. Statistics South Africa (Stats SA) annually publishes birth registration data at the aforementioned subnational levels. Census and Community Survey data, and some indirect methods will be used to check and assess the live births records.

There has been substantial attention on the study and analysis of birth registration data at national and provincial level with little focus on district data analysis. Statistics South Africa produces birth registration data for all areas of aggregation and it is made available on its website. The country's birth registration data is also summarised in Stats SA's Recorded live births statistical releases that are published every year. Stats SA reports that the national birth registration completeness is estimated to be 88.6% between the 2011 census and 2016 community survey period (Statistics South Africa 2020b). It was noted that for the years 2015-2019 births registrations are higher in the most populous provinces of South Africa and at district level the metropolitan district municipalities continue to lead with the highest number of birth registrations. This provincial and district distribution of births is reflective of the population size in the respective area of aggregation. The birth registration completeness particularly at district level has not been explored. This study will investigate completeness at subnational level and establish if there are any trends as well as spatial relationships between the areas of aggregation.

1.2 Problem Statement

Birth registration is a challenge in sub-Saharan African countries where vital registration is either non-existent or largely inadequate. This has given rise to the problem of accurately calculating the total fertility rate and birth registration completeness. In this study indirect methods will be used with the available birth registration data, to derive estimates of fertility and completeness at provincial and district level. Spatial distribution trends for fertility and birth registration completeness will be examined between the respective subnational areas of aggregation.

1.3 Aim and Objectives

The aim of this study is to assess the accuracy, validity and reliability with a view to ascertaining the usefulness of birth registration data at subnational level in South Africa, for the purposes of completeness of births and estimating Total Fertility Rate (TFRs). This study has five objectives. The first objective is to establish the completeness of birth registration data, that is, undercounts over time, using indirect methods. The second objective is to discover if there are any trends in the birth registration data such as seasonality of births and sex ratios. The third objective is to estimate the Age Specific Fertility Rates (ASFRs) and Total Fertility Rates (TFRs) for various subnational areas. The fourth objective will establish, with the use of GIS techniques, if there are any spatial variations of completeness and fertility at a subnational level. The fifth objective of the study seeks to find any problems encountered when dealing with subnational birth registration data.

1.4 Research Questions

To meet the aim and objectives outlined above, this research sets out to answer the following questions. The first is: By how much are the births undercounted at subnational level? The second and third questions are: Do births follow any seasonal pattern that can be used to project estimates of births and are sex ratios at subnational level consistent with national level? They will address the objective to establish trends in the birth registration data. The fourth research question will focus on measuring the ASFRs and TFRs: What are the ASFRs and TFRs for the available data? The fifth question is: What are the spatial characteristics of the birth registration data? The sixth research question seeks to establish the drawbacks of working with subnational birth registration data and the question is: What are the problems encountered when dealing with birth registration data at subnational level?

1.5 Justification for the study

The Sustainable Development Goal (SDG) 16.9 aims to achieve effective birth registration, which is important for legal identification, by 2030. In most societies this gives a child rights such as healthcare, education facilities, and as they progress into adulthood, rights to employment, property ownership and financial services (Danel and Bortman 2008; Mikkelsen, Lopez and Phillips 2015). Vital statistics form a crucial component of the health system and they assist policymakers in planning, decision making and distribution of resources (National Research Council 2009). Birth data by local areas, in particular, have not received enough attention in attempting to establish completeness and fertility indicators. Previous studies have not covered the estimation of birth registration completeness and total fertility rate at district level. Hence the study will shed light on the insights that can be drawn from live births registration data at subnational levels in South Africa.

Furthermore, the study seeks to address the lack of spatial analysis in the study of the distribution of demographic phenomena for example, the geographical correlation of district areas based on their birth completeness level. There has been a recent interest in the use of Geographical Information Systems techniques to investigate demographic questions. This study will contribute to this growing interest of spatial analysis within the context of birth registration data in South Africa

2. LITERATURE REVIEW

2.1 Overview of birth completeness in the world

Birth registration is a basic human right that affords a child an identity and benefits that come with having an official nationality; such as, healthcare, education and protection (UNICEF 2002). The Sustainable Development Goal (SDG) 16.9 targets the provision of a legal identity for everyone, as well as free birth registrations by 2030 (UN DESA 2016). Furthermore, birth registration data is crucial for the purpose of tracking fertility levels, monitoring family planning interventions and population projection (Andreev, Kantorová and Bongaarts 2013; Yaya, Data and Lindtjørn 2015). Well-maintained birth registration systems help in the measurement and achievement of the Sustainable Development Goals (Setel, AbouZahr, Karpati *et al.* 2018).

Birth registration is complete in some parts of the world such as Western Europe, North America, Australia, New Zealand and Japan (Phillips, Adair and Lopez 2018). It is almost complete in a number of Asian and Latin American countries, however, birth registration is still considerably less complete in many low- and middle-income countries (Garenne, Collinson, Kabudula *et al.* 2016; Phillips, Adair and Lopez 2018). Figure 2-1 shows the global distribution of birth registration coverage with, notably Sub-Saharan Africa, some parts of South America and those of Asia, displaying less complete live births records. The global northern countries and notably New Zealand and Australia show better (more than 90%) birth registration. Although some Asian and Latin American countries have a coverage of more than 90%, there are some which fall into the less than 90% group and others as low as between 40% and 60%. A greater number of Sub-Saharan African countries have less than 90% coverage where more than 35% of countries have less than 50% coverage and more than 60% of the countries have less than 75% coverage (UNSD 2017).

The State of the World's Children (UNICEF 2019) reports that the birth registration percentage for year 2018 for Western Europe, North America is 100%, for Eastern Europe and Central Asia 99%. There is a sharp difference in the level of birth registration in the global south with the Eastern and Southern African countries reported to have 40% and Western and Central African countries reported to have 53% registrations. The whole of Sub-Saharan Africa has an average birth registration completeness of 46%. However, South Africa is reported to have a relatively higher national completeness level of 89% (UNICEF 2019). The birth registration percentages are based on the number of children under the age of five who were registered at the time of the survey, with the numerator being the children who were reported to have a birth certificate. The main data sources used were data available from Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), censuses and vital registration systems.

Figure 2-1 Live births registration coverage. Source: UNDESA Statistics Division (2017)

Phillips, Adair and Lopez (2018) examined data from 145 countries to assess the availability, quality, and usefulness of birth registration data for the period 1948 to 2015, using data from the annual estimates of total births that were derived by the United Nations Population Division. These are divided into the observed number of births reported for each country in order to give the completeness of the national birth registration. It was noted that there are gaps in the birth data and only 8.6% of births are in countries with effective registration systems (Phillips, Adair and Lopez 2018). This leaves the rest of the countries with poor and incomplete registration systems and hence, the need to use indirect methods to estimate vital events. The study was the first (to the authors' knowledge) worldwide assessment of the completeness of birth registration data. Moreover, it was noted that unlike death registrations, the sharing, assessment and monitoring of birth registration data has not been well explored

(Phillips, Adair and Lopez 2018). Nannan, Dorrington and Bradshaw (2019) note that there has been a failure to maintain vital registration systems in developing countries to record births and deaths accurately and provide key health indicators for children.

2.2 Birth registration history and legislation in South Africa

The registration of births and deaths is regulated by the Births and Deaths Registration Act of 1992 (Act No 51 of 1992)[1]. The Act makes it mandatory for all citizens and permanent residents to register all births and deaths. Although there have been several amendments to the Act, the most notable was the 1998 amendment that introduced new birth and death notification forms. This improved the timeous registration of births since, during antenatal visits, women were provided with birth registration information and they were also encouraged to apply for their own identity documents in cases where they did not have them (Bradshaw, Kielkowski and Sitas 1998). The government's determination and will, together with the launch of Child Support Grants in 1998, improved birth registration since birth certificates are a pre-requisite for claiming the grant (Abrahams, Berry and Hendricks 2006; Garenne, Collinson, Kabudula *et al.* 2016).

2.3 Birth completeness in South Africa

Few studies exist which explore subnational birth registration data in South Africa. Nannan, Dorrington and Bradshaw (2019) sought to assess the completeness of birth registration for South Africa's nine provinces using South Africa's civil registration and vital statistics system from 1996 to 2011. The authors noted that earlier estimates of completeness had been derived by dividing the number of birth occurrences in a particular year by registrations in that same year. This assessed the timing of the registration instead of the completeness of birth registration (Nannan, Dorrington and Bradshaw 2019). As a result, a measure of completeness which relies on independent estimates of the true number of births was used in their study. These independent estimates were derived using the reverse survival indirect method with data from the 2011 Census South African citizens population. South African citizens' population was considered because birth registration data is recorded for South African citizens only.

Based on data from Statistics South Africa, UNICEF reports that in 2011 the country had a birth registration completeness of 95% within the same year of birth (Cappa and Wardlaw 2013). This was derived by dividing birth occurrences in 2012 by

---

[1] Government Gazette No. 13953: https://www.gov.za/documents/births-and-deaths-registration-act

birth registrations in 2012, instead of dividing by robust estimates of true births. The national household survey conducted in 2011 shows that 11% of the births of children aged three or under were not registered (Berry, Dawes and Biersteker 2013). In its 2011/2012 Annual Report, the Department of Home Affairs reported that birth registration had increased to 90% (DHA 2012). The DHA compares the births that occurred and were recorded in a particular year within 30 days after birth to the total number of births registered, including the late registered births. Nannan, Dorrington and Bradshaw (2019) note that the aforementioned proportions are a measure of timing instead of completeness since they do not use true births (as the denominator for completeness). Hence, the World Health Organization recommends a standard method for calculating birth registration completeness. The method uses "an independent estimate of the total number of births" and completeness is derived by dividing the actual registered births by the total estimated number of births for the corresponding period and express that as a percentage (WHO 2010). These birth estimates could be based on the available national estimates or derived from demographic estimation techniques.

Although there is still room for improvement, South Africa's civil registration and vital statistics system is well-established and has improved substantially since 1996 (AbouZahr, De Savigny, Mikkelsen *et al.* 2015; Bah 2009; Nannan, Dorrington and Bradshaw 2019). Nannan, Dorrington and Bradshaw (2019) used three methods to estimate the true number of births for the period 1996 to 2011. The first method used was the reverse survival method applied to the 2011 population census. For children in the 0-15 age group that were enumerated in the 2011 census, the number of births for each age cohort was determined by dividing a specific age cohort population by the suitable survival probabilities. The survival probabilities were obtained from the Actuarial Society of South Africa's population projection model of 2008 assuming no migration. The method estimates fertility without the need for asking direct questions about fertility. Assuming a population closed to migration, the number of people of any age x are the survivors of the births of those people in the previous x completed years. Hence, the number of births which occurred in the previous x years can be derived as long as the estimates of survival probabilities from birth to age x are available and the population counts are accurate. Moreover, total fertility can be estimated from the births and women according to age derived from the reverse survival method (Timæus and Moultrie 2013).

The second method to estimate the number of births in a particular year was the use of Total Fertility Rates (TFRs) based on previous research from 1996 to 2012. Age Specific Fertility Rates (ASFRs) and TFRs were derived from birth record data in the censuses of 1996, 2001 and 2011, the 2007 Community Survey and the 1998 Demographic and Health Survey. The ASFRs were then multiplied by annual estimates of the population of women of childbearing age, in order to get estimates of the number of births by age of mother. The resulting births were compared to those from the reverse survival method. An assumption was made that ASFRs change linearly from one data point to another (Nannan, Dorrington and Bradshaw 2019).

Third, births were estimated by using data from the District Health Information System in the public sector from 2004 to 2012. Pursuant to equitable and favourable health policies by the government post-1994, there was a notable surge of births in the public sector (Gilson and McIntyre 2007). Apart from that, Nannan, Dorrington and Bradshaw (2019) used data from the private sector or from home births to adjust for the data in the public sector. It was noted that private sector deliveries were more common in richer provinces such as the Western Cape and Gauteng. Poorer provinces, whereas Limpopo, Eastern Cape and Mpumalanga reported more home births (Nannan, Dorrington and Bradshaw 2019).

The Nannan, Dorrington and Bradshaw (2019) research focused on estimating the number of births nationally and at a provincial level. These estimates were then compared with the CRVS to estimate the completeness of birth registration. While the study used different approaches to derive correct estimates of births, it lays a base upon which provincial completeness of births can be compared. The current study uses the relational Gompertz fertility model and it goes a step further to smaller areas of aggregation and establishes whether meaningful conclusions can be made about the completeness at district level using the live births records data.

Rural areas have shown low levels of vital registration completeness (Garenne, Collinson, Kabudula *et al.* 2016; Jewkes and Wood 1998). For example, Agincourt, a largely rural area in Mpumalanga, has been under a health and socio-demographic surveillance program which started in 1992. Fieldworkers for the Agincourt Health and Demographic Surveillance System (HDSS) visit every household and collect data regularly once a year. Data is collected on births and deaths registration as well as in and out migration for the area. Moreover, related socioeconomic and demographic data are collected at individual and household levels. An analysis was done to investigate the factors and trends in births and deaths registration completeness from 1992 to 2014. There was an increase of births registration from 6.2%[2] in 1992-1993 to 89.1% in 2013-2014 and the study revealed that completeness was higher in households with higher levels of education as well as in wealthier communities of Agincourt (Garenne, Collinson, Kabudula et al. 2016).

2.4 Problems in measurement

Admittedly the process of estimating the births and deaths registration completeness is tough, particularly for South Africa. One of the reasons is that in the case of birth completeness, the denominator which is the true number of births is controversial. It is noted that the different estimates of true births that are derived from sources such as censuses, surveys or through application of models, differ by 10% or more (Garenne, Collinson, Kabudula *et al.* 2016).

The other reason is the confusion brought up by the numerator due to late or delayed registration of births. However, this problem was dealt with in the Nannan, Dorrington and Bradshaw (2019) study by referring to different ages of children by which the births are registered, for instance, births before age one or before age two. Moreover, for Agincourt, the problem of the denominator and enumerator does not apply since they have been recorded accurately from the year 1992. The undercounts for births and deaths in Agincourt were, according to the authors, "negligible" for the study period of 1992-2014 (Garenne, Collinson, Kabudula *et al.* 2016).

Since the Agincourt study focused on one area which is under a demographic surveillance program it managed to address accuracy problems associated with huge undertakings like a census since data is collected yearly for the HDSS. Higher levels of registration may also have been created as a result of interventions which may have encouraged people to register. Furthermore, the study forms a base upon which to compare similar areas of concern, i.e. other rural areas in the country. However, the current study seeks to unfold the irregularities of census and community survey birth registration data in estimating completeness.

---

[2] The authors give inconsistent completeness levels between the research's abstract (7.8% in 1992) and main text (6.2 in the first two years 1992-1993; implying that the year 1993 alone was 4.8%). However, an inference of Fig. 1 shows that completeness was 7.8% in 1992 and 6.2 in 1993.

2.5 Birth estimation methods

Despite compromised data quality in developing countries, indirect methods or statistical approaches can be used on the available data sources to produce robust birth estimates. This section will outline some of the methods that were used under settings where data was poor or incomplete.

Queiroz and Lima (2016) cite the same two factors that cause poor vital registration as noted by Jaspers (1994): environmental and institutional factors. Environmental factors result from lack of motivation from the public to report vital events; institutional factors come from the difficulty in accessing registration offices and the associated costs. The study focuses on Brazil as a case study to assess the completeness of birth registration. It highlighted that indirect methods for estimating fertility have been used by demographers in the Latin American region for the previous five decades. Census data from 1970 to 2010 was used for the Rio Grande do Norte (RN) State, a region with poor quality vital registration data and which underwent quick changes in fertility and mortality (Lima and Queiroz 2014; Queiroz and Lima 2016). The number of children ever born, births during the year before each census, and the number of women were all classified into five-year age groups.

The national analysis used birth registers which were made publicly accessible by the Ministry of Health as well as estimates from the National Statistical Office (The Brazilian Institute of Geography and Statistics) and Demographic Health Survey. Three indirect methods for estimating fertility were used: The P/F ratio, Synthetic Relational Gompertz model and the Own-Children method. The results obtained from these methods were compared to assess birth registration completeness as fertility decreased beyond replacement level. It was ascertained that the Synthetic Relational Gompertz model produced consistent fertility estimates irrespective of the data source. This showed the method's robustness. The Own-Children method produced slightly over-estimated fertility rates in the 25-29 age group compared with the Synthetic Relational Gompertz model. The P/F ratio and Own-Children methods produced comparatively similar results with small differences in the older ages of reproduction.

The study advocates for investing and promoting the accurate recording of information about mothers and investing in civil registration education of both fathers and mothers (Queiroz and Lima 2016). Apart from that it was concluded that the P/F ratio and Synthetic Relational Gompertz model were ideal in assessing births completeness particularly under circumstances of rapid fertility changes. However, the study focuses more on producing fertility estimates and comparisons of various methods and data sources. Nonetheless, the study gives assurance to the idea that indirect methods can produce reliable fertility estimates in circumstances where the quality of data is poor.

Using linear regression, P/F ratio and the Preston Integrated method to estimate fertility, Chen (2016) evaluated the completeness of birth registration in China. The data used was derived from census, primary school enrolment records, and household registration systems. It was noted that the fertility estimates from the different methods were largely consistent across the data sources. Nevertheless, the study highlights the negative impact of the one-child policy on the quality and completeness of birth registration. Chen (2016) highlights the two opposite opinions about under-reporting of birth data. One school of thought is that fertility drop since the late 1990s showed a true picture since the census and survey data upon which it was based is accurate, while the other school of thought argue that fertility has not been that low since there was underreporting, particularly of girls. There was a TFR of 1.23 and 1.18 for the 2000 and 2010 population censuses, respectively. Hence, the need for the study to use different sources of data and methods to evaluate birth reporting in China. The study concludes that in spite of highly consistent birth estimation results, the Preston Integrated approach is more suited for China (Chen 2016). The approach produces more robust fertility estimates which are, in addition, insensitive to relative changes in the completeness of coverage of two consecutive census age distributions. Unlike the corresponding changes in adult mortality, birth rate estimates respond negligibly to the substantial changes in census coverage (Preston 1983). This would be appropriate for China with censuses that have under-counts at young ages as well as to sub-national regions which have differences in under-reporting and migration (Chen 2016).

The study sought and managed to use three methods to estimate fertility from three sources of data. This was successfully done to show the reliability of the methods and how the different sources of data still give consistent estimates. However, the study also focuses on the whole country of China, although the author subsequently suggests the same study could be done at sub-national level especially with the suitability of the Preston Integrated method at that level. Hence, the Preston Integrated method could be used for another study in South Africa and compared with the results of methods which have already been used.

2.6 GIS in demography/Spatial distribution of birth registration

Demography maps have been used to show the spatial relationship of population features in given scenarios. Maps give a visual sense to demonstrate where a particular phenomenon tends to be more predominant in some areas than others. Moreover, maps help to illustrate the concentration of high and low values of the studied phenomenon. This leads to the identification of any socio-economic or political reasons or factors that are responsible for the spatial patterns (Oliveau and Guilmoto 2005).

Oliveau and Guilmoto (2005) argue that demographers have the proficiency to analyse the strength of statistical connections between phenomena, for example between fertility and education of mothers. However, demographers are less inquisitive about spatial relationships which can be shown when demographic phenomena are fully mapped, for example the geographical correlation of district areas based on their fertility level. One of the tools that is used to analyse the degree of spatial variations of demographic data is Moran's Index, an inferential statistic that describes spatial autocorrelation. Spatial autocorrelation is the degree of similarity between observations based on their location (Hijmans 2020). The Spatial Autocorrelation (Global Moran's I) is a GIS tool that measures spatial autocorrelation based on both the location of a geographic feature and its value at the same time. For instance, given a set of districts and a related attribute[3] or characteristic like birth registration completeness, the Moran's Index assesses whether the displayed pattern is random, dispersed or clustered. The Moran's Index ranges from -1 to +1. The extreme of +1 signifies that similar values are clustered together, 0 means that there is a random pattern and the other extreme of -1 indicates that high values are near low values.

