

**UNIVERSITY OF KWAZULU-NATAL**

**AN ASSESSMENT OF RETROSPECTIVE BIRTH  
HISTORY REPORTING FOR THE MEASUREMENT OF  
FERTILITY IN SOUTH AFRICA**

**NELOUFAR KHAN**

**2010**

## DECLARATION

Submitted in partial fulfillment of the requirements for the degree of Masters in Population Studies, in the Graduate Programme in

The School of Development Studies, University of KwaZulu-Natal,  
South Africa.

I declare that this dissertation is my own unaided work. All citations, references and borrowed ideas have been duly acknowledged. I confirm that an external editor was not used. It is being submitted for the degree of Masters in Population Studies, in the Faculty of Humanities, Development and Social Science, University of KwaZulu-Natal, South Africa. None of the present work has been submitted previously for any degree or examination in any other University.

---

Neloufar Khan

2010

## ***Abstract***

Fertility is one of the major tenets of demography. Its importance lies in the determination of fertility trends in a country, in a specific time period. These statistical inferences of fertility play an imperative role in population policy formation and planning. Thus the importance of the measurement of fertility remains undisputed.

Due to the significance of fertility, its measurement and its profound impact on societies, acknowledging and addressing the quality of fertility data is of great importance. This research study was conceived in response to the above concern. This study aims at addressing and providing insight into birth history data irregularities and determining interventions of working with this issue in the context of South Africa.

Through secondary analysis (i.e. descriptive exploratory and comparative analysis) the study sought to firstly establish a demographic profile of women associated with inconsistent and inaccurate reporting of their birth histories. Secondly the research attempted to ascertain a relationship between the socio-economic statuses of individuals and retrospective reporting. A third objective was to note the sex-selectiveness of reporting (i.e. were more girls or boys reported or misreported on in the retrospective birth histories).

The study has established that older, married women with some educational attainment, of rural areas from either the middle and lower income categories tend to misreport more frequently than their converse counterparts. Furthermore, a plausible relationship between the socio-economic statuses of individuals was observed. In terms of the sex-selectiveness of reporting, in general, boys were reported on more consistently than girls. However in certain cases, it was found that rural and middle income women reported accurately on girl children born alive and dead girl children.

Recommendations made with respect to improve the quality of fertility data for include the proper training of enumerators and data capturers, quality control during data collection, testing of questionnaires, dealing with social, cultural and language barriers and the reinforcement of publicity campaigns for censuses and surveys.

## **Table of Contents**

<b>Declaration</b> .....	<b>ii</b>
<b>Abstract</b> .....	<b>iii</b>
<b>Table of Contents</b> .....	<b>iv</b>
<b>Acronyms and Abbreviations</b> .....	<b>viii</b>
.....	
<b>Chapter 1: Introduction</b> .....	<b>1</b>
1.0 Introduction.....	1
1.1 Background to the Research .....	2
1.2 Statement of the Research Problem .....	3
1.3 Importance, Motivation and Aim of the Study .....	4
1.4 Scope of the Study.....	5
1.5 Research Problem and Objectives: Broader Issues to be Investigated .....	5
1.6 Research Problem and Objectives: Key Questions to be Asked.....	5
1.7 Research Design and Methodology.....	5
1.8 Theoretical Framework for the Study .....	6
1.9 Limitations to the Study .....	9
1.10 Organization of the Dissertation.....	10
.....	
<b>Chapter 2: Literature Review</b> .....	<b>11</b>
2.0 Introduction .....	11
2.1 Birth Histories: An Explanation .....	11
2.2 Irregularities in Retrospective Data .....	12
2.3 Process of Retrospective Reporting.....	13
2.4 Evidence of Response Errors from the Literature.....	15
2.4.1 Time and Recall Ability.....	15
2.4.2 Saliency .....	17
2.4.3 Telescoping .....	19
2.5 Response Errors Caused By Survey Instruments.....	20
2.5.1 Questionnaires as a Source of Measurement Error .....	20
2.5.2 Enumerators as a Source of Measurement Error .....	22
2.5.3 Timing Errors in Questionnaires .....	22
2.6 Socio-Cultural and Political Context of Reporting.....	23
2.7 Respondent Socio-Economic Characteristics .....	26
2.7.1 Education .....	27
2.7.2 Ethnicity (Population Group) .....	27
2.7.3 Age of Respondent .....	28
2.7.4 Place of Residence .....	29
2.8 Measurement and the Collection of Fertility Data .....	30
2.9 Collection of Fertility Data in South Africa.....	32
2.9.1 The 1996 Census.....	33
2.9.2 The 2001 Census.....	43
2.10 Summary .....	37