Marti-Henneberg, Franch-Auladell and Solanas-Jiménez (2016) explain how GIS tools particularly the Global Moran's Index are used to examine population density over a period of time based on municipal boundaries. Another study, by Salvacion, Magcale-Macandog and Cultures (2015), uses the Local and Global Moran's Indexes to assess the spatial distribution of population growth in the Philippines. The Local and Global Moran's Indexes are similar in that they both provide a measure of how similar geographical features are to their neighbours. The Global Moran's *I* compares how similar every feature (in this case a district) is to its neighbours, and then averages out all the comparisons to give an overall view about the spatial pattern of the variable. The

---

[3] In GIS an attribute is information about a geographic feature that is nonspatial for example its name, area, TFR and completeness. This is stored in a table linked to the feature. In our case the district is the geographic feature

Local Moran's *I* singles out the features that are similar or different to the features in their neighbourhood and compares individually without averaging out the comparisons of all features in the neighbourhood. Hence, each feature would obtain its own *I* value, and also its own variance, z value, expected *I*, and variance of *I*. The study by Salvacion, Magcale-Macandog and Cultures (2015) shows the importance of GIS in variations of population distribution over time and space. Demographers have in recent years developed an interest in the application of spatial analysis methods (Wachter 2005). Demography is often spatial in nature since it deals with population dynamics in specified geographical areas (Hachadoorian, Gaffin and Engelman 2011; Matthews and Parker 2013; Weeks 2004). Matthews and Parker (2013) argue that many questions in the field of demography can be investigated from a spatial perspective. This becomes perceptible as the increase in geocoded demographic data is coupled with new technologies and analysis methods (Matthews and Parker 2013).

In South Africa, a few studies have been done in the area of spatial demography, notably an inquiry into the spatial associations and distribution of infant mortality and the implication of policies (Sartorius, Sartorius, Chirwa et al. 2011). The Sartorius, Sartorius, Chirwa et al. (2011) study aimed to spatially analyse infant mortality for South Africa at sub-district level and particularly to establish the high risk areas of infant mortality. The current study will apply the Global Moran's Index GIS tool to assess the spatial relationships of district with respect to their birth registration completeness and TFR levels. The study seeks to establish if, based on completeness, there is any pattern between districts of similar values, and the same analysis will be done on district TFR values. Then, overall the study will enquire if completeness and TFR are spatially interdependent, that is , whether their separate spatial patterns affect each other.

The chapter gave an overview of birth completeness in the world and highlighted how the global south is still lagging behind. Moreover, it was noted that as compared to other Sub-Saharan countries, South Africa has higher birth completeness level. A description of how the birth registration history and the associated legislation evolved over the years in South Africa was given. It was noted that birth completeness levels have increased over time. Although there are problems in the measurement of birth completeness, the relational Gompertz model will be used since it produces reliable fertility estimates despite the fact that the source data could to some extent be incomplete and unreliable. A method by Hauer and Schmertmann (2020) that estimates fertility will be used to validate the results of the relational Gompertz model method.

The method of Hauer and Schmertmann (2020) will be described in detail in the following chapter. In addition, it should be noted that the use of GIS for spatial analysis is still in its infancy in demography. The current study will apply the GIS tool of spatial autocorrelation since there has been a gap in the use of GIS to investigate demographic measures on a subnational level.

3. DATA AND METHODS

This chapter will describe the data sources and methods that are used to produce fertility estimates which will be used for calculating completeness. Three main data sources are used for this study: the vital registration births records, the 2011 Census, and the 2016 Community Survey. For the years between 2011 and 2016, the completeness analysis will be done using midyear population estimates and population projections.

3.1 Recorded Live Births

When a baby is born in South Africa, either parent or a legal guardian gives a notice of birth through the completion of a DHA-24 form for births within 30 days or, DHA-24/LRB for those reported after 30 days. As a result of the completion of the Notice of Birth form, a birth certificate is issued by the Department of Home Affairs (DHA). Stats SA then acquires the birth registration data captured by the DHA for processing. Notably the recorded live births do not include births from parents that are not South African citizens or non-permanent residents.

Stats SA has published on their website two sets of births records: the collation of statistical releases from 1998 to 2017 and a standalone release of 2018 birth records. For easy handling, the two datasets were combined to form one dataset using the “append” Stata command.

The live births data contains 10 variables:

- The year of the birth registration
- The year of birth occurrence
- The month of birth occurrence
- Sex of the child
- Age of the father
- Age of the mother
- Birth registration status
- Magisterial district of birth occurrence
- District municipality of birth occurrence
- Province of birth occurrence

3.2 Birth registration completeness

The study seeks to establish the completeness of the recorded live births from vital registration. This is achieved through comparing the recorded live births (from vital registration) and the births that actually occurred (estimates from the relational Gompertz model). The completeness is calculated as follows:

$$\frac{LB}{EB} \times 100$$

[3.1]

Where:

LB are the live births records which are derived from the Stats SA's Recorded Live Births and,

EB are the actual births that occurred which are estimated from the relational Gompertz Model.

The completeness is calculated for each area of disaggregation, that is, at national, provincial and district level.

3.3 The relational Gompertz model

The relational Gompertz Model uses fertility estimates that come from the births of recent months (e.g. 12 months) before a census or survey to determine the shape of fertility distribution and the lifetime average parities to establish the fertility level (Moultrie 2013). The goal is to find more robust estimates of age-specific fertility rates and total fertility rates by establishing the shape of the fertility schedule from recent birth data while establishing the level of the fertility schedule from average parities of younger mothers. Older women tend to forget to mention any children who might have died or have left home which leads to a low rate of children ever born while younger women tend to exaggerate their age which leads to inflated fertility rates for older women. The relational Gompertz model corrects such commonly inherent errors in fertility data including too few or too many births reported in a reference period, underreported lifetime fertility, and age misreporting among older women.

The relational Gompertz Model uses the sigmoidal Gompertz distribution, namely:

$$G(x) = \exp\left(a.\exp(bx)\right)$$

[3.2]

A double-negative logarithmic transformation is applied to the function G(x) to yield a relationship known as a 'gompit'. This transformation is the basis for the relationship between the cumulated fertility and the average parities:

$$Y(x) = -\ln\left(-\ln\left(G(x)\right)\right)$$

[3.3]

Brass (1978) noted that there is an approximately linear fit between the gompits of the observed fertility data and the gompits of a defined standard fertility schedule. This is expressed simply as:

$$Y(x) = \alpha + \beta Y_s(x)$$

[3.4]

Where $Y_s(x)$ is the gompit of the standard cumulated fertilities as derived and presented by Booth (1984).

If $\alpha=0$ and $\beta=1$, then the shape of the fertility schedule is the same as the standard fertility schedule. Alpha (α) is the extent to which the mean child-bearing age in the population is different from the standard. Negative values of alpha indicate that the mean childbearing age is greater than the standard. Beta (β) shows the extent to which fertility distribution is spread (if beta is greater than 1 it shows a narrower distribution than the standard).

The relational Gompertz model has two weaknesses. The first one is that the total fertility is needed as an input while it is also the result that the analyst is attempting to derive. This leads to circularity. Secondly the model assumes that fertility is constant overtime. These weaknesses were dealt with by Zaba (1981), who reformulated the method to avoid circularity and also eliminating the assumption of constant fertility.

In this study the relational Gompertz Model will be applied to the recent fertility reported in the 2011 Census and 2016 Community Survey. The model rectifies the irregularities in fertility data, by evaluating the parity and recent fertility data while providing the actual correct fertility level. Assuming that the model produces correct estimates of births for the particular years, the correct birth estimates are then compared with the vital registration births data to determine completeness (Equation [3.1]). The model is used to correct the births for the 12 months before the 2011 Census (1

October 2010 to 30 September 2011) and the 12 months before the 2016 Community Survey (1 March 2015 to 29 February 2016).

3.4 Derivation of birth data from the 2011 Census and the 2016 Community Survey

The census and community surveys provide recent and lifetime fertility data. The births data is used to calculate the age-specific fertility rates (ASFRs) and parities that are needed as an input into the relational Gompertz model to estimate the correct age-specific and total fertility rates. The data from the 2011 Census and 2016 Community Survey data were subjected to editing and correction before being published (Section 3.9).

3.5 Age Specific Fertility Rates (ASFRs)

The recent fertility data is derived from the 12 calendar months before the 2011 census. Since the census was done on 9 and 10 October 2011, the recent births were considered from 1 October 2010 up until 30 September 2011. Similarly, the Community Survey was conducted over a six-week period (7 March to 22 April 2016) hence, the 12 calendar months before the survey are from 1 March 2015 up until 29 February 2016. The last-child month and year of birth variables are used to determine the births that occurred within the aforementioned 12-month time frames.

Furthermore, the births data extracted from the 2011 Census and 2016 Community Survey is only for South African citizens to enable comparison with the vital registration births data. The population counts are either from the 10 per cent sample of the census or from the community survey and are adjusted for the undercount using a weighting variable. The population is for female South African citizens of childbearing age presented in five-year age groups and is used as the denominator for calculating the ASFRs.

3.6 Parities

The parity data was derived from the total children ever born variable. The data shows that it is already edited since there are no implausible parities. For the 2011 Census, parity data correction was not possible since the unknown parities were negative numbers. The value of beta is negative (e.g. -0.008 at national level). The 2016 census survey parity data does not need an el-Badry correction since the proportion of women in each age group with unknown parity (Ui) is less than two per cent (Table 3-1):

Table 3-1 Proportion of women with unknown parity, 2016 CS

Age	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Ui	0.22%	0.22%	0.25%	0.36%	0.37%	0.46%	0.47%	0.62%

3.7 Mid-Year Estimates (MYE)

The Stats SA mid-year population estimates were used to calculate estimates of births for the years between the 2011 census and 2016 community survey (i.e. 2012, 2013, 2014 and 2015). The MYE are disaggregated down to district municipality level. Assuming that age-specific fertility rates change linearly between two data points (Nannan, Dorrington and Bradshaw 2019), that is, between 2011 and 2016, the ASFRs were multiplied by the MYE to give estimates of births. The MYE are as of June while the ASFRs applied are six months before the census or community survey. This results in a misalignment of two months. However, given that the ASFRs are inadequately accurate and for practical purposes, the rates were not interpolated by 1/6 of a year. Consequently, the births derived were compared with the vital registration births to assess the completeness.

3.8 Hierarchies of disaggregation

Stats SA recodes the DHA birth data into three levels of geographical disaggregation based on the 2016 municipal boundaries. There are 226 local municipalities, 52 district municipalities and nine provinces. This study is based on the district municipal and provincial levels. A variable of unique district codes was created in Stata to identify the 52 district municipalities. The nine provinces already had unique codes assigned to them. The methods used and results obtained thereof can be similarly applied and extended to the local municipal level in further studies, although data at that level of disaggregation becomes less reliable (Statistics South Africa 2012a)

The 2011 and 2016 years were chosen since their geographical data is comparable because of their matching boundaries. There have not been significant changes between the 2011 and 2016 boundaries. The preceding census of 2001 and community survey of 2007 have been subjected to significant boundary changes when compared to the 2016 municipal demarcation (Statistics South Africa 2012b). This makes it difficult to compare the same areas over different periods of time.

Moreover, to make comparisons easier, the alignment process was done in order to match the boundaries of the censuses and community surveys . On one hand, both the 2016 Community Survey and 2011 Census are compatible for comparison purposes since the 2016 Community Survey was done based on the 2011 census municipal boundaries. On the other hand, the 2007 Community Survey has been recoded so that it becomes comparable with the 2001 census. As such, the 2007 Community Survey is presented in both the 2001 and 2005 boundaries. However, there is a difficulty in comparing the 2007 Community Survey and 2001 Census to the current boundaries of the 2016 Community Survey and 2011 census. Hence, a cautionary note by Statistics South Africa (2012b) to align the boundaries.

3.9 Data editing

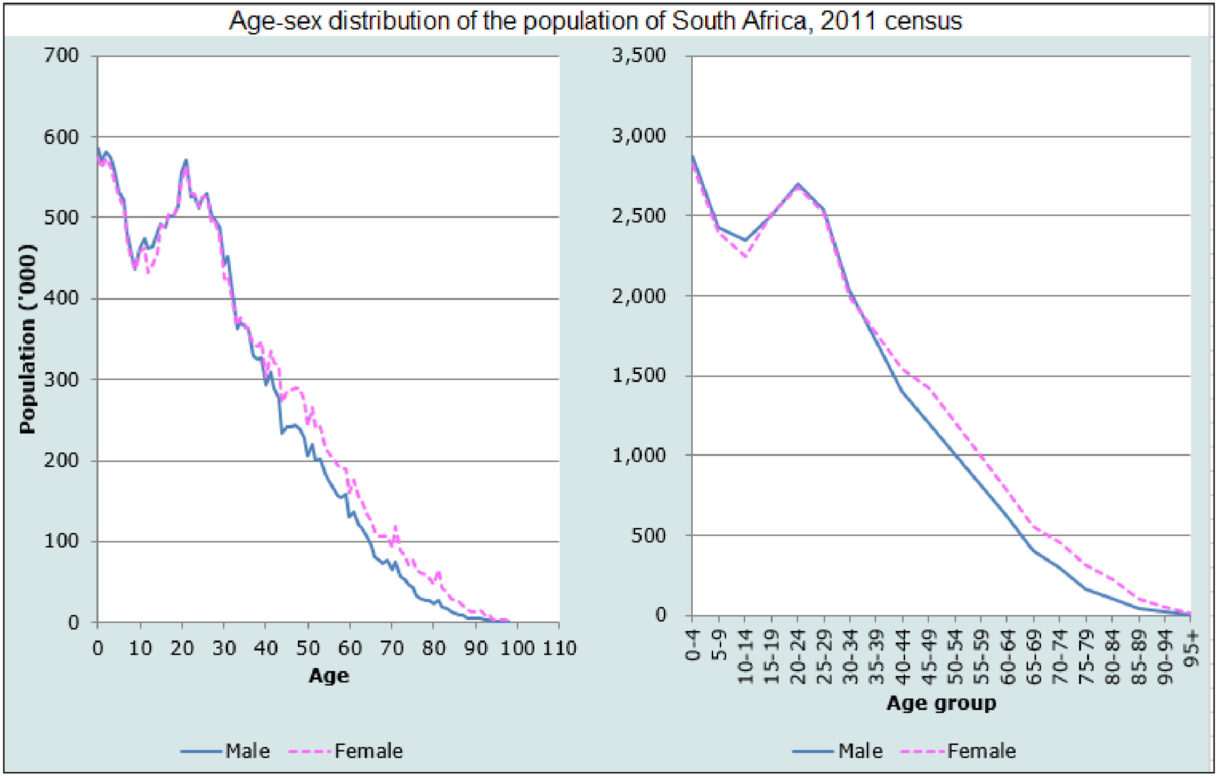
Data quality assessment for the 2016 Community Survey was mainly automated and done in two stages by Stats SA. In the first stage the electronic questionnaires underwent a set of conditions and validation processes. This removed unnecessary irregularities found during data collection, for example unrealistic data entries. If a questionnaire did not meet the minimum conditions it would be labelled REJECTED, otherwise, it was labelled ACCEPTED. The rejected questionnaires were sent back to the fieldworkers to correct the mistakes after which they would be declared “Complete”. Consequently, the process led to a remarkable reduction of missing values.

In the 2011 Census, data was corrected through logical imputation and dynamic or hot-decking. Logical imputation is a data editing process where missing values are derived from related information in the household. Hot-decking is replacing missing values with data from the nearest neighbour.

3.10 Evaluating data quality- age and sex distribution

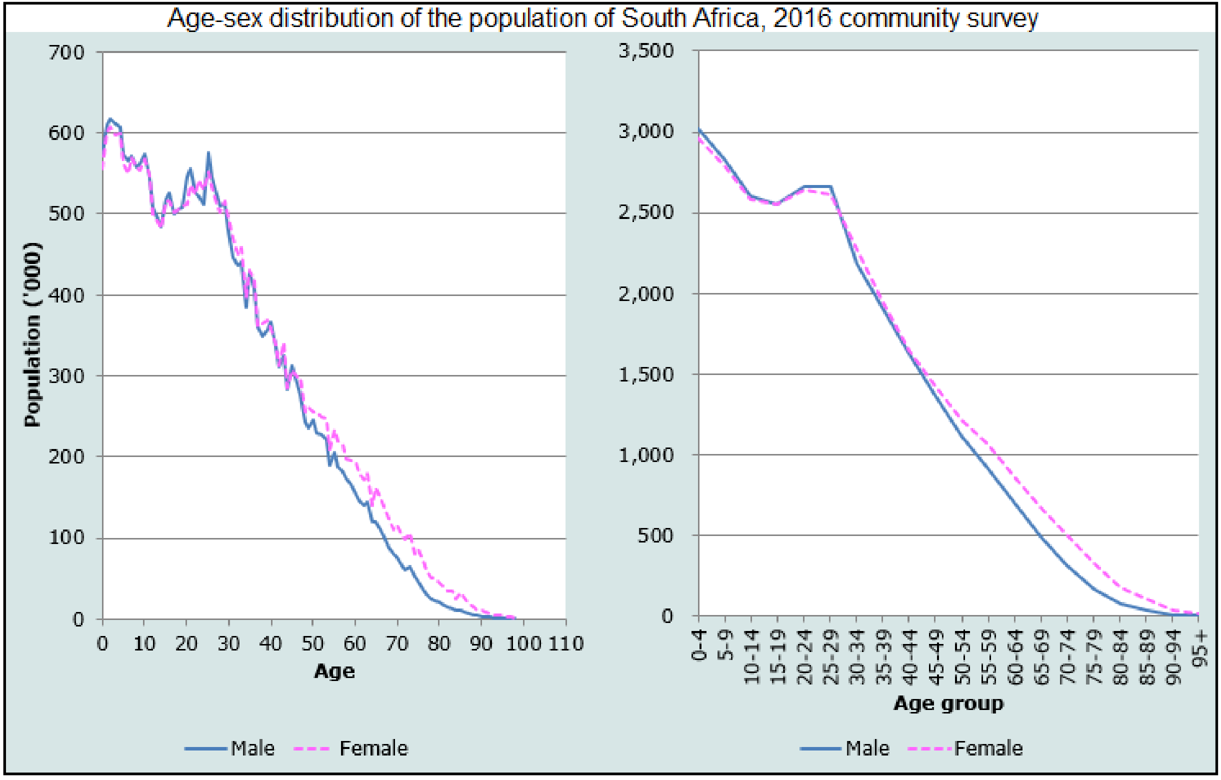
In this section the data is evaluated using procedures that check for consistency internally or externally. Internal checks show whether the age and sex distribution of the population complies with general standards. External checks are done to check consistency with historical data of the same area. If there are major differences from known demographic patterns of births, deaths and migration, without any interference of significant disasters, then the data may have some errors.

Figure 3-1 Age and sex distribution of South Africa, 2011 Census



The left-hand graph (Figure 3-1) shows a general matching trend between male and female population. At age twelve there is a trough for females suggesting that there may have been under-counting of girls at that age. There are also concerns as to the dip in children, which may have been caused by omission of the children who were starting to attend school. After age 40 we have more females than males. This conforms to the general idea that females have a higher life expectancy than men. Typical of a developing country, the younger ages in the 0-5 age group are more than the older ones, reflecting the mortality pattern. Moreover, there are subtle hints of digit preference for ages ending with 1 and 6. The right-hand side shows a smoothed graph in five-year age groups. Similar patterns are observed for females at age twelve and after 40 years. The 2016 population (Figure 3-2) shows a general pattern similar to the 2011 population. However, there are more defined spikes in the 2016 population suggesting age heaping around ages which end with 0 and 5.

Figure 3-2 Age and sex distribution of South Africa, 2016 Community Survey



3.11 Sex ratios

In the 2011 Census the national sex ratio at age zero is 101.94 and over the twelve months prior to the census a sex ratio of 101.53 of live births was recorded. The 2016 community survey sex ratio for age zero is 102.76 while in the prior year the recorded live births is 101.59. This is consistent with a typical sex ratio at birth which should be higher than 100 showing that more males are born than females. The national sex ratio at birth will be used in comparison with that of the different provinces. Then in turn the sex ratios at birth for the individual provinces will be compared with the districts which fall under them. This is done to check any major deviations from the overall sex ratio, and which will lead to an explanation and insights of the underlying data.

3.12 Seasonality of births

The total births for each calendar month were retrieved for all the years from 2011 to 2016. The years were selected since they are consistent with the time frame under study. The total births were obtained from Stats SA's statistical release of the 1998 to 2018 birth records. In order to establish a trend over time, a line graph was used to display each year's births with the horizontal axis showing calendar months and the vertical axis displaying the number of births. In addition, the monthly data was aggregated based on the seasons of South Africa, that is, summer (December to February), autumn (March to May), winter (June to August) and spring (September to November). Then, a line graph was used to show the variations between seasons for all six years from 2011 to 2016.

3.13 Spatial Analysis

The 52 districts of South Africa have different levels of completeness and fertility. This prompts the idea of spatial variation analysis to inquire into the reasons behind the spatial distribution of completeness and fertility across districts. Geographical Information Systems (GIS) concepts will be used to map the differences in the fertility indicators observed and to investigate the spatial correlation of the variables. GIS offers visual and quantitative methods that help to describe the spatial characteristics of the indicators. In a GIS environment data can be explored in a way that was previously not easily attainable (Oliveau and Guilmoto 2005).

Spatial autocorrelation shows the degree to which one object is similar to other nearby objects. Notably, the Moran's *I* (Index) measures spatial autocorrelation. This GIS tool assesses whether the pattern of, in this instance births, completeness is clustered, dispersed, or random. The null hypothesis states that the births completeness of districts are randomly distributed across South Africa. Two models of Spatial Relationships were applied and there was consistency in the results. The models are the Inverse Distance and the Contiguity Edges Only.

The Inverse Distance method renders more weight to the districts that are closest to the target district than to the districts that are further away. Hence, a spatial relationship is sought between districts that are within a certain range of the target district. In the Contiguity Edges Only method, only the neighbouring districts which share a boundary with the target district will influence the computations of the spatial relationship.

The z-scores and p-values were used to determine statistical significance towards clustering or dispersion and whether we accept the null hypothesis of random distribution of births completeness by district. The Moran's *I* index ranges between extremes of -1 and +1. The closer the value is to +1, the more it will show the clustering of similar values, and the closer it is to –1, more clustering of dissimilar values will be shown. This index helped analyse the spatial variation and dependency of the completeness and TFRs of districts of South Africa.

Moran's Index is calculated as follows:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{j=1}^{n} z^2_{\ i}}$$

[3.5]

Where:

z_i is the deviation of an attribute for geographic feature i from its mean: $\left(x_i - \bar{X}\right)$,

$w_{i,j}$ is the spatial weight[4] between geographic feature i and j,

n is the total number of geographical features, and

S_0 is aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}$$

[3.6]

The z_I -score for the statistic is calculated as follows:

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}}$$

[3.7]

where:

$$E[I] = -1/(n-1)$$

[3.8]

$$V[I] = E\left[I^2\right] - E[I]^2$$

[3.9]

---

[4] A spatial weight is a variable that quantifies the relative strength of a relationship between pairs of locations (Griffith, 2017).

The Global Moran's *I* statistic calculates the mean and variance for a given attribute under assessment, in this instance birth completeness. After that, for each geographic feature (a district in this case) it finds the difference between the attribute value and the mean, resulting in a deviation from the mean. A cross-product of the deviations from the mean of all the neighbouring geographic features/districts (for example districts within a specific distance range) is calculated. Notably, the summation of the cross-products are on the numerator of the Global Moran's *I* statistic. The variance normalises the numerator so that the Index stays within the -1 to +1 band. Consequently, the Moran's Index value is compared to the Expected Index value. Considering the number of geographic features/districts in a dataset and the variance for the data, a z-score and p-value are calculated to determine whether the difference between the observed and expected index values is statistically significant or not. The conclusion is reached within the context of a null hypothesis which states that there is a random distribution between the geographic features/districts based on birth completeness level (Getis and Ord 2010; Goodchild 1986; Mitchell 2005).

The shapefiles for spatial analysis were sourced from the Municipal Demarcation Board. For each of the three categories of birth registration lateness, the completeness and TFRs were shown on a geographical map with district boundaries. The different completeness and TFRs are classified into intervals and the intervals are assigned different shades of colour displayed on the map, known as a choropleth map. A choropleth map is a thematic map which is coloured in relation to a range of statistical data values presented in a key.

3.14 xTFR$^+$

After estimating the total fertility rates using births that are corrected by the relational Gompertz model for national, provincial and district level, for comparison and validation purposes the Hauer and Schmertmann (2020) population pyramids will be employed in the current study. This method does not rely on the conventional way of calculating TFR by using births presented by age of mother, which makes it suitable for areas with minimal data available. Even though other indirect methods were suggested (Bogue and Palmore 1964; Rele 1967), they still require variables such as the mean age at marriage and percentage of women ever married which may not be available in census or survey data. Hence, these indirect methods would only apply to areas or circumstances with detailed population data akin to the direct TFR method.

Hauer and Schmertmann (2020) demonstrate five TFR variants that circumvent the aforementioned challenges by only relying on age and sex population data from a census or a survey. The method takes advantage of the demographic link between population pyramids and TFR. In earlier research by Hauer, Baker and Brown (2013) that compares this method and the Bogue-Palmore method, it has been illustrated that the errors produced by the Hauer and Schmertmann (2020) method are less than those from other indirect methods. The five variant method is tested and evaluated for accuracy under many different demographic circumstances (Hauer and Schmertmann 2020). The method's accuracy was tested on 2403 fertility schedules from a period of 124 years.

The method is based on the demographic relationship between TFR and Age-Sex distributions Equation [3.10]:

$$TFR = n.\sum_{a=\alpha}^{\beta-n} F_a = n.\sum_{a=\alpha}^{\beta-n} \frac{B_a}{W_a}$$

[3.10]

Where:

[α, β) is the reproductive age range for women,

W_a is the midyear population of women in the age interval; [a, a + n),

B_a is the annual number of births to those women, and

F_a is the average fertility rate of the women.

Demographers frequently adopt (α, β, n) = (15, 50, 5) which gives seven age groups that have fertility rates F_a where $a = 15, 20, \ldots, 45$ and

$$TFR = 5.\sum F_a$$

[3.11]

Furthermore, in the relationship of TFR and the Child to Woman Ratio (CWR), there are three main demographic factors to be considered: Some of the children who are born in the previous *n* years will have died by the time the population is counted. Some of the women who gave birth over the past *n* years will die before they get counted; and the women who make it for enumeration in a specific *n*-year age group were in that age group only for a part of the past *n* years. Hence, the expected number of children under age 5 who survive per woman who survive in an age group a at the end of a five-year period as approximated by cohort-component projection methods; which are a slight variation of the standard Leslie matrix formulae (Schmertmann and Hauer 2019; Wachter 2014) is

$$C_a = TFR.s.p_a$$

[3.12]

and therefore, the expected total number of children under age 5 who survive is:

$$C = \sum_{a=15}^{45} W_a C_a = W.p.s.TFR$$

[3.13]

Where:

W_a is the enumerated women in age group a

W is the total number of women counted at childbearing ages 15 to 50, assuming that fertility is zero outside this age range

p_a is the fraction of lifetime fertility that females in age group a experienced over the past five years

$p = \sum W_a p_a / W$ is the population-weighted mean of p_a values, and

$s = L_0 / 5$ is the expected fraction of children born in the past five years who are still alive.

By rearranging the equation we get:

$$TFR = \frac{1}{s}.\frac{1}{p}.\frac{C}{W}$$

[3.14]

The TFR derived from this equation fundamentally assumes that the number of children under age five in an age pyramid can be a substitute of recent births to the women in the same age pyramid, after appropriately adjusting for child mortality.

The five variants of the method are *iTFR, xTFR,* $iTFR^{+}$*,* $xTFR^{+}$ and *bTFR*. The *iTFR* and *xTFR* variants only require age pyramid data as input, while the $iTFR^{+}$, $xTFR^{+}$ and *bTFR* variants adjust for child mortality (*q5*). Of the later three variants, the $iTFR^{+}$ does not use the age distribution detail for women (15–49) and the *bTFR* is a probabilistic approach that handles demographic quantities in Equation [3.12] and Equation [3.13] as uncertain. The current study will use the $xTFR^{+}$ (Section 3.14) variant because the method adjusts for both child mortality and the uneven distribution of women within reproductive ages, that is, women are more concentrated in the 25-34 age groups (Hauer and Schmertmann 2020).

When the Hauer and Schmertmann (2020) general equation for TFR (Equation [3.14]) takes into account child mortality and the different distribution of women in reproductive ages, it reduces to:

$$xTFR^{+} = \left(\frac{10.65 - 12.55\pi_{25-34}}{1 - 0.75q_5} \right) \cdot \frac{C}{W}$$

[3.15]

Where:

π_{25-34} is the proportion of women aged 25–34 among those who are aged 15–49,

$1 - 0.75q_5$ is an approximation of 1/s in Equation [3.14] established on empirical evidence from life tables.

The coefficients values 10.65 and –12.55 were derived based on *W* which includes women aged 15-50. Different coefficients would apply for different reproductive age ranges. The Statistics South Africa (2020a)'s District Council projection by sex and age (2002-2020) was used as a source for the 2011 and 2016 population at district level.

For the *q5* component of Equation [3.15] at district level, estimates from Zewdie (2014) of municipality level were to derive *q5* values to be used for the years 2011 and 2016. Zewdie (2014) derived U5MR estimates at municipal level for the year 2011. These estimates were used in the current study to deduce 2011 U5MR at areas of higher disaggregation, that is, district level using Equation [3.15] and adding together the results for all the municipalities in a district:

$$\frac{mp}{DP} \times mU5MR$$

[3.16]

Where:

mp is the 2011 population for a particular municipality,

DP is the 2011 population for the district that contains the municipality, and

$mU5MR$ is the U5MR for the municipality.

For the year 2016, it was assumed that the relative differentials have not changed. Hence, using the provincial U5MRs of 2011 (Zewdie 2014) and 2016 based on the Thembisa 4.2 model (Johnson, Dorrington and Moolla 2017), the district U5MR for the year 2016 were calculated as (Equation [3.17]):

$$\frac{dU5MR_{2011}}{pU5MR_{2011}} \times pU5MR_{2016}$$

[3.17]

Where:

$dU5MR_{2011}$ is the U5MR for a district in 2011 (derived from Equation [3.16]),

$pU5MR_{2011}$ is the 2011 U5MR for the province within which the district is found, and

$pU5MR_{2016}$ is the 2016 U5MR for the province within which the same district is found.

3.15 Software

Data querying and analysis was mostly done with Stata version 15.1 (StataCorp 2017) and further analyses of data obtained through Stata were carried out in Microsoft Excel. The district birth completeness and Total fertility Rates were spatially analysed in a GIS environment. All maps throughout this research were created using ArcGIS pro 2.5.0 software by Environmental Systems Research Institute (ESRI 2020).

4. RESULTS AND ANALYSIS

The objective of this chapter is to present and analyse the results derived from birth registration data over time. Changes in the completeness of birth registration are examined from 2011 to 2016. The chapter also explores the spatial variation and distribution of completeness levels as well as TFR at district level. Thereafter a comparison is done between TFRs from the relational Gompertz model and the Hauer and Schmertmann (2020) method.

4.1 Derivation of the 2011 and 2016 births data from Recorded Live Births

The age of the mother was converted into 5-year age groups through the creation of a new variable. The birth month and year were expressed in the Century-Month Code (CMC) date system format for simplicity of calculation (Croft, Marshall, Allen *et al.* 2018). Furthermore, unknown and unspecified births were proportionally apportioned across the age of the mother for a complete representation of all the recorded births. The unknown or unspecified births are less than 2% of the total births which is a reasonably acceptable error percentage (Table 4-1 and Table 4-2).

Table 4-1 Percentage of unknown/unspecified births for 12 months before 2011 Census within 1, 2 or 5 years of registration

Age	Within 1 year	Within 2 years	Within 5 years
15	113,042	136,514	150,882
20	248,462	268,047	277,394
25	241,246	255,653	261,886
30	167,283	177,555	182,021
35	98,495	105,467	108,372
40	29,528	32,075	33,161
45	3,147	3,474	3,757
Unknown / Unspecified	341	601	2,207
Total	901,544	979,386	1,019,680
Unknown / Unspecified	0.04%	0.06%	0.22%

Table 4-2 Percentage of unknown/unspecified births for 12 months before 2016 Census within 1, 2 or 5 years of registration

Age	Within 1 year	Within 2 years	Within 5 years
15	121,501	133,451	139,047
20	236,018	242,671	244,937
25	233,113	238,328	240,125
30	184,470	188,750	190,245
35	97,575	100,076	100,934
40	28,877	29,810	30,154
45	2,155	2,258	2,327
Unknown /Unspecified	598	938	1,626
Total	904,307	936,282	949,395
Unknown/Unspecified(%)	0.07%	0.10%	0.17%

To determine the lateness of birth registration, the live births records were examined in three categories:

1. the births that were recorded within the year of birth,
2. the births that were recorded before end of the following year and,
3. the births that were recorded before the child's fifth birthday.

Stats SA derives the status of registration from the year of birth registration and the year of birth occurrence variables. Registration year is the year in which the birth was processed and the birth year is the year in which the birth actually occurred. Stats SA defines current births as those which were registered within the same calendar year that they occurred or registered up to the end of February of the year following the calendar year of birth. Hence, late births (according to Stats SA) are births which were registered after February of the year following the calendar year in which the births occurred. It is worth noting that the Births and Deaths Registration Act, 1992 (Act No. 51 of 1992), requires a child to be registered within the first 30 days of birth. According to the Act, births registered after 30 days are considered to be late,.

The Stats SA live births records data used in the current study only gives the year of registration as the time the births were registered without month and day component data. The study sought to show the exact date of birth registration, i.e. day, month and year, so that it is possible to measure and compare accurately the time of birth registration based on the available definitions' late registration. As a result, the aforementioned three categories that represent lateness of birth registration are based only on the year a birth was registered.

For the first category, the criteria used to retrieve the births that were recorded within the year of birth was to match the year of registration to the year of birth. However, some births that occurred and were registered within the same 12-month period were excluded because they were registered in the subsequent calendar year. For example, there is a chance that some births that occurred in December 2010 were registered in January 2011. These births would be excluded from a 12-month period that spans two calendar years even if their registration falls within the first 30 days of birth and hence considered current. Consequently, for instance, this approach underestimates the births that occurred in the 12-month period before the 2011 Census. This in turn underestimates the completeness level for the October 2010-September 2011 period.

The second category of births that were recorded before the end of the next year was calculated by adding one year to the year of birth to get the number of births that were registered before they reached age two. If we consider the 12-month period before the 2011 Census, births that occurred for instance in October 2010, would turn two in October 2012. However, the criteria used for this category still derives births that occurred in October 2010 but registered after they turn two, say, in December 2012 and hence, that overestimates the births. This is because the data does not provide the day and month of registration.

In the third category the births were calculated by adding four years to the year of birth. The births that occurred, for example, between October 2010 and September 2011 (12 months before the 2011 Census) and were registered in the period of January 2016 to September 2016 are not included yet they still fall within the less than fifth birthday category. This is because the registration years added to the birth year do not allow for the specific months and days in the year that some children would have turned five. Therefore, the calculations underestimated the births for this category.

The retrieved births from the vital registration data are used as a numerator to calculate the completeness levels for the years of 2011 and 2016 (Equation 3.10). The births for the denominator are acquired from the 2011 Census and the 2016 Community Survey data and are corrected through the relational Gompertz model.

4.2 Sex Ratio Analysis

The live births records have a national sex ratio of 101.53 for the 12-month period before 2011 Census, and for the same period before 2016 Community Survey the sex ratio is 101.59. The provincial sex ratios (Table 4-3) are reasonable since they do not deviate significantly from the national levels and that there are more boys than girls.

Table 4-3 Provincial sex ratios for the 12 months before 2011 Census and 2016 CS respectively

Province	Sex Ratio 2011	Sex Ratio 2016
1. Western Cape	102.7	102.2
2. Eastern Cape	101.1	100.9
3. Northern Cape	102.8	102.9
4. Free State	100.0	102.3
5. KwaZulu-Natal	100.7	101.8
6. North West	101.7	100.9
7. Gauteng	101.8	101.5
8. Mpumalanga	101.6	100.4
9. Limpopo	102.3	102.0

At district level there is a general consistency of sex ratios for both 12-month periods before the 2011 Census and the 2016 CS respectively (Figure 7-3 of Annexure 2). However, there are districts which deviate from their respective province's average (Table 4-4). Notably, the Central Karoo has a sex ratio of 95 in 2016, Xhariep has 91 in 2016 and Zululand has 95 in 2011. Possible reasons for these irregularities could be errors in data that may be as a result of undercounting boys.

Table 4-4 District examples with at least one inconsistent sex ratio

District	Sex Ratio 2011	Sex Ratio 2016
105. Central Karoo	113.1	95.3
212. Amathole	98.8	99.4
244. Alfred Nzo	97.8	100.5
307. Pixley ka Seme	95.6	106.9
416. Xhariep	100.6	91.2
556. Zululand	94.8	103.7

4.3 Seasonality of births

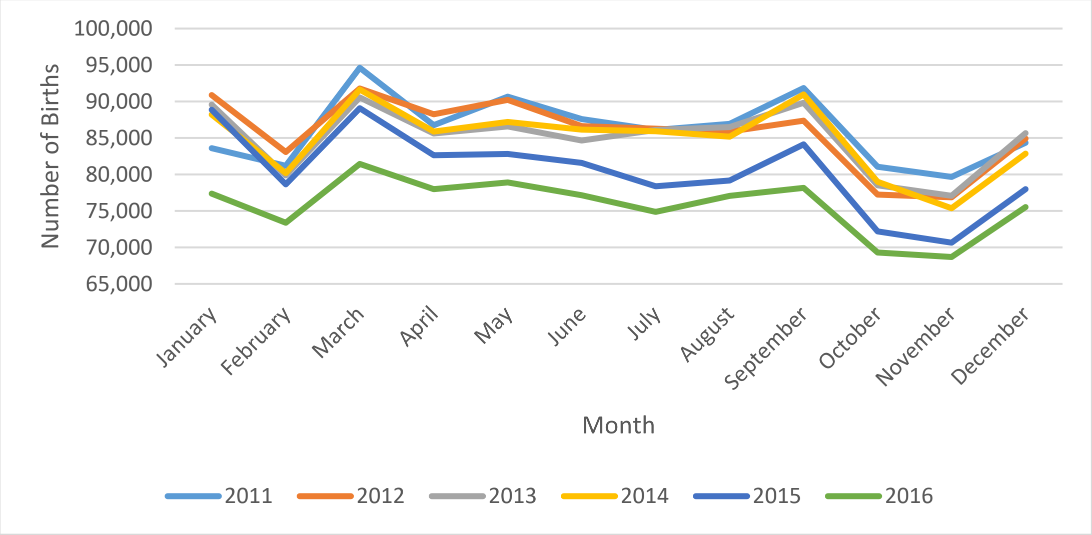
The number of births changes noticeably by season. Based on the live birth records one of the objectives of this study is to establish a pattern that births follow the different seasons. The pattern can be used as a general guideline to predict births for purposes such as administration and planning. Data on live births registered in South Africa in 2011-2016 was used to analyse the pattern (Table 4-5 and Figure 4-1). The month of March has the highest peak, followed by another peak in September. Births tend to fall steadily between April and August albeit with a slight bump in May. The month of

February and October-November are characterised by troughs. A limitation of the live births records is that they do not include socio-demographic factors that may assist in understanding the varying patterns. There is generally a similar pattern for the different years in the distribution of births across the months. However, in 2015 and 2016 there seems to be a lower number of births, most likely reflecting recent births that are not yet registered.

Table 4-5 Births occurrence by month and year (2011-2016)

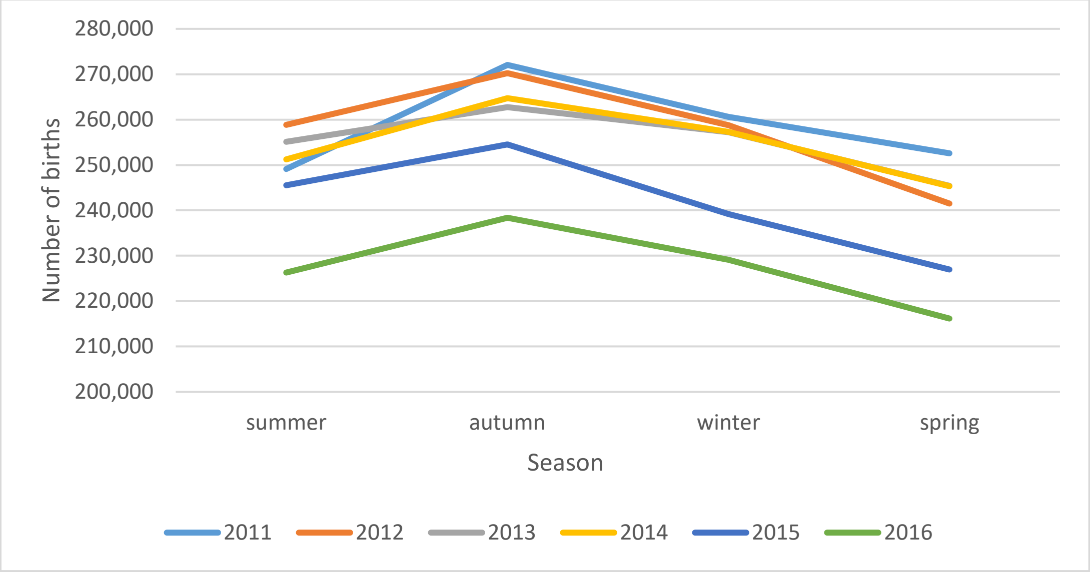
Month	2011	2012	2013	2014	2015	2016
January	83,593	90,877	89,587	88,222	88,883	77,355
February	81,172	83,077	79,877	80,162	78,667	73,388
March	94,625	91,788	90,556	91,636	89,081	81,446
April	86,758	88,244	85,626	85,902	82,621	78,008
May	90,672	90,213	86,567	87,190	82,829	78,917
June	87,597	86,621	84,641	86,167	81,588	77,164
July	86,094	86,299	86,110	85,947	78,401	74,884
August	86,917	85,868	86,512	85,173	79,182	77,052
September	91,865	87,368	89,828	90,962	84,129	78,182
October	81,045	77,260	78,509	78,998	72,181	69,283
November	79,663	76,870	77,077	75,370	70,651	68,688
December	84,350	84,901	85,647	82,877	77,989	75,548

Figure 4-1 Births by month of occurrence (2011-2016)



When the births are aggregated into the seasons of South Africa, Figure 4-2 shows that births are mostly prominent in the autumn season of all the years (i.e. March, April and May). The winter and summer season show roughly the same births while spring has the lowest births for each year. The analysis by the three-month seasons gives smoother variations than the analysis by each month which covers up details of the underlying individual months. For instance, even if the autumn season has the highest births out of all seasons, the month of April generally has lower births than the month of September which is in spring.