<b>Chapter 3: Research Design and Data Analysis .....</b>	<b>38</b>
3.0 Introduction .....	38
3.1 Brief History and Socio-Demographic Profile of South Africa.....	38
3.2 Context of the Research.....	40
3.3 Research Design .....	40
3.4 Questionnaire Design.....	41
3.5 Sampling Frame.....	42
3.6 Data Analysis Procedures.....	43
3.7 Method of Analysis .....	44
3.7.1 Grouping of Variables .....	44
3.7.2 Stages of Analysis.....	45
3.8 Summary .....	46
<b>Chapter 4: Results of the Study.....</b>	<b>47</b>
4.0 Introduction .....	47
4.1 Analysis of Background Characteristics for Sample One and Sample Two of Women.....	47
4.2 Descriptive Exploratory Analysis.....	50
4.2.1 Part One: Establishing a Demographic Profile of Women with Inconsistent Reporting ....	51
4.3 Summary of Part One.....	65
4.4 Part Two: Establishing a Link between the Socio-Economic Status of Individuals and Retrospective Reporting of Birth Histories .....	66
4.4.1 Theoretical Expectations for Socio-Economic Status of Individuals and Retrospective Reporting: Living Standard Measure .....	66
4.4.2 Theoretical Expectations for Socio-Economic Status of Individuals and Retrospective Reporting: Geographical Location .....	67
4.4.3 Comparison of Records for Dead Children and Socio-Economic Status for Sample One and Sample Two of Women .....	74
4.5 The Link Between Socio-Economic Status and the Sex-Selectiveness of Reporting.....	81
4.6 Summary .....	83
<b>Chapter 5: Discussion of Results .....</b>	<b>84</b>
5.0 Introduction .....	84
5.1 Findings of the Study.....	84
5.2 Other Explanations.....	89
5.3 Summary .....	90
<b>Chapter 6: Conclusion .....</b>	<b>91</b>

## **List of Tables and Figures**

### **Tables**

#### **Chapter 2: Literature Review**

2.1 Census 1996 Fertility Questions.....	33
2.2 Census 2001 Fertility Questions.....	34

#### **Chapter 4: Results and Discussion**

4.1 Background Characteristics of Women Aged 15-49 – Sample One and Two of Women ....	47-48
4.2 Inconsistencies Observed for Women with Live Births by Background Characteristic .....	52
4.3 Inconsistencies Observed for Women with Children Still Alive and Living In and Not Living in the Household by Background Characteristic.....	54
4.4 Inconsistencies Observed for Women with Children Dead by Background Characteristic .....	55

### **Figures**

#### **Chapter 1: Introduction**

1.1 Event Location Distributions for the Reported Dates of the First Two Births .....	8
---	---

#### **Chapter 2: Literature Review**

2.1 Forgettive Curve .....	16
----------------------------	----

#### **Chapter 3: Research Design and Data Analysis**

3.1 Map of South Africa .....	38
3.2 Percentage Distribution of Households for Data Collected by Sample and Across Provinces ..	43
3.3 Percentage Distribution of Women Used in Analysis by Sample and Across Provinces.....	43