Figure 4-2 Births by season of occurrence (2011-2016)

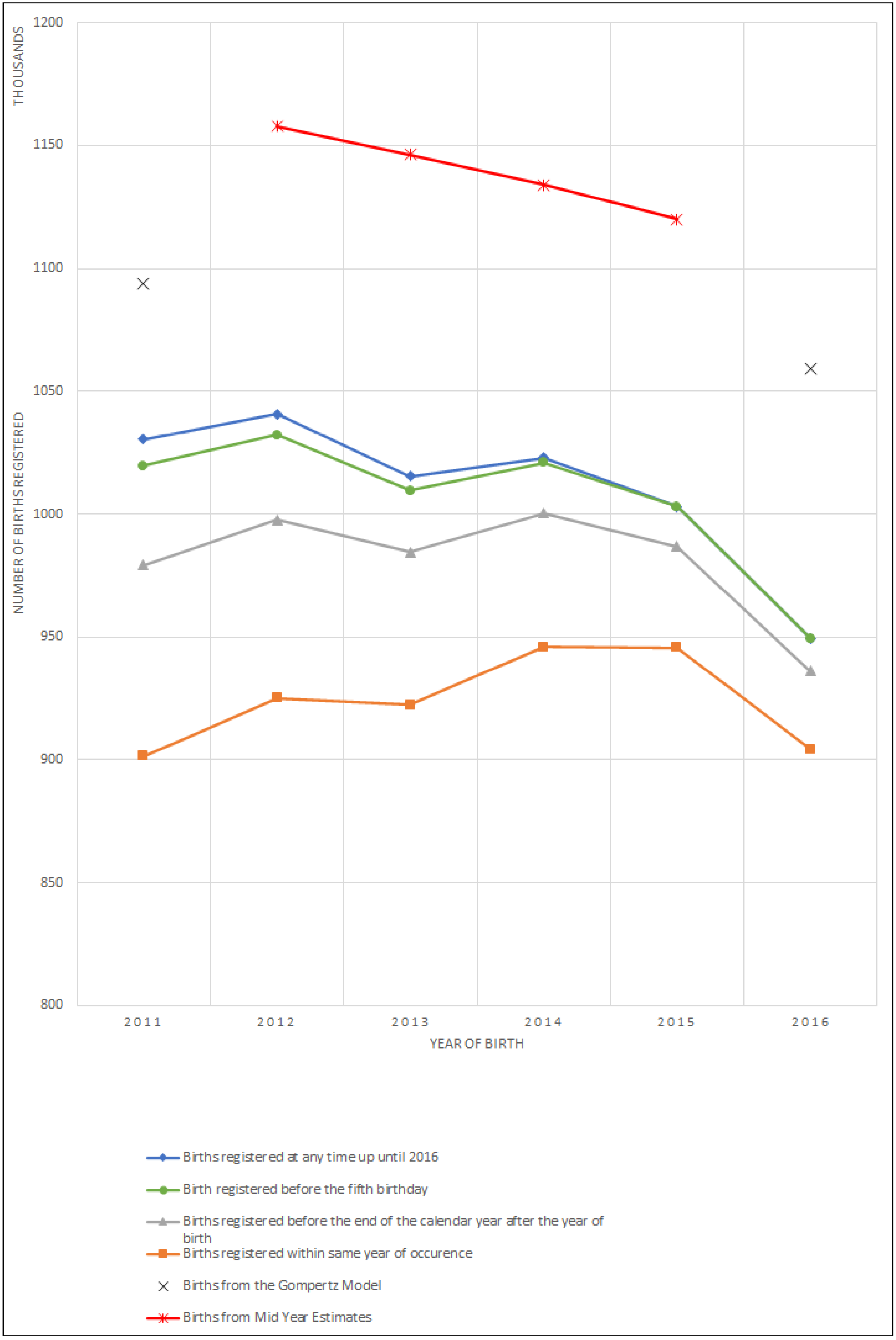


4.4 Birth registration and time of registration

Figure 4-3 shows the number of registered births recorded per year in the vital registration from 2011 to 2016. The figure further shows the estimated births derived from the Stats SA mid-year population estimates are as a result of applying the linearly decreasing ASFRs from the 2011 and 2016 Gompertz model. Moreover, the 2011 and 2016 Gompertz model birth estimates are represented in point form. The number of any time births generally increased from 2011 to 2012 from 1 030 501 to 1 040 754 before decreasing steadily down to 949 395 in 2016. The decline in births records over the years, notably from 2015 to 2016, could be due to birth registrations that are not yet captured in the more recent years' data. Apart from that, it could be evidence that fertility has been declining over the years in South Africa (TFR is 2.55 in 2011 and 2.28 in 2016) (Table 7-1 and Table 7-2 of Annexure 1)

Birth registration distinctly improves between 2011 and 2016. This is seen from the progressive increase of registered births starting with births registered within the same year of occurrence, followed with those that are registered before the end of the year following the year of birth and those registered before the child's fifth birthday. The gap between these birth registration categories decreases between 2011 and 2016 (Figure 4-3). This shows that more births were registered earlier after occurrence than in previous years. As such the births registered before the fifth birthday are much closer in number to the anytime births registered up until 2016. In spite of the improvements in births registration over the years, under-registration is noticeable when the registered births are compared to more robust estimates of births from the relational Gompertz model.

Figure 4-3 Birth Registration in South Africa 2011-2016



The point estimates of the 2011 and 2016 births that were derived from the relational Gompertz model are greater than the recorded births in the same years. This shows that the births data is under-recorded. This could be due to births that have not yet been recorded in the reference period. The estimated births are 1 094 000 for 2011 and 1 059 000 for 2016. Since it was assumed that the ASFRs decreased linearly from 2011 to 2016, the births for 2012 to 2015 were calculated through applying the respective ASFRs to the mid-year population estimates. The births derived (for 2012-2015) are also more than those acquired from the vital registration. The completeness of births registration is explored in the next section, applying the approach to estimate completeness of registration set out in Section 3.2.

4.5 Completeness of birth registration

Figure 4-4 Completeness of birth registration 2011-2016

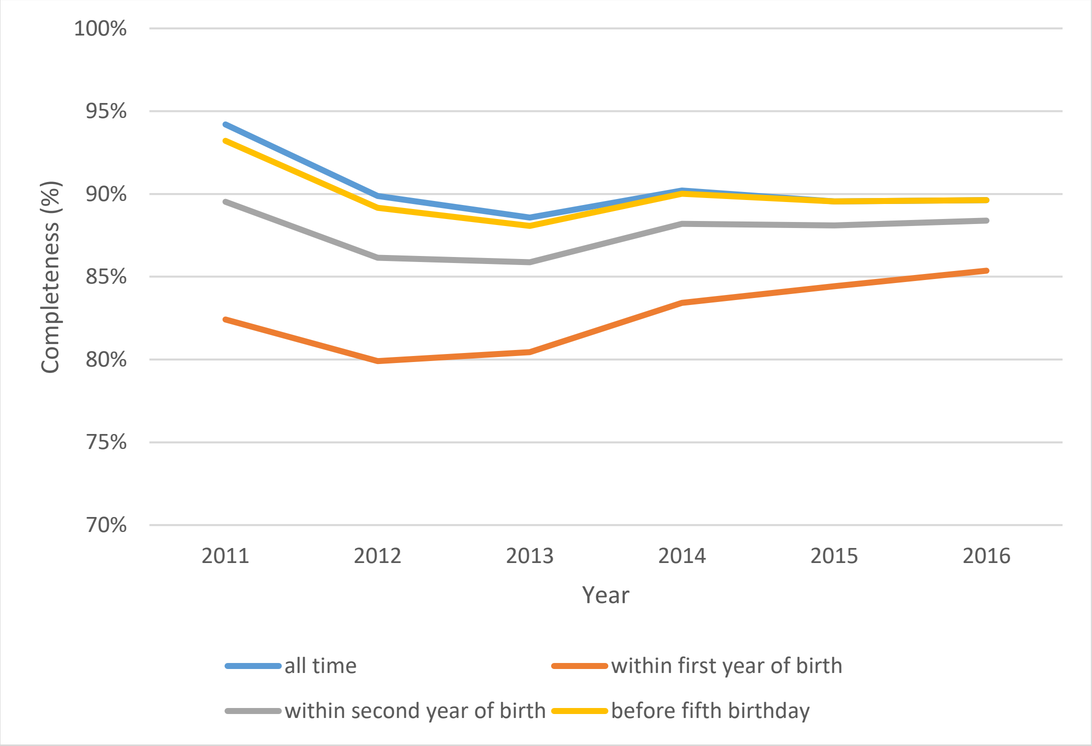


Figure 4-4 shows the estimated national completeness of births registration between 2011 and 2016. The completeness is shown by the year in which the registration occurred and it is calculated from the estimated number of births derived from the relational Gompertz model.

An improvement is noticed in birth registrations that were done within the same year that the births occurred. In 2011 about 82% of births were registered in the same year of occurrence and this increased to about 85% in 2016. There is a general steady decrease in the completeness of registrations that were done before the end of the second and fourth year after the year of birth. In 2011 about 90% births were registered before the second year after birth and this decreased to about 88% in 2016. The completeness for the registrations done before the end of the fourth year after birth decreased from about 93% down to an estimated 90%. The decrease in completeness as children get older and the increase in completeness for registration in the same year of birth occurrence shows an improvement in birth registration at younger ages.

Figure 4-5 Provincial completeness of birth registration 2011-2016

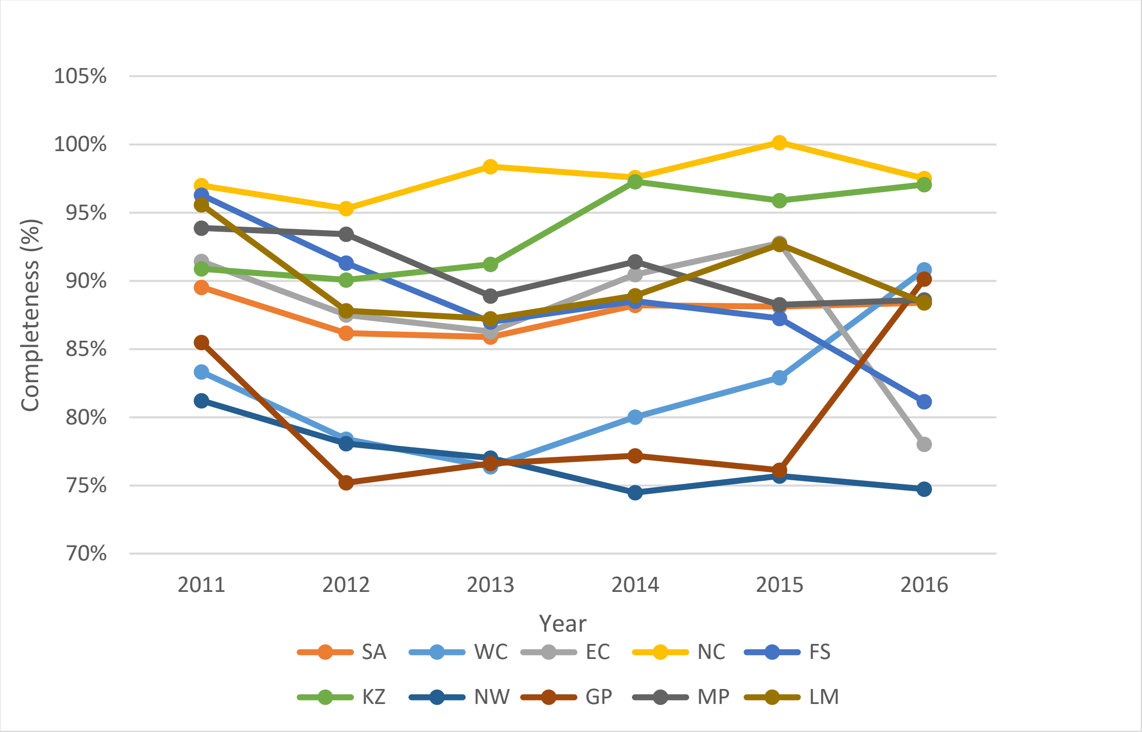


Figure 4-5 and Table 4-6 show the birth registration completeness before the end of the second calendar year after the year of birth per province derived using the relational Gompertz model. The national completeness decreased from 90% to 88% between 2011 and 2016. However, variations between provinces are evidenced. In the early years starting from 2011, the completeness for Gauteng, Western Cape and North West were below the national average. Western Cape and Gauteng increased above the national completeness level in 2016 while North West remains below with a decreasing trend. These observations are likely due to the migration of women to the big metropolitan centres. The women may have reported to live in one province while registering births in another. Free State and Eastern Cape are above the national average at first but they show a decrease below the national average after 2015. It is possible that recent births may not yet have been registered for the provinces in the period of 2015 and 2016. Northern Cape, Limpopo, Mpumalanga and KwaZulu-Natal maintain higher completeness levels above the national average throughout the period under study.

In addition, as an example, there is higher completeness of birth registration in the 2015-2016 period in Gauteng province while there is a lower completeness in the Free State province. Possibly, some mothers from Free State may have migrated to and delivered their babies in Gauteng. A similar scenario is noticed between the Eastern Cape and Western Cape, with the Western Cape as the receiving end.

Table 4-6 Completeness of birth registration (%)

Province	2011	2012	2013	2014	2015	2016
Western Cape	83%	78%	76%	80%	83%	91%
Eastern Cape	91%	87%	86%	90%	93%	78%
Northern Cape	97%	95%	98%	98%	100%	97%
Free State	96%	91%	87%	89%	87%	81%
KwaZulu-Natal	91%	90%	91%	97%	96%	97%
North West	81%	78%	77%	74%	76%	75%
Gauteng	85%	75%	77%	77%	76%	90%
Mpumalanga	94%	93%	89%	91%	88%	89%
Limpopo	96%	88%	87%	89%	93%	88%
National	90%	86%	86%	88%	88%	88%

4.6 Spatial Autocorrelation for birth registration completeness: Global Moran's *I*

To answer the question of whether or not the districts with similar completeness are clustered, randomly distributed or dispersed, the Global Moran's *I* statistic was used. Before calculating the statistic to determine the relationship of a district to its neighbours, the neighbours were defined in two different ways. One way is the Inverse Distance. This method renders more weight to the districts that are closest to the target district than the districts that are further away. Hence, a spatial relationship is sought between districts that are within a certain range of the target district. A distance of 300 kilometres was used to determine the neighbours of a target district. The distance was used because it allows every district to have at least one neighbour when measured from its centroid. The other way to define a neighbour is the Contiguity Edges Only. This approach only considers those districts which share a boundary with the target district to influence the computations of the spatial relationship.

The Global Moran's *I* analysis null hypothesis states that the births completeness of districts are randomly distributed across South Africa. Since the z-scores and p-values do not show statistical significance in clustering or dispersion, we fail to reject the null hypothesis that there is random distribution of births completeness by district. The Global Moran's Index is closer to zero than it is to -1 or +1 which further confirms the random distribution of births completeness (Figure 4-9 and Figure 4-10).

4.7 Spatial analysis of completeness in districts

There is a general tendency to higher completeness for districts in major cities (Figure 4-6 to Figure 4-8 and Figure 4-11 to Figure 4-13). This could be due to better transport and communication networks which make it easier for people to visit registration offices . Despite that observation, the neighbouring districts do not show any spatial pattern in relation to major cities as evidenced by the Global Moran's *I* results. For instance, the district of iLembe in KwaZulu-Natal which is closest to the city of Durban has low completeness (e.g. completeness within the year of birth =41.2%, before the age of two =47.0% and before the child's fifth birthday =49.7% in 2011). The district of Xhariep in the Free State despite being close to the city of Bloemfontein has low completeness levels (completeness with the year of birth =45.3%, before the age of two =48.5% and before the child's fifth birthday =52.1% in 2011). These examples may suggest that a significant number of births from these districts were recorded in the closest city where the mothers might have gone to deliver.

The national average completeness within the year of birth increased from 82% in 2011 to 85% in 2016. However, the completeness before age two and five decreased from 90% to 88% and from 93% to 90% , respectively. This may show that births were more promptly registered in 2011 than they were in 2016 as evidenced by higher earlier completeness in 2011. Notably there is a huge drop in the individual districts completeness as shown by the 2016 maps (Figure 4-11 to Figure 4-13) when compared to 2011 (Figure 4-6 to Figure 4-8). Half of the districts in the country had a percentage drop between 2011 and 2016 in the completeness with the year of birth (Table 7-3 of Annexure 1). Out of the 52 districts, 32 had a percentage drop in the completeness before age two and 37 of 52 districts dropped completeness percentage points for completeness before age five. The decline in completeness in individual districts in the 2016 CS could be attributed to the late registrations that were not yet included in the 2016 data.

There is not much change in the distribution of birth completeness between completeness before age two and age five, unlike between completeness within the year of birth and before age two for both the 2011 and 2016 data. This could suggest that more births are registered in the subsequent year after birth than they are just before the fifth birthday.

Figure 4-6 Completeness % within the year of birth (2011 census)

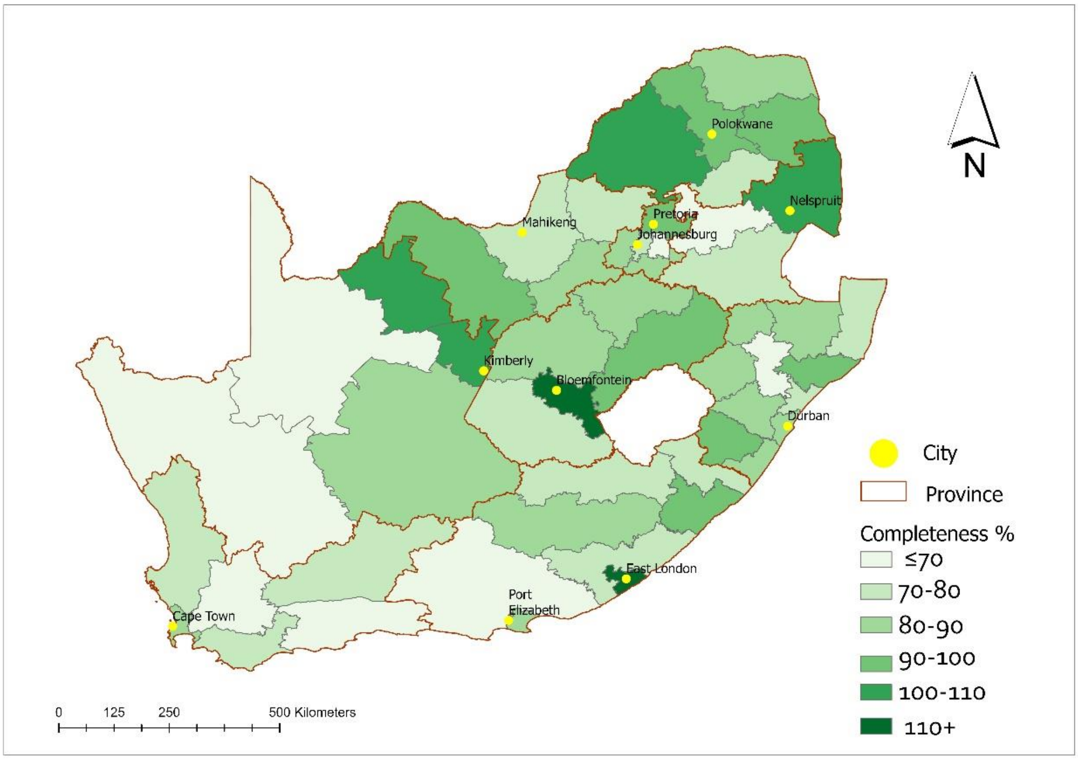


Figure 4-7 Completeness % before age 2 (2011 census)

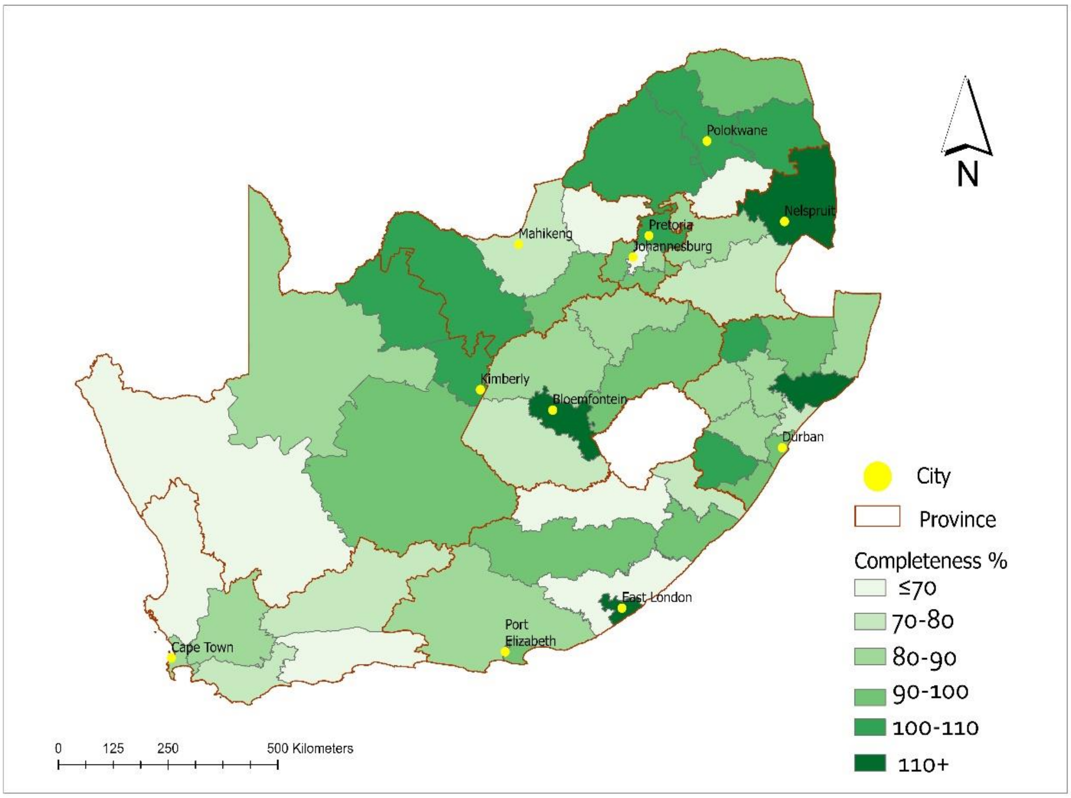


Figure 4-8 Completeness % before 5 years (2011 census)

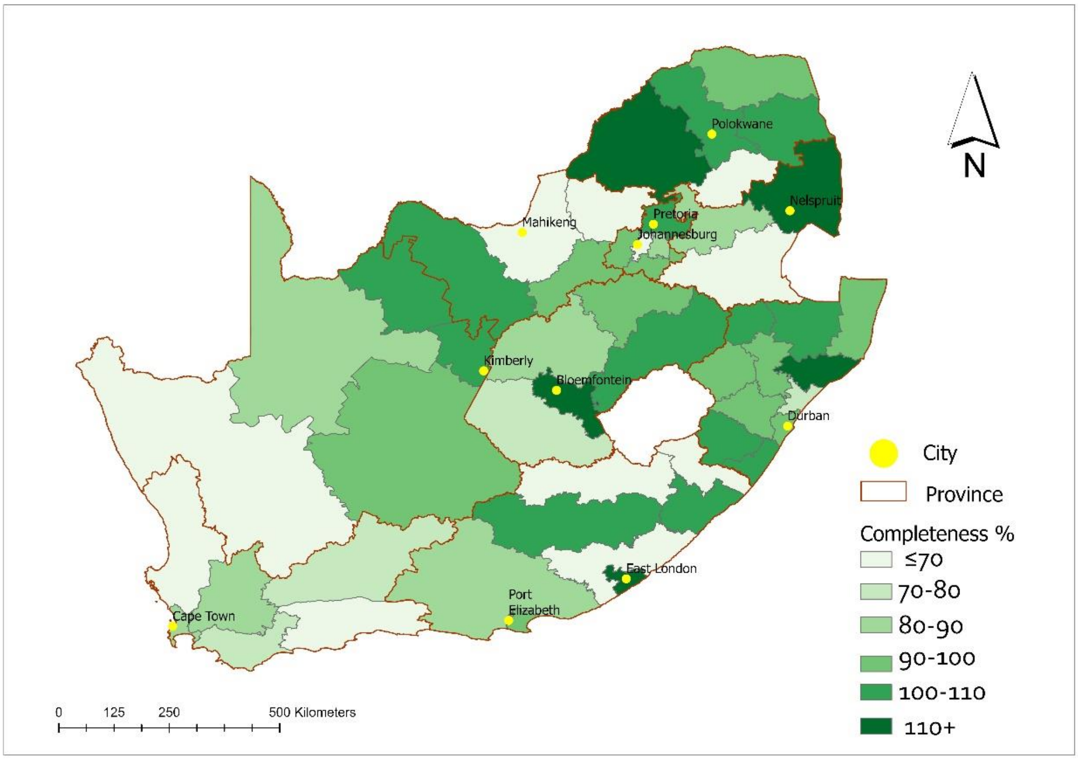


Figure 4-9: Global Moran's Index results (2011 census)

2011

Inverse distance

Completeness 1

```
Global Moran's I Summary
Moran's Index:  -0.121463
Expected Index: -0.019608
Variance:    0.005212
z-score:     -1.410882
p-value:     0.158279
```

Completeness 2

```
Global Moran's I Summary
Moran's Index:  -0.095919
Expected Index: -0.019608
Variance:    0.005171
z-score:     -1.061177
p-value:     0.288609
```

Completeness 5

```
Global Moran's I Summary
Moran's Index:  -0.079046
Expected Index: -0.019608
Variance:    0.005150
z-score:     -0.828245
p-value:     0.407532
```

Contiguity edges only

Completeness 1

```
Global Moran's I Summary
Moran's Index:  -0.126934
Expected Index: -0.019608
Variance:    0.007846
z-score:     -1.211683
p-value:     0.225634
```

Completeness 2

```
Global Moran's I Summary
Moran's Index:  -0.100370
Expected Index: -0.019608
Variance:    0.007785
z-score:     -0.915340
p-value:     0.360013
```

Completeness 5

```
Global Moran's I Summary
Moran's Index:  -0.080399
Expected Index: -0.019608
Variance:    0.007753
z-score:     -0.690410
p-value:     0.489936
```

Figure 4-10: Global Moran's Index results (2016 CS)

2016

Inverse distance

Completeness 1

```
Global Moran's I Summary
Moran's Index:  -0.087280
Expected Index: -0.019608
Variance:    0.005282
z-score:     -0.931111
p-value:     0.351796
```

Completeness 2

```
Global Moran's I Summary
Moran's Index:  -0.065608
Expected Index: -0.019608
Variance:    0.005278
z-score:     -0.633192
p-value:     0.526608
```

Completeness 5

```
Global Moran's I Summary
Moran's Index:  -0.060955
Expected Index: -0.019608
Variance:    0.005278
z-score:     -0.569129
p-value:     0.569269
```

Contiguity edges only

Completeness 1

```
Global Moran's I Summary
Moran's Index:  -0.057736
Expected Index: -0.019608
Variance:    0.007952
z-score:     -0.427575
p-value:     0.668961
```

Completeness 2

```
Global Moran's I Summary
Moran's Index:  -0.035518
Expected Index: -0.019608
Variance:    0.007945
z-score:     -0.178494
p-value:     0.858335
```

Completeness 5

```
Global Moran's I Summary
Moran's Index:  -0.030727
Expected Index: -0.019608
Variance:    0.007945
z-score:     -0.124748
p-value:     0.900723
```

Figure 4-11 Completeness % within the year of birth (2016 CS)

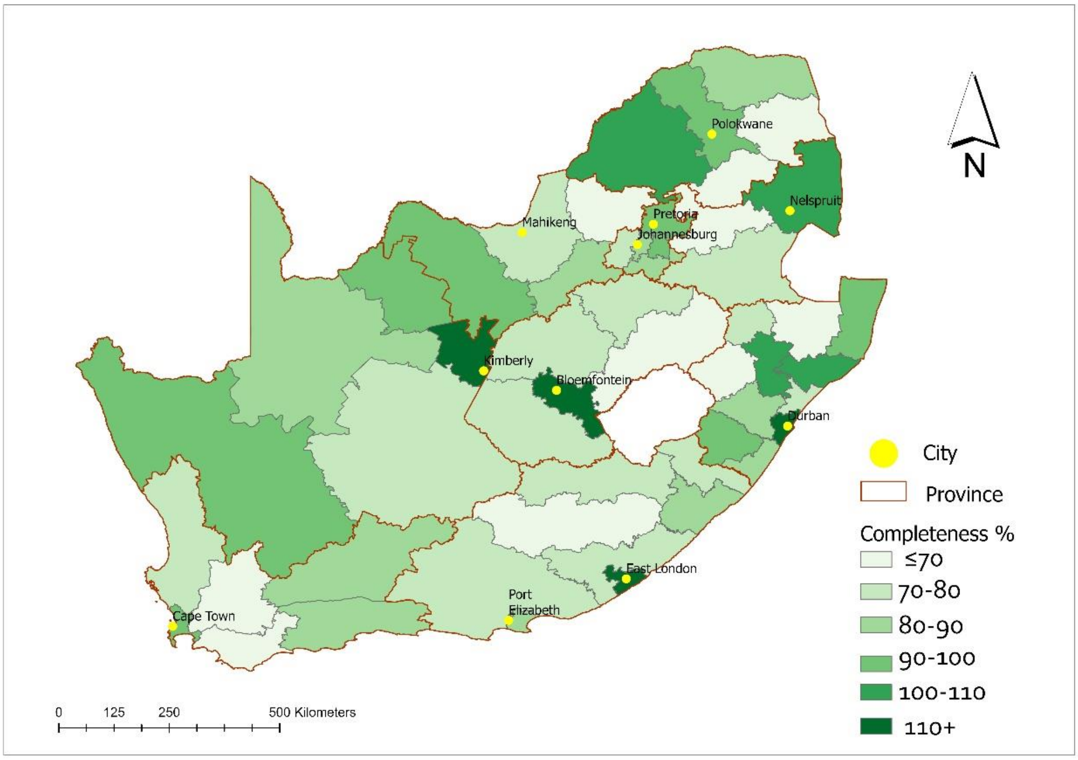


Figure 4-12 Completeness % before age 2 (2016 CS)

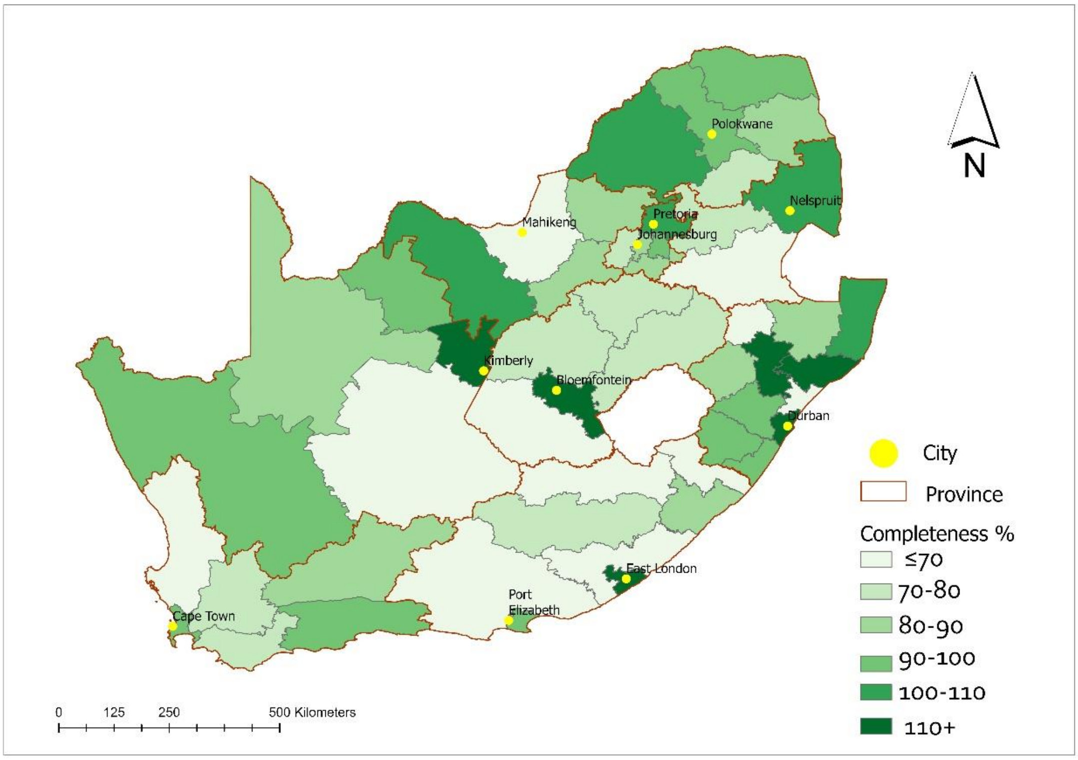
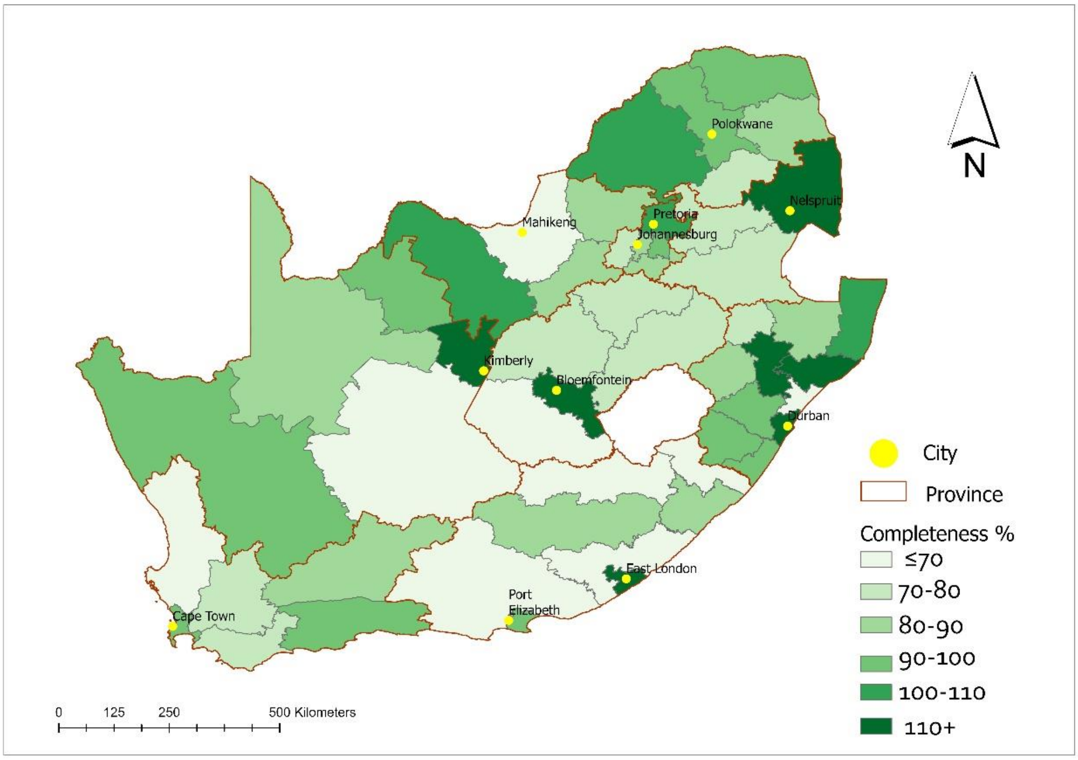


Figure 4-13 Completeness % before 5 years (2016 CS)



4.8 Spatial Autocorrelation for TFR: Global Moran's I

The null hypothesis postulates that there is a random distribution of TFR in the districts of South Africa or that there is no spatial relationship between the districts in terms of TFR. After running the Global Moran's I pattern analysis tool, the z-scores and p-values for both the 2011 and 2016 TFR show statistically significant differences in the TFR spatial patterns. In 2011, the Inverse Distance method indicates at 95% confidence level a failure to accept the null hypothesis while in the Contiguity Edges Only method (which considers shared boundaries instead of distance from district centroid), a similar decision is reached at an improved confidence level of 99%. In 2016, both methods show a failure to accept the null hypothesis at 99% confidence level. Overall, the results suggest that there are underlying spatial patterns. With a Moran's Index of 0.3014 (2011) and 0.3362 (2016) tending towards +1 this is evidence that there is clustering of similar TFR values (Table 4-7). These results prompt further investigation into the relationship between distribution of TFR and birth registration completeness (Section 4.9).

Figure 4-14 TFR distribution in 2011

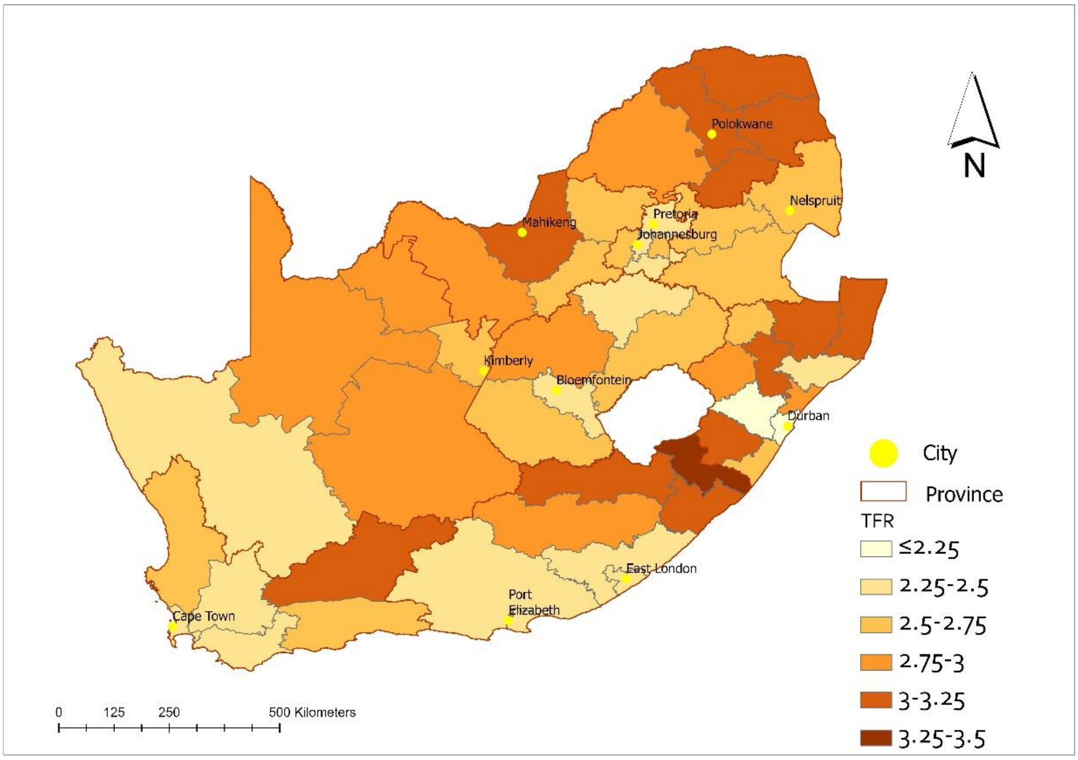


Figure 4-15 TFR distribution in 2016

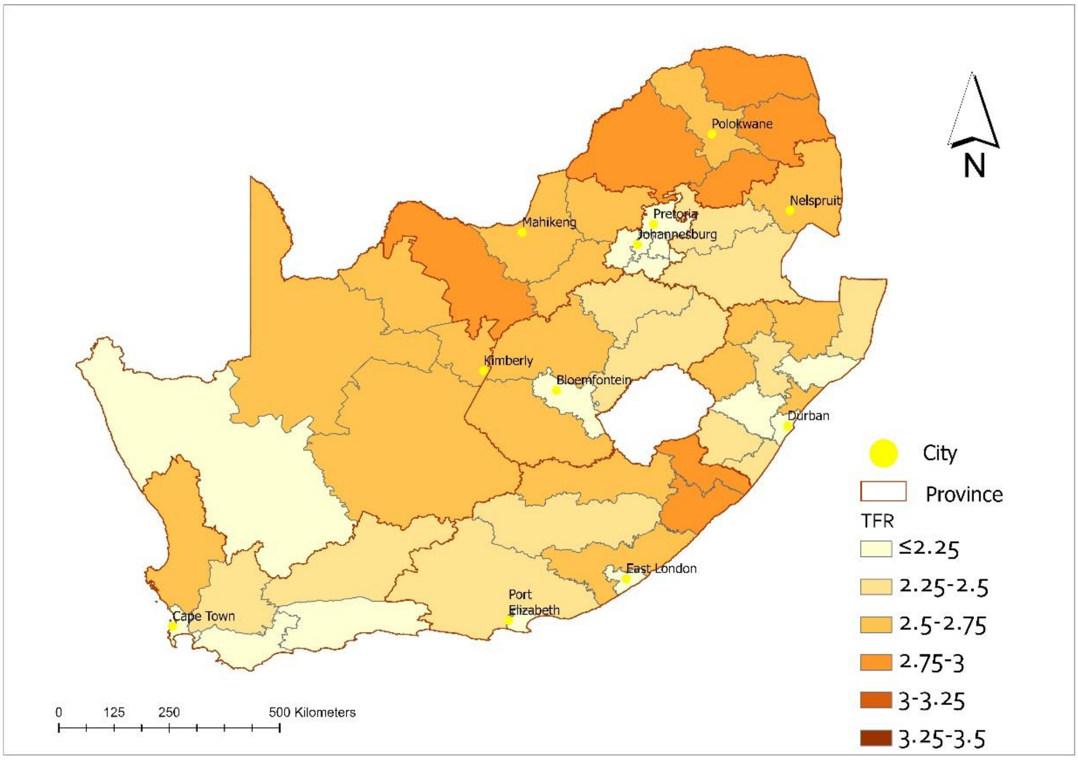


Table 4-7 TFR Global Moran's Index results (2011 and 2016)