#### **Chapter 4: Results of the Study**

4.1 Comparison of Records for Live Births for Sample One and Sample Two of Women - National Trend .....	57
4.2 Comparison of Records for Boy Children Born Alive for Sample One and Sample Two of Women - National Trend .....	58
4.3 Comparison of Records for Girl Children Born Alive for Sample One and Sample Two of Women - National Trend .....	59
4.4 Comparison of Records for Children Still Alive for Sample One and Sample Two of Women - National Trend .....	60
4.5 Comparison of Records for Boy Children Still Alive for Sample One and Sample Two of Women - National Trend .....	61
4.6 Comparison of Records for Girl Children Still Alive for Sample One and Sample Two of Women - National Trend .....	62
4.7 Comparison of Records for Children Dead for Sample One and Sample Two of Women- National Trend .....	63
4.8 Comparison of Records for Boy Children Dead for Sample One and Sample Two of Women - National Trend .....	64
4.9 Comparison of Records for Girl Children Dead for Sample One and Sample Two of Women-National Trend .....	65
4.10 Comparison of Records for Live Births for Sample One and Sample Two of Women by Living Standard Measure .....	68

4.11 Comparison of Records for Live Births for Sample One and Sample Two of Women by Geographical Type .....	69
4.12 Comparison of Records for Boy Children Born Alive for Sample One and Sample Two of Women by Living Standard Measure .....	70
4.13 Comparison of Records for Boy Children Born Alive for Sample One and Sample Two of Women by Geographical Type .....	71
4.14 Comparison of Records for Girl Children Born Alive for Sample One and Sample Two of Women by Living Standard Measure .....	73
4.15 Comparison of Records for Girl Children Born Alive for Sample One and Sample Two of Women by Geographical Type .....	74
4.16 Comparison of Records for Dead Children for Sample One and Sample Two of Women by Living Standard Measure .....	75
4.17 Comparison of Records for Dead Children for Sample One and Sample Two of Women by Geographical Type .....	76
4.18 Comparison of Records for Boy Children Dead for Sample One and Sample Two of Women by Living Standard Measure .....	77
4.19 Comparison of Records for Boy Children Dead for Sample One and Sample Two of Women by Geographical Type .....	78
4.20 Comparison of Records for Girl Children Dead for Sample One and Sample Two of Women by Living Standard Measure .....	79
4.21 Comparison of Records for Girl Children Dead for Sample One and Sample Two of Women by Geographical Type .....	80

**List of Appendices**

Appendix A: Questionnaires Administered in the Survey .....	96
Appendix B: Fertility Questionnaire Version One .....	97
Appendix C: Fertility Questionnaire Version Two .....	98
Appendix D: Female Questionnaire: Version One and Version Two (Roster) .....	99
Appendix E: Motivation for the Usage of the Living Standard Measure .....	102

<b>References.....</b>	<b>104</b>
------------------------	------------

## **Acronyms and Abbreviations**

AIDS	Acquired Immuno Deficiency Syndrome
AMPS	All Media and Products Survey
ASFR	Age Specific Fertility Rates
DHS	Demographic and Health Survey
EA	Enumerator Area
HIV	Human Immunodeficiency Virus
LSM	Living Standard Measure
MFLS	Malaysian Family Life Survey
N	Sample Size
PRB	Population Reference Bureau
PSU	Primary Sampling Unit
SAARF	South African Advertising Research Foundation
TFR	Total Fertility Rate
WFS	World Fertility Survey



## CHAPTER ONE Introduction

\*\*\*\*\*

### **1.0 Introduction**

In their publication titled *Population: A Lively Introduction*, the Population Reference Bureau (PRB) (2007) opens with an interesting statement which reads as follows: “Most people think demography is just math in disguise— a sort of dry social accounting” (McFalls, 2007:3).

Individuals fail to realize the profound impact that demographic forces have on their daily lives. People view the field as the mere counting of people. Demography is light years more than that. It is one of the very many dynamic and challenging fields of academia. Demography is a fascinating subject to be a part of as its reach extends to many personally relevant topics, such as birth, marriage, childbearing, retirement and death (PRB, 2007). Through the interplay of three demographic processes (fertility, mortality and migration) the demographic landscape of a population changes and evolves creating a new set of demographic forces that exerts its influence on the population which in turn sparks demographic change and so the process continues

The interest of the study lies with the demographic process of fertility. According to Shyrock and Siegal (1976) fertility is defined as the number of live births that occur to an individual or to a population. Fertility is an absorbing sub-field of demography. It would be considered an insult to simply think of fertility as just the bearing of children by women. The keenness associated with fertility is generated through the fact that fertility mainly concerns itself with women. Decisions made by women are influenced by their individual complexities and those that surround them. These influences impact on their decisions about fertility as well.