	TFR 2011	TFR 2016
Inverse distance	Global Moran's I Summary Moran's Index: 0.163106 Expected Index: -0.019608 Variance: 0.005682 z-score: 2.423962 p-value: 0.015352	Global Moran's I Summary Moran's Index: 0.213531 Expected Index: -0.019608 Variance: 0.005630 z-score: 3.107027 p-value: 0.001890
Contiguity edges only	Global Moran's I Summary Moran's Index: 0.301407 Expected Index: -0.019608 Variance: 0.008795 z-score: 3.423018 p-value: 0.000619	Global Moran's I Summary Moran's Index: 0.336180 Expected Index: -0.019608 Variance: 0.008715 z-score: 3.811076 p-value: 0.000138

4.9 Spatial analysis of TFR in districts

There is a general decrease of TFR between 2011 and 2016, with notable changes in the provinces of Limpopo, KwaZulu-Natal, Eastern Cape and Western Cape (Table 7-1 and Table 7-2 of Annexure 1). The maps in (Figure 4-14 and Figure 4-15) show a comparison of the 2011 and 2016 TFR with visible drops in the Vhembe, Mopani, Sekhukhune, and Capricorn districts of Limpopo, the Ngaka Modiri Molema district in North West and Central Karoo district, Western Cape. Moreover, there are visually noticeable decreases in TFR in Alfred Nzo, Harry Gwala, O.R.Tambo, Joe Gqabi districts of the Eastern Cape and, Umzinyathi, Umkhanyakude and Zululand districts of KwaZulu-Natal.

There is a noticeable general pattern of higher fertility in districts that have relatively lower completeness levels. This suggests a trend of lower completeness as TFR increases (Figure 4-16 and Figure 4-17). In 2011, except the provinces of Northern Cape and Mpumalanga, the rest of South Africa's provinces exhibit a general upward trend of TFR as completeness decreases. In the Northern Cape, the trend is affected by the district of Namakwa which is the least populous in the province and has the lowest TFR of 2.35. Apart from that, the Northern Cape shows a picture of higher completeness with a decreased TFR. Mpumalanga only has three districts which have about the same TFR; however, the district of Ehlanzeni is an outlier with a completeness of 123.4%, most likely because that is where the capital Nelspruit is located. As a result, births of mothers from the other two districts may have been registered in the capital.

In 2016, the trend is visible in all the provinces except North West and Mpumalanga. The district of Ngaka Modiri Molema, where the capital of Mahikeng is situated, has the highest completeness of 103.9% and highest TFR of 2.98 in the province. Mpumalanga shows similar characteristics as in 2011.

However, as the scatter plots in (Figure 4-16 and Figure 4-17) show, there is a weak relationship between TFR and completeness at district level, leading to a conclusion that that there is no real connection between TFR and completeness. The lower fertility in major cities is likely due to urbanisation and higher completeness in most cities is because urban areas are associated with better reporting due to, for instance, better transport networks.

Figure 4-16 Relationship between TFR and completeness-2011

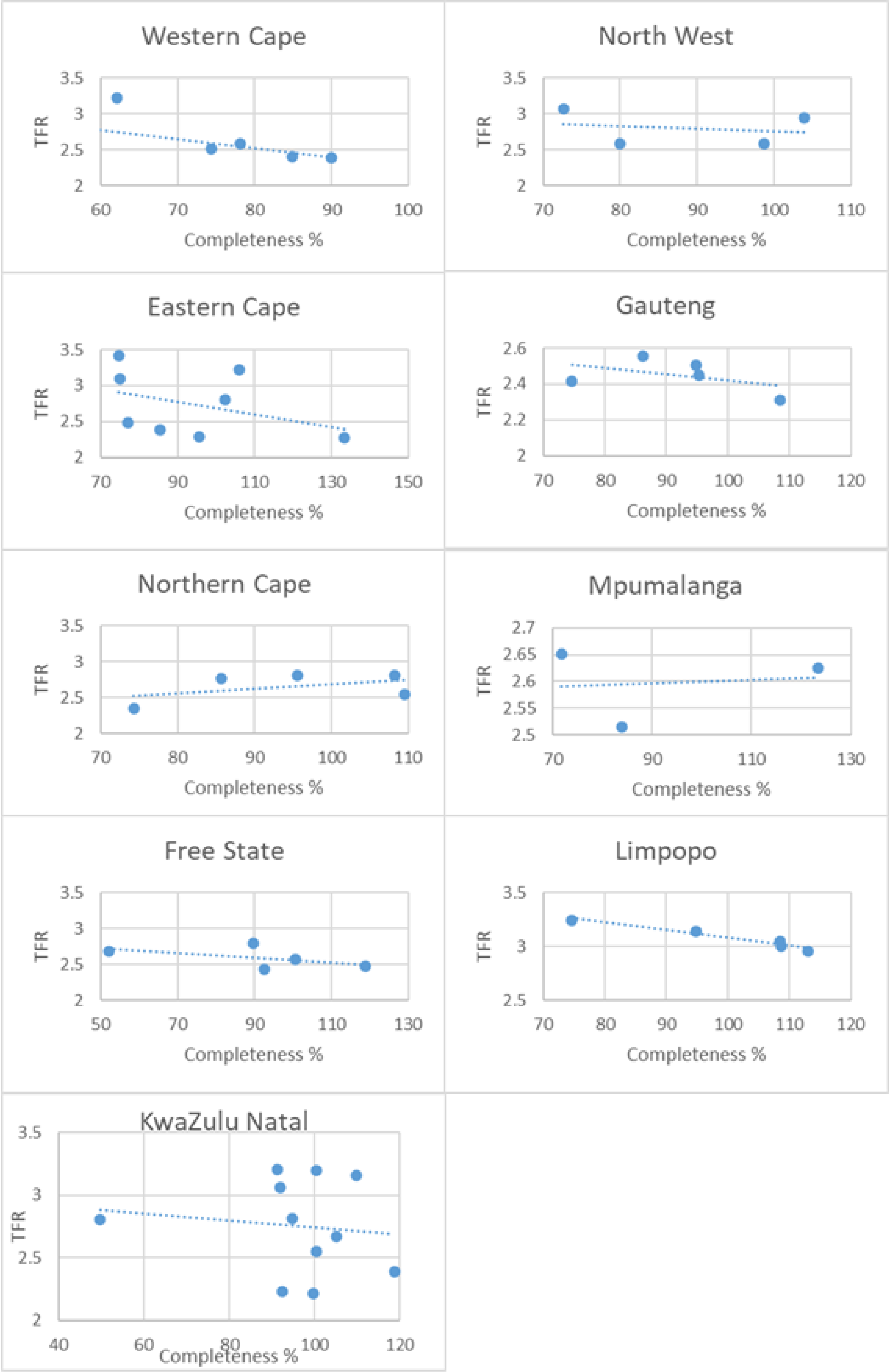
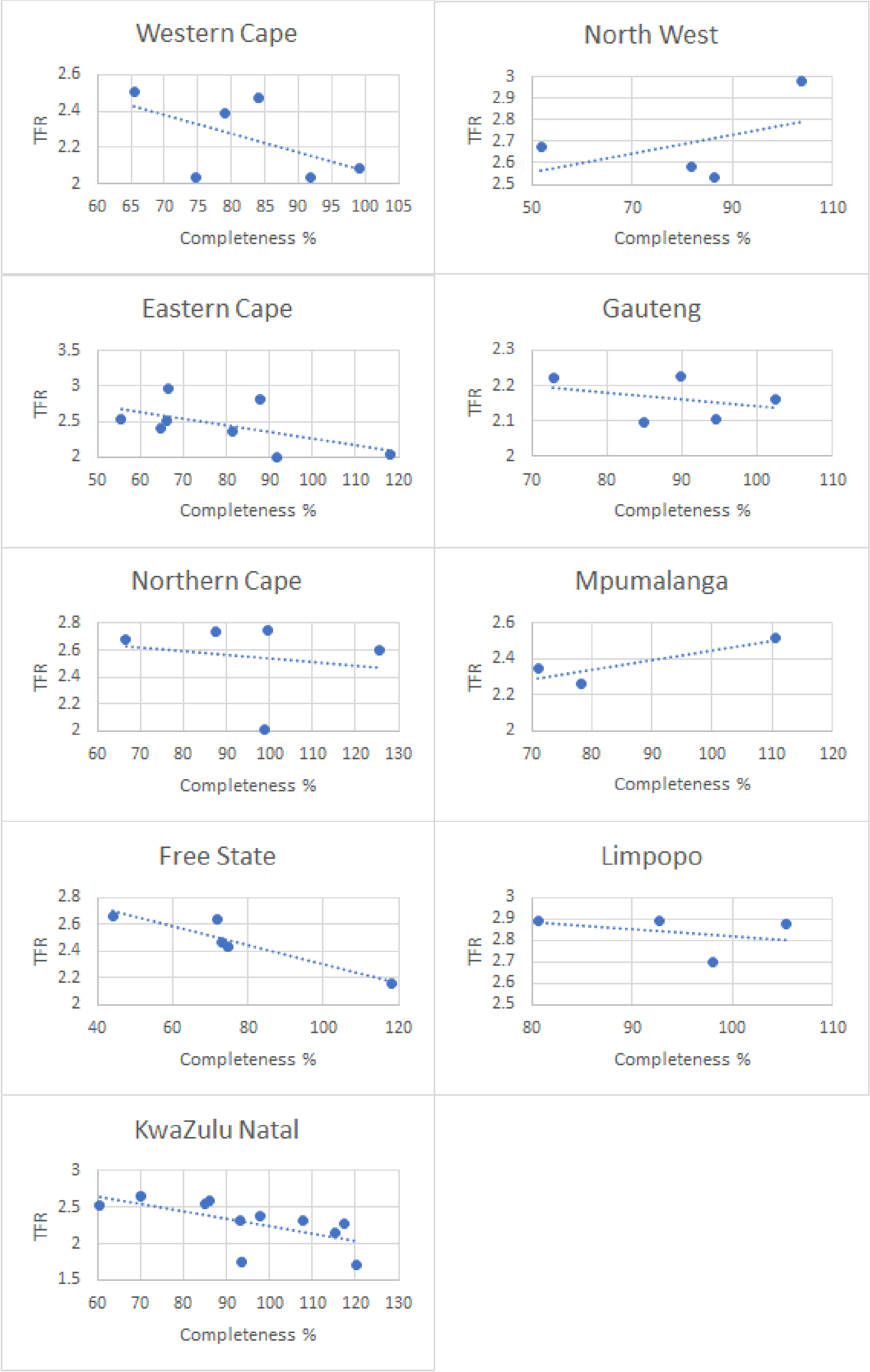


Figure 4-17 Relationship between TFR and completeness-2016



4.10 xTFR$^+$ variant of Hauer and Schmertmann's TFR results

Hauer and Schmertmann (2020)'s *xTFR*$^+$ variant method of calculating TFR was used to calculate TFRs for the 52 districts. An alternative method was necessary to evaluate and validate the TFRs derived from the relational Gompertz model. This was done to assess the reliability and verify the robustness of the results of the relational Gompertz model. A comparison of the *xTFR*$^+$ and Gompertz TFRs in 2011 shows that out of the 52 districts, 16 have a difference of more than 20% while 31 districts have more than a 10% difference in TFR (Table 7-4 of Annexure 2). In 2016, 15 districts have a more than 20% gap and 30 districts have a more than 10% gap in TFR (Table 7-5 of Annexure 2). There is a direct change of *xTFR+* as a result of an error in the Child-to-Woman Ratio. For instance, if the Child-to-Woman Ratio has an error of 20% there will be a corresponding 20% error in the *xTFR*$^+$. A 20% error in the proportion of women aged 25–34 among those who are aged 15–49 (π_{25-34}) leads to approximately 10% error in *xTFR*$^+$ and a 20% error in *q5* has almost no effect (about 1% error). This shows that the Hauer and Schmertmann (2020) *xTFR*$^+$ method requires accurate counts of Children (C), Women (W) and the age distribution of women. Hence, as a metric the method highlights errors in the data when compared with other methods.

However, for both years of 2011 and 2016 the districts which have the most differences (more than 10%) between *xTFR*$^+$ and the Gompertz model are mainly in the Eastern Cape, KwaZulu-Natal and Gauteng. In 2011 KwaZulu-Natal had 10 out of 31 (32%) districts with more than a 10% error, Eastern Cape had six out of 31 (19%) and Gauteng had four of 31 (13%) giving a total of more than 60% of the 31 districts from the three provinces (Table 7-4 of Annexure 2). In 2016, 11 out of 30 (37%) districts from KwaZulu-Natal with more than a 10% error, six out of 30 (20%) from the Eastern Cape and two out of 30 (7%) from Gauteng, amounting to more than 60% of the 30 districts (Table 7-5 of Annexure 2). These provinces are in turn, when *xTFR*$^+$ is calculated at provincial level, the top three provinces with the most differences for both 2011 and 2016 (Table 7-6 of Annexure 2).

This raises questions about the accuracy of the mid-year estimates for KwaZulu-Natal, the Eastern Cape and Gauteng which affects the Child-to-Woman Ratio for these provinces. The mid-year estimates could be affected by interprovincial migration resulting in babies being registered in a different province from the one in which their respective mothers were counted. This causes a mismatch in the number of babies against the number of mothers in a particular province and, therefore, the Child-to-

Woman Ratio is affected. In addition, at national level there are relatively smaller differences between the *xTFR*$^{+}$ and the relational Gompertz model methods. The error margin is less than 10%, i.e., 7.25% in 2011 and 4.6% in 2016. Hence, at national level the methods produce fairly similar results (Table 7-6 of Annexure 2).

The chapter analysed birth registration completeness data and the respective spatial distribution of completeness and TFR. A background check was carried out on the accuracy of the live birth records through investigating sex ratios. This was done to determine the reliability of the data and to see if there were any anomalies. The data showed some consistency with theoretically expected values. With a national sex ratio of more than 101.5 for both 2011 and 2016, there were only six out of 52 districts which had at least one sex ratio being less than 100. The data was assumed to be adequately suitable for determining reliable demographic measures, that is, completeness and TFR.

In addition, an analysis was done on the seasonality of births with notable spikes in March and September and dips in February and October-November. The birth registration data shows a similar distribution throughout a calendar year. There is a lack of an explanation to the way births differ with time due to missing data on the socio-demographic factors associated.

Likewise, more births in the recent years are most likely to not have been recorded, hence progressively lower births shown from 2012 to 2016. A decline in fertility may also have contributed towards reducing births. However, there was an increase of the number of births between 2011 and 2012. The analysis shows an increase in birth registrations between 2011 and 2016. Notably there was an improvement in the promptness of birth registration as shown by more births being registered in earlier years of life. Differences in the estimated and registered number of births between provinces were noticed and they could be as a result of inter-provincial migration.

A spatial analysis was carried out on birth completeness and TFR. Although the spatial distribution of districts based on birth completeness is not significantly different from the null hypothesis (i.e. no spatial pattern), a visual inspection shows that districts in major cities have higher birth completeness. This means that when these districts are compared to their neighbours within a specified radius, there is no apparent spatial relationship. Furthermore, the analysis results show that there is spatial clustering of districts based on their TFRs. A visual inspection also shows that the clustering is around major cities. The chapter concludes with a comparison of Hauer and

Schmertmann (2020)'s $xTFR^+$ variant and relational Gompertz TFRs. It is noted that the $xTFR^+$ method is sensitive to errors in the Child-to-Woman Ratio which directly affects the TFR result.

5. CONCLUSION

This chapter details the extent to which the aim and objectives of the research were met. In addition, an account is given of the limitations and problems encountered. The findings are discussed, establishing new insights and suggesting further studies that can be undertaken to further understand birth registration as data granularity increases.

5.1 Discussion

The study aimed to evaluate the accuracy and usefulness of birth registration data at district level in South Africa. The study sought to assess if the live births records can be of any use in calculating and analysing completeness of births and estimating Total Fertility Rate (TFR). The first objective was to establish the completeness of birth registration data. This was achieved by finding the births undercounts for 2011 and 2016 using indirect methods. Two main data sources were used for calculating the birth completeness level. The numerator was the live births records which were obtained from Statistics South Africa Datasets portal. The denominator was obtained from the births for the 12-months before the 2011 Census as well as the 2016 Community Survey, respectively. These births were corrected using the relational Gompertz model to give a more robust count of births for each of the 52 districts of South Africa. The births that were derived from the relational Gompertz model were scaled so that the district totals match their respective provinces and in turn the total of all provinces matches the national total.

The second objective was to discover if there are any trends in the birth registration data on the seasonality of births and sex ratios. The study was able to show the different seasonal patterns of birth occurrence. The main highlights are peaks in March and September and troughs in February and October-November, with a steady decline between April and August. The pattern for the different years is similar. The birth registration data, however, lacks socio-demographic data that could give insights into understanding the different patterns. Future studies can explore ways of incorporating available socio-demographic data into assessing seasonality of birth registration data

In addition, the district sex ratios were consistent with provincial and national values. There were only a few districts which slightly deviated from the normal sex ratio of at least 100, with the lowest being Xhariep which had 91.2 in 2016. Such variations could reflect under-recording of baby boys in certain districts. The objective to calculate sex ratios was to appraise the accuracy of birth registration data, to ascertain if the data is adequately accurate to produce valid and reliable results. As expected, the majority of districts had satisfactory sex ratios.

The third objective was to estimate the Age Specific Fertility Rates (ASFRs) and Total Fertility Rates (TFRs) for various subnational areas. The study illustrates the functionality of the indirect method of the relational Gompertz model in producing robust births estimates using census data. The method was validated, particularly for its ability to provide estimates of TFR at district level, by using the Hauer and Schmertmann (2020) method. However, it was noted that the Hauer and Schmertmann (2020) method relies on accurate numbers of children and women, since any error in the Child-to-Woman Ratio reflects directly in the TFR result. This sheds more light on the Hauer and Schmertmann (2020) method's ability to detect errors in data particularly the Child-to-Woman Ratio when compared to other methods.