Researching these decisions and uncovering these complexities (whether social, cultural or political) is pivotal in understanding what makes a population tick. Below standard quality of data quality can greatly impact on the interest in such areas of research. The discussion to follow elaborates on this aspect.

## **1.1 Background to the Research**

It is a well known fact that fertility is one of the major tenets of demography. Its importance lies in the determination of fertility trends in a country, in a specific time period. These statistical inferences of fertility play an imperative role in population policy formation. Thus the importance of the measurement of fertility remains undisputed. This scenario however, is only attainable if the data collected is of good quality. Often censuses and surveys are conducted without adequately testing for quality control with regards to the questions to be asked. This sadly impacts on the data collected, results produced and most importantly, user confidentiality.

The United Nations (2006) document titled the *Principles and Recommendations for Population and Housing Censuses* stresses the importance and usefulness of testing various aspects of a census plan prior to the anticipated enumeration period. It emphasizes greatly the suitability of intended census questions, testing the formulation as well as the instructions provided. Locally, lessons learnt during the last two censuses conducted in 1996 and 2001 have emphasized the urgent need for intensive research to inform the development of data collection tools, i.e. data items, questions to be asked and the format of the questionnaire (Statistics South Africa, 2007). This process requires constant engagement with data providers and users (Statistics South Africa, 2007).

In 2002 Statistics South Africa established a permanent structure dedicated to further developing methodologies pertaining to the execution of population censuses. The component *Census Research and Methodology* is tasked with the responsibility of identifying and filling the gaps regarding pertinent census research issues. The structure was formulated with the purpose of co-coordinating and implementing strategic plans of action for future population censuses. As a result, it has developed a comprehensive research programme which aims at further developing methodologies regarding the execution of future population censuses. It is in this context that content research<sup>1</sup> on issues pertaining to fertility was undertaken and which gives impetus to my study.

---

<sup>1</sup> Content research is the continuous review of census methodologies and procedures with the aim of ensuring cost effectiveness and quality. In other words, content research unpacks the topics to be covered in the census.

## **1.2 Statement of the Research Problem**

The *Final Data Handover Report of Project: Census 2001 Data Processing* (Statistics South Africa, 2003) highlighted several issues with regard to census content that required research. Amongst them was fertility. The method of collecting data on fertility in Census 2001 was conducted via retrospective birth histories. In their report, Moultrie and Dorrington (2004) identified the following two major concerns with regard to the fertility data that was collected in Census 2001. The first concern voiced was that all births that occurred in the preceding twelve months were not captured accurately and secondly the data on the number of children born to women of reproductive age were deficient (Moultrie and Dorrington, 2004).

To elaborate, in the Census 2001, there was a deficiency of dead children. This was due to women of childbearing ages misunderstanding the questions posed to them during enumeration. As noted by Moultrie and Dorrington (2004), data on the proportion of children ever born, indicated that women interpreted the question on *children ever born* as asking about *children still alive*. Consequently, data on the number of children born to women of reproductive ages were seriously deficient (Moultrie and Dorrington, 2004). Hence proportions of dead children were severely undercounted.

The above irregularities stem from a range of factors, such as retrospective data recall by women of reproductive ages, proxy responses given by a member (typically the head) of the household in the absence of woman in question, inadequate enumerator training and publicity of the census.

Following on from the acknowledgement of the significance of fertility, its measurement and its profound impact on societies, there is an urgent gap in this chain that needs to be filled. The concern at this stage is whether information on fertility can be obtained accurately and, if not, what are the consequences for the measurement of fertility? A logical assumption that is therefore made is that barriers cannot be overcome unless they are identified. Thus justification for this research is found in the need to identify the factors relating to fertility data irregularities. The question that remains, is *how*?