The fourth objective sought to establish, with the use of GIS techniques, if there are any spatial variations of completeness and fertility at a subnational level. The study managed to obtain completeness percentage levels and TFRs for each district. Further analysis was done to determine if any patterns and trends existed in the results obtained. One such analysis was to examine the spatial relationships between districts based on their respective birth completeness as well as TFR. Using the GIS technique of Global Moran's *I* spatial autocorrelation, the results showed that there were no geographical patterns to be found between districts as a result of their completeness. Although the Global Moran's *I* statistic does not show any clustering of similar or dissimilar completeness values in districts, a visual inspection of the produced maps shows that completeness tends to be high in most major metropolitan areas. Examples of such areas are the districts of Buffalo City, Mangaung, Frances Baard, Ehlanzeni and City of Tshwane. Overall at district level, however, the data shows that there is no spatial pattern of birth completeness.

Further analysis was done on the spatial distribution of TFR. The Global Moran's *I* results show a clustering of similar TFR values at district level for both 2011 and 2016. Urban areas were most likely associated with clusters of low TFR levels because women in urban areas have a high contraceptive usage and higher levels of education. Although in most urban areas where TFR is low and completeness is high, this does not explain a connection between TFR and completeness. The high completeness is mainly due to better reporting associated with urban areas.

The fifth objective of the study seeks to find any problems encountered when dealing with subnational birth registration data. One of the limitations that was encountered early in the study was that the birth registration data only records the year part of the time that births were registered. The absence of an exact date of registration that includes day and month limited the study's ability to investigate the timing of birth registrations in comparison to the DHA's requirement of 30 days to register a birth. This limited the completeness analysis to only be done using complete years after the year of birth. As a result, the child's age by which birth completeness was calculated is estimated. It is noted that there has not been a clear definition of the timeliness of birth registration with respect to birth year. However, as underscored in Nannan, Dorrington and Bradshaw (2019), apart from adopting the WHO definition of completeness that compares registered births against the actual births, the current study also includes the timing of registration based on a particular age of child i.e. within the year of birth, before age two and before age five.

Apart from that, the possible migration of children before they are registered poses a challenge in assessing completeness especially between neighbouring provinces or districts. The study assumed no migration since there was no migration data and the process of documenting migrants is not straightforward.

A further limitation of the study was the level of disaggregation on which it mainly focused. The population data at district level should be carefully interpreted due to its sensitivity to errors that may not be adequately addressed by a post-enumeration survey (PES). A PES collects high quality data only for a relatively small number of enumeration areas (EA)(e.g. 600 EAs out of 89 305 in 2011) after a census is completed. The PES quantifies coverage and content errors for national, provincial and urban and non-urban geographic levels. The matching process is used to determine and hence adjust the undercounts by comparing the PES and the census. However, the estimates provided by the PES can "only be reliable at national and provincial levels"(Statistics South Africa 2012a: p.7). Since the current study was focused mainly on district levels, the census data's reliability may be questionable as the population numbers are not appropriately adjusted for coverage and content errors.

5.2 Conclusion

The study noted that at district level there was a shift of growth in the completeness levels between completeness within the year of birth, before the age of two, and before age five from 2011 to 2016. Completeness within the year of birth had the highest overall growth (3%), while completeness from birth to age two dropped by 1% and completeness from birth to age 5 dropped by 4% (Table 7-3 of Annexure 1). This demonstrates that the promptness of birth registration improved between 2011 and 2016 since births tend to be registered in earlier days of life.

Moreover, while Nannan, Dorrington and Bradshaw (2019) noted that by 2011, there was an 83% birth registration completeness before the end of the year after the calendar year of birth, the current study found out that post-2011 there was an increase in birth registration completeness at a younger child age. That is, the completeness within the year of birth increased from 82% to 85% for the 2011-2016 period (Table 7-1 and Table 7-2 of Annexure 1). This shows that mothers are registering their babies at younger ages than before. Since the Nannan, Dorrington and Bradshaw (2019) study was based on provincial levels, they suggest a deeper look into problems that particular communities encounter which may deter them from achieving better completeness. The current study managed to identify at a lower level of disaggregation the districts that have low completeness levels. This can help policy makers to identify the problems that those specific districts are facing and how they can best channel resources to solve the problems.

As noted by Jaspers (1994) two factors affect birth registration. It is either that the public does not have enough willingness to report vital events (environmental) or, that the registration offices are hard to access with related costs (institutional). Notably the South African government has made successful attempts towards dealing with these factors (AbouZahr, De Savigny, Mikkelsen *et al.* 2015; Bah 2009; Nannan, Dorrington and Bradshaw 2019). Different legislation has been put in place to encourage the public to register births. The amendment of the Births and Deaths Registration Act of 1992, readily available registration information during antenatal visits and the introduction of Child Support Grants worked together in hastening the birth registration process.

5.3 Recommendations

The results of this study are envisioned to assist authorities in their future endeavours to improve birth registration through paying specific attention to districts with low completeness which therefore need the most help. The study can form a backdrop for similar studies for areas at lower geography hierarchies such as municipalities in South Africa.

Further studies could focus on what determines the varying seasonal patterns in South Africa and how available data can be used to predict births. The studies can consider birth seasonality in the context of factors as noted by Ellison, Valeggia and Sherry (2005) i.e. social and cultural, climatological and energetic factors. GIS overlay analysis can be implemented in integrating different factors by combining information from one GIS layer (for example, temperature with another GIS layer like rainfall) to deduce any possible spatial relationships between births and associated factors.

Completeness and TFR could not be evaluated for earlier data, that is the census of 2001 and community survey of 2007, since there were boundary changes over time. This would need the data to be aligned so that the boundaries can match the 2016 municipal demarcation. The process of boundary alignment can be managed by using areal interpolation (Gregory, Dorling and Southall 2001; Weir-Smith 2016), however, some degree of error which varies from one spatial unit to another would be introduced. Nonetheless, since the 2007 Community Survey was recoded so that it becomes comparable with the 2001 Census, a similar study to the current one can investigate and establish the birth registration completeness in those earlier years at a district level. This could give an insight into the usefulness of earlier data sources in determining how birth registration evolved for individual districts.

6. REFERENCES

AbouZahr, C., D. De Savigny, L. Mikkelsen, P.W. Setel *et al.* 2015. "Civil registration and vital statistics: Progress in the data revolution for counting and accountability", *The Lancet* **386**(10001):1373-1385.

Abrahams, K., L. Berry and M.J. Hendricks. 2006. "Child health: The general context.," in Monson J, Hall K, Smith C, Shung-King M (ed). *South African child gauge 2006*. Cape Town: Children's Institute, University of Cape Town, pp. 74-75.

Andreev, K., V. Kantorová and J.J. Bongaarts. 2013. "Demographic components of future population growth", *New York: Population Division, Department of Economic Social Affairs, United Nations, New York*

Bah, S.J. 2009. "Multiple forces working in unison: the case of rapid improvement of vital statistics in South Africa post-1996", *World Health & Population* **11**(1):50-59.

Berry, L., A. Dawes and L. Biersteker. 2013. "Getting the basics right: an essential package of services and support for ECD," in Lizette Berry, Linda Biersteker, Andrew Dawes, and Lori Lake and Charmaine Smith (eds). *South African Child Gauge 2013*. Cape Town: Children's Institute, University of Cape Town, pp. 26-33.

Bogue, D.J. and J.A.J. Palmore. 1964. "Some empirical and analytic relations among demographic fertility measures, with regression models for fertility estimation", *Demography* **1**(1):316-338.

Booth, H.J. 1984. "Transforming Gompertz's function for fertility analysis: The development of a standard for the relational Gompertz function", *Population Studies* **38**(3):495-506.

Bradshaw, D., D. Kielkowski and F.J. Sitas. 1998. "New birth and death registration forms-a foundation for the future, a challenge for health workers?", *South African Medical Journal* **88**(8):971-974.

Brass, W.J. 1978. *The relational Gompertz model of fertility by age of woman*. London: Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Cappa, C. and T. Wardlaw. 2013. *Every child's birth right: inequities and trends in birth registration*. UNICEF, Data and Analytics Section, Division of Policy and Strategy. https://www.cmsny.org/wp-content/uploads/Cappa.pdf.

Chen, W. 2016. "Evaluation of the Completeness of Birth Registration in China Using Analytical Methods and Multiple Sources of Data," Paper presented at United Nations Expert Group Meeting on "Methodology and lessons learned to evaluate the completeness and quality of vital statistics data from civil registration. New York, 3-4 November 2016.

Croft, T.N., A.M. Marshall, C.K. Allen, F. Arnold *et al.* 2018. *Guide to DHS statistics.* Rockville, Maryland, USA ICF. https://dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_DHS-7.pdf.

Danel, I. and M. Bortman. 2008. *An assessment of LAC's vital statistics system: The foundation of maternal and infant mortality monitoring*. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/206651468177844686/pdf/448620WP0Box3210paper01LACVitalStat.pdf

DHA. 2012. *Department of Home Affairs annual report 2011–2012*. Pretoria: Department of Home Affairs. http://www.dha.gov.za/files/Annual%20Reports/Annual_Report_2011-2012.pdf

Ellison, P., C. Valeggia and D. Sherry. 2005. "Human birth seasonality," in Brockman D, Schaik C (ed). *Seasonality in Primates: Studies of Living and Extinct Human and Non-Human Primates*. Vol. 44. New York: Cambridge University Press, pp. 379–400.

ESRI. 2020. *ArcGIS pro Release 2.5.0* Redlands, CA: Environmental Systems Research Institute.

Garenne, M., M.A. Collinson, C.W. Kabudula, F.X. Gómez-Olivé *et al.* 2016. "Completeness of birth and death registration in a rural area of South Africa: the Agincourt health and demographic surveillance, 1992–2014", *Global Health Action* **9**(1):32795.

Getis, A. and J.K. Ord. 2010. "The analysis of spatial association by use of distance statistics," in Rey, Luc AnselinSergio J. (ed). *Perspectives on spatial data analysis*. Berlin, Heidelberg: Springer, pp. 127-145.

Gilson, L. and D.J. McIntyre. 2007. "Post-apartheid challenges: household access and use of health care in South Africa", *International Journal of Health Services* **37**(4):673-691.

Goodchild, M.F. 1986. *Spatial autocorrelation.* USA: Geo Books. https://books.google.co.za/books?id=2BYnAQAAIAAJ.

Gregory, I., D. Dorling and H.J. Southall. 2001. "A century of inequality in England and Wales using standardized geographical units", *Area* **33**(3):297-311.

Griffith, D.A. 2017. "Spatial Weights," in D. Richardson, N. Castree, M.F. Goodchild, A. Kobayashi, W. Liu and R.A. Marston (ed). *International Encyclopedia of Geography: People, the Earth, Environment and Technology* Chichester, UK: John & Sons, pp. 1-7.

Hachadoorian, L., S.R. Gaffin and R. Engelman. 2011. "Projecting a gridded population of the world using ratio methods of trend extrapolation," in Gorenflo, Richard P. CincottaLarry J. (ed). *Human population*. Berlin, Heidelberg: Springer, pp. 13-25.

Hauer, M., J. Baker and W. Brown. 2013. "Indirect Estimates of Total Fertility Rate Using Child Woman/Ratio: A Comparison with the Bogue-Palmore Method", *PLoS One* **8**(6):e67226.

Hauer, M.E. and C.P.J. Schmertmann. 2020. "Population pyramids yield accurate estimates of total fertility rates", *Demography* **57**(1):221-241.

Hijmans, R.J. 2020. *Spatial Data Science with R*. https://rspatial.org/raster/analysis/3-spauto.html. Accessed: 17 December 2020.

Jaspers, D. 1994. "La calidad de las estadísticas vitales en América latina," Paper presented at Seminario Internacional de Calidad Estadística. Bogota, 03 June 1994.

Jewkes, R. and K.J. Wood. 1998. "Competing discourses of vital registration and personhood: perspectives from rural South Africa", *Social Science Medicine* **46**(8):1043-1056.

Johnson, L.F., R.E. Dorrington and H.J. Moolla. 2017. "Progress towards the 2020 targets for HIV diagnosis and antiretroviral treatment in South Africa. Southern African journal of HIV medicine", *Sothern African Journal of HIV* **18**(1):a694.

Lima, E.E.C.d. and B.L.J. Queiroz. 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**(2014):1721-1730.

Marti-Henneberg, J., X. Franch-Auladell and J.J. Solanas-Jiménez. 2016. "The Use of Digital Tools for spatial analysis in Population Geography", *Frontiers in Digital Humanities* **3**(2016):9.

Matthews, S.A. and D.M.J. Parker. 2013. "Progress in spatial demography", *Demographic Research* **28**:271-312.

Mikkelsen, L., A. Lopez and D.J. Phillips. 2015. *Why Birth and Death Registration Really Are Vital Statistics Development. Human development reports*. UNDP. http://hdr.undp.org/en/content/why-birth-and-death-registration-really-are-%E2%80%9Cvital%E2%80%9D-statistics-development

Mitchell, A. 2005. *The ESRI Guide to GIS Analysis (Volume 2)*. Redlands, CA: Esri Press. https://esripress.esri.com/display/index.cfm?fuseaction=display&websiteID=86&moduleID=0.

Moultrie, T.A.J. 2013. "The relational Gompertz model," in Moultrie TA, RE Dorrington, AG Hill, K Hill, IM Timæus and B Zaba (ed). *Tools for Demographic Estimation*. Vol. 1621459. Paris: International Union for the Scientific Study of Population. demographicestimation.iussp.org, pp. 54-68.

Nannan, N., R. Dorrington and D.J. Bradshaw. 2019. "Estimating completeness of birth registration in South Africa, 1996–2011", *Bulletin of the World Health Organization* **97**(7):468-476.

National Research Council. 2009. *Vital statistics: Summary of a workshop. Michael J. Siri and Daniel L. Cork, rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education*. Washington, DC: National Academies Press.

Oliveau, S. and C.J. Guilmoto. 2005. "Spatial correlation and demography Exploring India's demographic patterns," Paper presented at XXVe Congrès International de la Population. Tours, France, July 2005.

PAHO. 2017. "Plan of action for the strengthening of vital statistics 2017-2022 " Paper presented at 29th Pan American Sanitary Conference. 69th Session of the Regional Committee of WHO for the Americas. Washington, D.C., 25-29 September 2017.

Phillips, D.E., T. Adair and A.D.J. Lopez. 2018. "How useful are registered birth statistics for health and social policy? A global systematic assessment of the availability and quality of birth registration data", *Population health metrics* **16**(1):21.

Preston, S.H.J. 1983. "An integrated system for demographic estimation from two age distributions", *Demography* **20**(2):213-226.

Queiroz, B.L. and E.E. Lima. 2016. *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* https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/unpd_egm_nov2016_s2_lima-queiroz-fertility_2016.pdf

Rele, J.R. 1967. *Fertility analysis through extension of stable population concepts*. Berkeley: Institute of International Studies, University of California.

Salvacion, A.R., D.B.J. Magcale-Macandog and I. Cultures. 2015. "Spatial analysis of human population distribution and growth in Marinduque Island, Philippines", *Journal of Marine* **4**(1):27-33.

Sartorius, B.K., K. Sartorius, T.F. Chirwa and S.J. Fonn. 2011. "Infant mortality in South Africa-distribution, associations and policy implications, 2007: an ecological spatial analysis", *International journal of health geographics* **10**(1):61.

Schmertmann, C.P. and M.E.J. Hauer. 2019. "Bayesian estimation of total fertility from a population's age–sex structure", *Statistical Modelling* **19**(3):225-247.

Setel, P.W., C. AbouZahr, A. Karpati and M. Bratschi. 2018. "Civil Registration and Vital Statistics (CRVS), and the Sustainable Development Goals (SDGs)," Paper presented at International Conference on Civil Registration and Vital Statistics. Dhaka, Bangladesh, 23-25 January 2018

StataCorp. 2017. *Stata Statistical Software: Release 15* College Station, TX: StataCorp LLC.

Statistics South Africa. 2012a. *Post-enumeration Survey (PES): Results and Methodolody*. Pretoria: Statistics South Africa. http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_PES.pdf

Statistics South Africa. 2012b. *Statistical Release P0301.4 (Revised): Census 2011*. Pretoria: Statistics South Africa. https://www.statssa.gov.za/publications/P03014/P030142011.pdf

Statistics South Africa. 2020a. *Statistical Release P0302: Mid-year population estimates*. Pretoria: Statistics South Africa. http://www.statssa.gov.za/publications/P0302/District%20Council%20projection%20by%20sex%20and%20age%20(2002-2020).xlsx

Statistics South Africa. 2020b. *Statistical Release P0305: Recorded live births 2019*. Pretoria: Statistics South Africa. http://www.statssa.gov.za/publications/P0305/P03052019.pdf

Timæus, I.M. and T.A.J. Moultrie. 2013. "Estimation of fertility by reverse survival," in Moultrie TA, RE Dorrington, AG Hill, K Hill, IM Timæus and B Zaba (ed). *Tools for Demographic Estimation*. Vol. 1621459. Paris: International Union for the Scientific Study of Population. demographicestimation.iussp.org, pp. 82-93.

UN DESA. 2016. *Sustainable Development Goals*. https://sdgs.un.org/goals/goal16. Accessed: 05 October 2020.

UNICEF. 1989. *Convention on the Rights of the Child*. New York: UN Commission on Human Rights. Available from: https://www.unicef.org/child-rights-convention/convention-text#

UNICEF. 2002. *Birth registration: right from the start*. Florence, Italy: UNICEF Innocenti Research Centre. https://www.unicef-irc.org/publications/pdf/digest9e.pdf.

UNICEF. 2019. *The State of the World's Children 2019. Growing Well in a Changing World. Children, Food and Nutrition*. Geneva, Switzerland: UNICEF. https://www.unicef.org/media/63016/file/SOWC-2019.pdf.

UNSD. 2017. *Demographic and Social Statistics: Civil Registration and Vital Statistics, Coverage of Birth and Death Registration*. https://unstats.un.org/unsd/demographic-social/crvs/index.cshtml. Accessed: 05 February 2021.

Wachter, K.W. 2014. *Essential demographic methods*. Cambridge, Massachusetts: Harvard University Press. http://www.jstor.org/stable/j.ctt6wps5v.

Wachter, K.W.J. 2005. "Spatial demography", *Proceedings of the National Academy of Sciences* **102**(43):15299-15300.

Weeks, J.R.J. 2004. "The role of spatial analysis in demographic research," in Michael F. Goodchild, Donald G. Janelle (ed). *Spatially integrated social science*. New York: Oxford University Press, pp. 381-399.

Weir-Smith, G.J. 2016. "Changing boundaries: Overcoming modifiable areal unit problems related to unemployment data in South Africa", *South African Journal of Science* **112**(3-4):1-8.

WHO. 2010. *Improving the quality and use of birth, death and cause-of-death information: guidance for a standards-based review of country practices*. Geneva, Switzerland: World Health Organization. https://apps.who.int/iris/handle/10665/44274.

Yaya, Y., T. Data and B.J. Lindtjørn. 2015. "Maternal mortality in rural south Ethiopia: outcomes of community-based birth registration by health extension workers", *PLoS One* **10**(3):e0119321.

Zaba, B. 1981. *Use of the relational Gompertz model in analysing fertility data collected in retrospective surveys*. Centre for Population Studies, London School of Hygiene & Tropical Medicine.

Zewdie, S.A. 2014. "Spatial analysis of child mortality in South Africa in relation to poverty and inequality: evidences from the 2011 census." Unpublished MPhil thesis, Cape Town, South Africa: University of Cape Town.

7. ANNEXURES

Annexure 1: TABLES

Table 7-1 Completeness and TFR results (2011)

District	completeness1	completeness2	completeness5	TFR (2011)
101.WestCoast	66.3	72.1	74.3	2.52
102.Cape Winelands	78.9	82.9	84.9	2.41
103.Overberg	49.8	54.6	56.9	2.45
104.Eden	72.5	76.6	78.1	2.58
105.CentralKaroo	58.3	61.0	62.1	3.22
199.CityOfCapeTown	83.3	87.8	89.9	2.39
210.Cacadu	74.8	81.5	85.5	2.39
212.Amathole	64.2	71.6	77.0	2.48
213.ChrisHani	89.2	97.7	102.5	2.80
214.Ukhahlamba	63.7	70.7	74.9	3.10
215.ORTambo	90.0	100.0	106.1	3.22
244.AlfredNzo	62.9	69.7	74.7	3.42
260.BuffaloCity	124.2	130.5	133.5	2.28
299.NelsonMandelaBay	86.7	92.5	95.7	2.29
306. Namakwa	71.6	73.1	74.3	2.35
307. Pixley_ka_Seme	84.9	93.0	95.5	2.81
308. Siyanda	79.0	83.8	85.7	2.76
309. Frances Baard	103.1	107.2	109.5	2.54
345. John Taolo Gaets	102.2	106.2	108.3	2.81
416.Xhariep	45.3	48.5	52.1	2.68
418. Lejweleputswa	82.9	87.1	89.8	2.79
419.ThaboMofutsanya	94.0	98.3	100.5	2.58
420.Fezile Dabi	84.1	89.6	92.5	2.43
499.Mangaung	112.6	116.6	118.8	2.47
521.Ugu	88.4	95.8	100.3	2.55
522.UMgungundlovu	81.7	88.4	92.5	2.23
523.Uthukela	80.6	89.7	94.8	2.82
527.Umkhanyakude	69.1	83.5	92.0	3.06
528.Uthungulu	97.0	111.3	118.7	2.39
543.Sisonke	95.0	105.1	109.9	3.16
554.Umzinyathi	75.4	85.6	91.4	3.21
555.Amajuba	89.7	100.9	105.0	2.67
556.Zululand	80.8	92.6	100.4	3.20
559.iLembe	41.2	47.0	49.7	2.81
599.eThekwini Metro	87.3	95.1	99.7	2.21
637.Bojanala	69.2	75.9	79.9	2.58
638.NgakaModiriMolema	62.1	69.7	72.6	3.07
639.DrRuthSegomotsi	90.4	100.0	103.9	2.95
640.DrKennethKaunda	87.0	94.6	98.7	2.59
742.Sedibeng	88.0	92.7	95.3	2.45
748.WestRand	87.6	93.0	94.8	2.50
797.Ekurhuleni	78.2	82.9	86.2	2.55
798.CityOfJohannesburg	68.2	72.3	74.6	2.42
799.CityofTshwane	96.6	104.0	108.5	2.31
830.GertSibande	58.6	67.8	71.7	2.65
831.Nkangala	72.5	80.7	84.0	2.51
832.Ehlanzeni	105.6	118.6	123.4	2.63
933.Mopani	99.7	106.5	108.6	3.00
934.Vhembe	82.9	91.6	94.8	3.14
935.Capricorn	97.7	105.6	108.5	3.05
936.Waterberg	100.3	109.2	113.1	2.95
947.GreaterSekhukhu	64.5	71.6	74.6	3.25
National	82.4	89.5	93.2	2.55

Table 7-2 Completeness and TFR results (2016)

District	completeness1	completeness2	completeness5	TFR (2016)
101.WestCoast	62.4	64.9	65.4	2.50
102.Cape Winelands	77.0	78.6	79.0	2.39
103.Overberg	72.1	73.8	74.7	2.04
104.Eden	89.6	91.3	91.9	2.04
105.CentralKaroo	82.0	83.4	84.0	2.48
199.CityOfCapeTown	96.3	98.4	99.1	2.08
210.Cacadu	60.9	63.4	64.6	2.40
212.Amathole	52.5	54.4	55.5	2.53
213.ChrisHani	76.4	79.3	81.3	2.37
214.Ukhahlamba	62.4	65.0	65.9	2.50
215.ORTambo	82.3	85.5	87.7	2.81
244.AlfredNzo	61.6	64.7	66.4	2.97
260.BuffaloCity	114.0	116.5	118.1	2.05
299.NelsonMandelaBay	88.2	90.6	91.8	2.00
306. Namakwa	97.2	98.4	98.8	2.01
307. Pixley_ka_Seme	63.7	65.8	66.6	2.68
308. Siyanda	84.9	86.6	87.4	2.74
309. Frances Baard	122.2	124.8	125.6	2.60
345. John Taolo Gaets	97.7	98.9	99.4	2.74
416.Xhariep	40.6	42.8	44.0	2.66
418. Lejweleputswa	69.6	71.2	71.8	2.63
419.ThaboMofutsanya	72.2	74.0	74.8	2.43
420.Fezile Dabi	69.6	71.9	73.0	2.46
499.Mangaung	115.3	117.0	118.1	2.16
521.Ugu	87.0	91.4	93.1	2.32
522.UMgungundlovu	87.9	92.1	93.4	1.74
523.Uthukela	78.7	83.5	85.0	2.55
527.Umkhanyakude	96.8	104.9	107.9	2.31
528.Uthungulu	105.4	112.8	115.4	2.14
543.Sisonke	92.4	96.7	97.7	2.39
554.Umzinyathi	107.5	115.5	117.5	2.27
555.Amajuba	64.9	68.6	70.0	2.64
556.Zululand	78.2	83.7	86.0	2.59
559.iLembe	53.7	58.6	60.5	2.52
599.eThekwini Metro	111.0	117.3	120.3	1.71
637.Bojanala	78.9	81.3	81.9	2.58
638.NgakaModiriMolema	50.0	51.1	51.8	2.67
639.DrRuthSegomotsi	79.7	82.7	83.4	2.98
640.DrKennethKaunda	83.8	85.7	86.5	2.53
742.Sedibeng	86.7	88.7	89.8	2.23
748.WestRand	69.6	72.2	72.9	2.22
797.Ekurhuleni	90.6	93.2	94.5	2.11
798.CityOfJohannesburg	82.1	83.9	85.0	2.10
799.CityofTshwane	97.7	101.0	102.4	2.16
830.GertSibande	66.5	69.8	71.1	2.35
831.Nkangala	74.8	77.8	78.3	2.26
832.Ehlanzeni	106.9	109.7	110.6	2.51
933.Mopani	78.7	80.2	80.7	2.89
934.Vhembe	89.0	91.6	92.7	2.89
935.Capricorn	94.5	97.1	98.1	2.70
936.Waterberg	100.3	104.1	105.4	2.88
947.GreaterSekhukhu	74.6	76.5	77.2	2.91
National	85.4	88.4	89.6	2.28

Table 7-3 Percentage points change of completeness from 2011 to 2016.

District	completeness1	completeness2	completeness5
101.WestCoast	-4	-7	-9
102.Cape Winelands	-2	-4	-6
103.Overberg	22	19	18
104.Eden	17	15	14
105.CentralKaroo	24	22	22
199.CityOfCapeTown	13	11	9
210.Cacadu	-14	-18	-21
212.Amathole	-12	-17	-21
213.ChrisHani	-13	-18	-21
214.Ukhahlamba	-1	-6	-9
215.ORTambo	-8	-14	-18
244.AlfredNzo	-1	-5	-8
260.BuffaloCity	-10	-14	-15
299.NelsonMandelaBay	2	-2	-4
306. Namakwa	26	25	24
307. Pixley_ka_Seme	-21	-27	-29
308. Siyanda	6	3	2
309. Frances Baard	19	18	16
345. John Taolo Gaets	-5	-7	-9
416.Xhariep	-5	-6	-8
418. Lejweleputswa	-13	-16	-18
419.ThaboMofutsanya	-22	-24	-26
420.Fezile Dabi	-14	-18	-20
499.Mangaung	3	0	-1
521.Ugu	-1	-4	-7
522.UMgungundlovu	6	4	1
523.Uthukela	-2	-6	-10
527.Umkhanyakude	28	21	16
528.Uthungulu	8	2	-3
543.Sisonke	-3	-8	-12
554.Umzinyathi	32	30	26
555.Amajuba	-25	-32	-35
556.Zululand	-3	-9	-14
559.iLembe	13	12	11
599.eThekwini Metro	24	22	21
637.Bojanala	10	5	2
638.NgakaModiriMolema	-12	-19	-21
639.DrRuthSegomotsi	-11	-17	-21
640.DrKennethKaunda	-3	-9	-12
742.Sedibeng	-1	-4	-6
748.WestRand	-18	-21	-22
797.Ekurhuleni	12	10	8
798.CityOfJohannesburg	14	12	10
799.CityofTshwane	1	-3	-6
830.GertSibande	8	2	-1
831.Nkangala	2	-3	-6
832.Ehlanzeni	1	-9	-13
933.Mopani	-21	-26	-28
934.Vhembe	6	0	-2
935.Capricorn	-3	-8	-10
936.Waterberg	0	-5	-8
947.GreaterSekhukhu	10	5	3
National	3	-1	-4

Table 7-4 Differences between *xTFR+* and Gompertz model (2011)

District	*xTFR+* (2011)	Gompertz (2011)	Difference (%)
212.Amathole	4.11	2.48	65.6%
528.Uthungulu	3.68	2.39	54.0%
521.Ugu	3.73	2.55	46.3%
639.DrRuthSegomotsi	4.22	2.95	42.8%
213.ChrisHani	3.96	2.80	41.1%
416.Xhariep	3.75	2.68	39.8%
523.Uthukela	3.91	2.82	38.8%
215.ORTambo	4.28	3.22	32.6%
244.AlfredNzo	4.42	3.42	29.2%
556.Zululand	4.13	3.20	28.9%
543.Sisonke	4.07	3.16	28.6%
527.Umkhanyakude	3.94	3.06	28.6%
554.Umzinyathi	4.10	3.21	27.8%
214.Ukhahlamba	3.83	3.10	23.6%
798.CityOfJohannesburg	1.88	2.42	22.5%
832.Ehlanzeni	3.17	2.63	20.6%
522.UMgungundlovu	2.66	2.23	19.3%
309. Frances Baard	3.03	2.54	19.1%
307. Pixley_ka_Seme	3.33	2.81	18.5%
345. John Taolo Gaets	3.28	2.81	17.0%
210.Cacadu	2.79	2.39	17.0%
638.NgakaModiriMolema	3.55	3.07	15.7%
559.iLembe	3.24	2.81	15.5%
830.GertSibande	3.06	2.65	15.3%
419.ThaboMofutsanya	2.96	2.58	14.9%
420.Fezile Dabi	2.77	2.43	14.0%
555.Amajuba	3.04	2.67	13.9%
797.Ekurhuleni	2.24	2.55	12.3%
637.Bojanala	2.89	2.58	11.7%
748.WestRand	2.22	2.50	11.5%
799.CityofTshwane	2.05	2.31	11.5%
933.Mopani	3.30	3.00	9.9%
935.Capricorn	3.34	3.05	9.4%
306. Namakwa	2.53	2.35	8.0%
199.CityOfCapeTown	2.20	2.39	7.8%
308. Siyanda	2.55	2.76	7.6%
260.BuffaloCity	2.43	2.28	6.6%
640.DrKennethKaunda	2.75	2.59	6.4%
599.eThekwini Metro	2.09	2.21	5.7%
499.Mangaung	2.60	2.47	5.0%
831.Nkangala	2.64	2.51	4.9%
947.GreaterSekhukhu	3.40	3.25	4.7%
103.Overberg	2.55	2.45	4.2%
105.CentralKaroo	3.34	3.22	3.9%
101.WestCoast	2.47	2.52	1.8%
104.Eden	2.61	2.58	1.3%
936.Waterberg	2.99	2.95	1.3%
299.NelsonMandelaBay	2.26	2.29	1.2%
102.Cape Winelands	2.38	2.41	1.0%
742.Sedibeng	2.43	2.45	1.0%
934.Vhembe	3.12	3.14	0.7%
418. Lejweleputswa	2.80	2.79	0.1%

Table 7-5 Differences between *xTFR+* and Gompertz model (2016)

District	*xTFR+* (2016)	Gompertz (2016)	Difference (%)
213.ChrisHani	3.74	2.37	58.2%
554.Umzinyathi	3.46	2.27	52.2%
528.Uthungulu	3.19	2.14	49.0%
212.Amathole	3.75	2.53	48.1%
527.Umkhanyakude	3.41	2.31	47.3%
543.Sisonke	3.48	2.39	45.8%
214.Ukhahlamba	3.55	2.50	41.8%
522.UMgungundlovu	2.43	1.74	39.4%
215.ORTambo	3.91	2.81	39.2%
523.Uthukela	3.53	2.55	38.6%
556.Zululand	3.58	2.59	37.9%
521.Ugu	3.14	2.32	35.1%
244.AlfredNzo	3.96	2.97	33.3%
416.Xhariep	3.36	2.66	26.3%
799.CityofTshwane	1.73	2.16	20.1%
798.CityOfJohannesburg	1.68	2.10	19.9%
308. Siyanda	2.20	2.74	19.6%
832.Ehlanzeni	3.00	2.51	19.3%
105.CentralKaroo	2.95	2.48	19.0%
935.Capricorn	3.17	2.70	17.4%
599.eThekwini Metro	1.97	1.71	15.6%
639.DrRuthSegomotsi	3.43	2.98	15.2%
638.NgakaModiriMolema	3.06	2.67	14.7%
104.Eden	2.32	2.04	13.7%
260.BuffaloCity	2.32	2.05	13.7%
103.Overberg	2.28	2.04	12.0%
102.Cape Winelands	2.10	2.39	11.9%
559.iLembe	2.80	2.52	11.4%
499.Mangaung	2.37	2.16	10.0%
555.Amajuba	2.91	2.64	10.0%
299.NelsonMandelaBay	2.20	2.00	9.7%
101.WestCoast	2.26	2.50	9.6%
419.ThaboMofutsanya	2.64	2.43	8.7%
947.GreaterSekhukhu	3.13	2.91	7.4%
307. Pixley_ka_Seme	2.88	2.68	7.4%
418. Lejweleputswa	2.45	2.63	6.9%
748.WestRand	2.07	2.22	6.9%
199.CityOfCapeTown	1.96	2.08	5.9%
210.Cacadu	2.53	2.40	5.4%
797.Ekurhuleni	2.01	2.11	4.7%
742.Sedibeng	2.13	2.23	4.6%
309. Frances Baard	2.71	2.60	4.4%
934.Vhembe	3.02	2.89	4.4%
345. John Taolo Gaets	2.86	2.74	4.3%
306. Namakwa	2.09	2.01	3.9%
933.Mopani	2.99	2.89	3.5%
830.GertSibande	2.42	2.35	3.1%
640.DrKennethKaunda	2.46	2.53	2.8%
637.Bojanala	2.53	2.58	2.1%
831.Nkangala	2.22	2.26	1.6%
936.Waterberg	2.91	2.88	1.2%
420.Fezile Dabi	2.45	2.46	0.4%

Table 7-6 Provincial percentage differences between xTFR and Gompertz (2011 & 2016)

Province	*xTFR+* (2011)	Gompertz (2011)	Difference (%)	*xTFR+* (2016)	Gompertz (2016)	Difference (%)
Western Cape	2.30	2.68	14.0%	1.98	2.14	7.4%
Eastern Cape	3.46	2.65	30.6%	3.13	2.43	28.8%
Northern Cape	3.09	2.75	12.5%	2.61	2.56	1.9%
Free State	2.82	2.48	13.8%	2.42	2.44	0.8%
KwaZulu-Natal	3.07	2.52	21.7%	2.82	2.11	33.3%
North West	2.87	2.76	4.0%	2.67	2.57	4.0%
Gauteng	1.96	2.46	20.3%	1.82	2.13	14.5%
Mpumalanga	2.96	2.61	13.3%	2.49	2.37	5.0%
Limpopo	3.44	3.06	12.4%	3.14	2.88	9.2%
South Africa	2.74	2.55	7.2%	2.44	2.55	4.6%

Table 7-7 District sex ratio for 12 months before 2011 Census & 2016 CS

District	Sex Ratio 2011	Sex Ratio 2016
101. West Coast	103	103
102. Cape Winelands	102	102
103. Overberg	107	101
104. Eden	103	105
105. Central Karoo	113	95
199. City of Cape Tow	102	102
210. Cacadu	107	104
212. Amathole	99	99
213. Chris Hani	100	103
214. Ukhahlamba	101	101
215. O.R.Tambo	101	101
244. Alfred Nzo	98	101
260. Buffalo City	102	100
299. Nelson Mandela B	103	102
306. Namakwa	107	102
307. Pixley ka Seme	96	107
308. Siyanda	102	103
309. Frances Baard	104	101
345. John Taolo Gaets	106	104
416. Xhariep	101	91
418. Lejweleputswa	101	102
419. Thabo Mofutsanya	100	103
420. Fezile Dabi	99	99
499. Mangaung	100	104
521. Ugu	100	103
522. UMgungundlovu	102	100
523. Uthukela	103	101
527. Umkhanyakude	102	101
528. Uthungulu	99	102
543. Sisonke	100	100
554. Umzinyathi	99	101
555. Amajuba	102	99

556. Zululand	95	104
559. iLembe	102	97
599. eThekwini Metrop	102	103
637. Bojanala	102	100
638. Ngaka Modiri Mol	100	101
639. Dr Ruth Segomots	102	101
640. Dr Kenneth Kaund	103	103
742. Sedibeng	101	104
748. West Rand	103	105
797. Ekurhuleni	100	101
798. City of Johannes	102	100
799. City of Tshwane	102	102
830. Gert Sibande	100	99
831. Nkangala	100	101
832. Ehlanzeni	103	100
933. Mopani	102	103
934. Vhembe	102	102
935. Capricorn	103	100
936. Waterberg	100	101
947. Greater Sekhukhu	105	104

Annexure 2: GRAPHS

Figure 7-1 Percentage differences between *xTFR+* and Gompertz (2011)

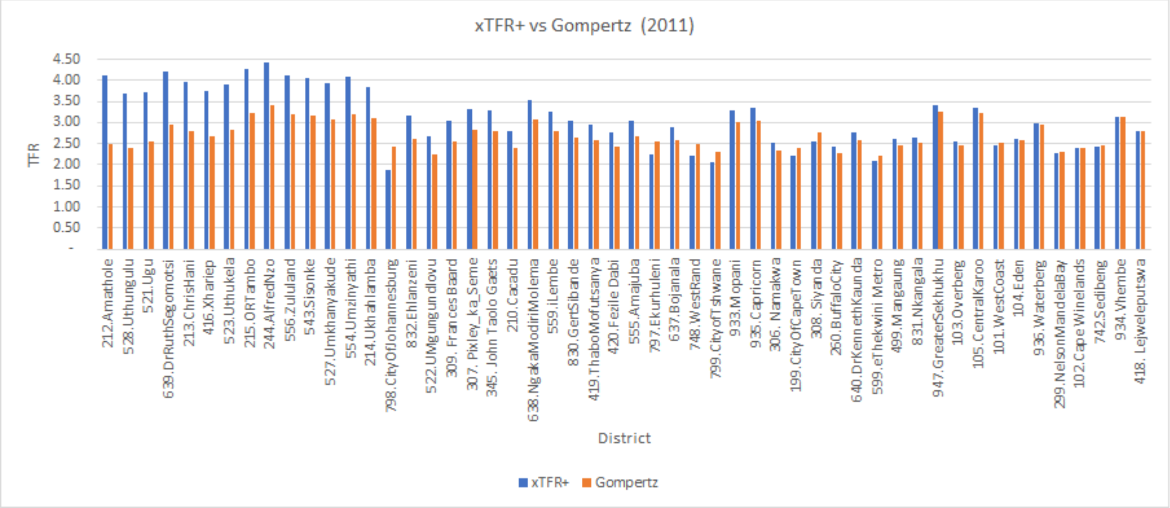


Figure 7-2 Percentage differences between *xTFR+* and Gompertz (2016)

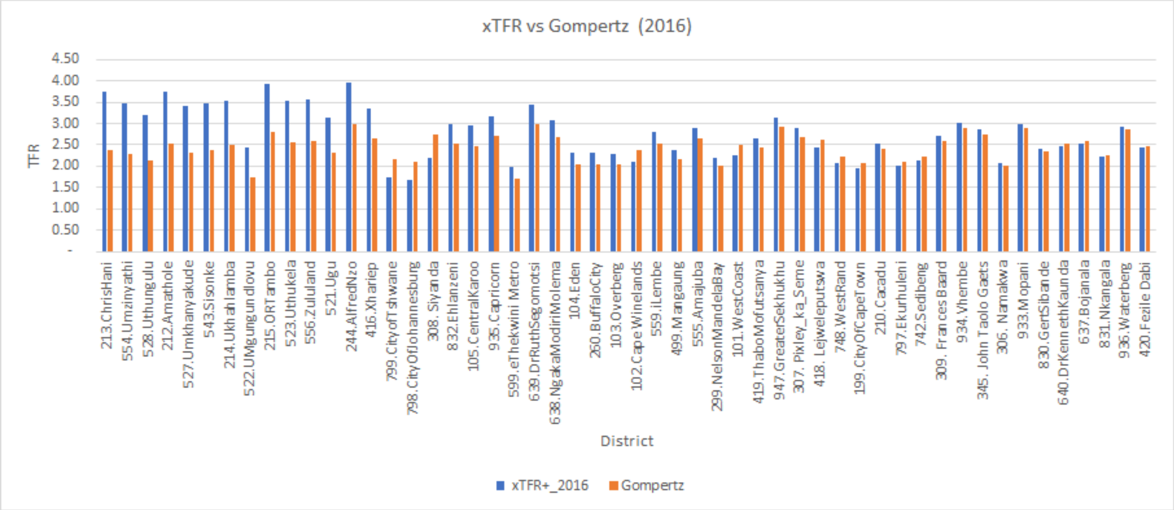


Figure 7-3 District sex ratio for 12 months before the 2011 Census & 2016CS respectively