### **1.3 Importance, Motivation and Aim of the Study**

The momentum that drives the research is as a result of the content research activities conducted at the Census Research and Methodology Component at Statistics South Africa. With regards to census content research on fertility, their primary aim is to examine the benefits of testing an alternative fertility schedule of questions and evaluate its outcomes. My study feeds on this by taking the analysis one step further.

The study will focus on the quality of reporting between the Census 2001 fertility schedule of questions and an alternative fertility schedule of questions.<sup>2</sup> Whilst the aim of the study is *not* to determine which schedule of fertility questions is a better option for the collection of fertility data, the study will however draw comparisons between the two schedules of questions with respect to the quality of data reporting.

Against the background of the statement of the research problem, the primary objective of the study is to investigate and tease out the factors associated with data irregularities observed in retrospective birth histories. In other words, *who* is misreporting and *what* are they misreporting. By this, the study aims to highlight a profile of women that are associated with inconsistent reporting as well as what these inconsistencies are. A secondary objective of the study will then be to recommend appropriate strategic plans of action in terms of ways to combat these data irregularities for future surveys and censuses.

In general, approaches in the past have focused on 'snapshot' data in order to make their conclusions. In other words, the data derived is a singular view of the phenomena at a particular instance in time. As a result, the conclusions arrived at are applicable to that singular analysis. The strength in my approach however lies in its comparative nature. In other words, the study will identify and evaluate similarities and differences between the two schedules and thus derive conclusions regarding birth history irregularities. A benefit for such an approach is foremost attaining a broader perspective of the issue at hand. Secondly, comparative analysis allows for more in-depth investigation

---

<sup>2</sup> The alternative fertility schedule of questions replicates some of the Census 2001 fertility schedule of questions except for a few additional questions. The inclusion of these additional questions is based on the recommendations made by the United Nations for population and housing censuses.

#### **1.4 Scope of the Study**

This study aims at addressing and providing insight into birth history data irregularities and determining interventions of working with this issue in the context of South Africa. The results of the study as well as that of the literature based search will hopefully assist in putting forward the need to pay attention to data quality issues of this nature. It is with this that the study envisages that this piece of research be of interest and simulate awareness between and among academics, research organizations and other stakeholders.

#### **1.5 Research Problems and Objectives: Broader Issues to be Investigated**

The aim of the study is two folded. The study seeks to:

- Ascertain the relationship between socio-economic statuses of individuals and retrospective birth reporting.
- Outlining intervention strategies in terms of enumerator training and publicity for the census. In other words, by knowing what kinds of socio-economic statuses are associated with certain kinds of reporting, how this information sheds some light on the different aspects of training and publicity.

#### **1.6 Research Problem and Objectives: Key Questions to be Asked**

With respect to the Census 2001 fertility schedule of questions and the alternative fertility schedule of questions:

- For those women whose birth histories have irregularities, what are their individual characteristics?
- Is there a link between socio-economic status of individuals and retrospective birth reporting? That is, who are the individuals who are likely to omit and misreport births?
- Is there a link between socio-economic status of individuals and the aspects of the birth that are misreported? For example sex selectiveness of births and deaths that is misreported or omitted?

#### **1.7 Research Design and Methodology**

##### **1.7.1 Conception of the Study**

The *Final Data Handover Report of Project: Census 2001 Data Processing* recommended that in light of the problems experienced with regard to the quality of the fertility data, Statistics South

Africa should explore the viability of asking extensive questions on fertility for the next census (Statistics South Africa, 2003). This can be achieved if research is conducted to determine the best way in which to ask questions on fertility.

Following on from this, in November 2005, a census content research study was conducted by Statistics South Africa where an alternative schedule of fertility questions for measuring fertility was tested. The Census 2001 fertility schedule of questions was also tested. Data was collected for two samples of women. The Census 2001 fertility schedule and the alternative fertility schedule was administered to each of the respective samples. Accompanying each of the schedules was a roster, in which women had to report on their children in greater detail. This will be explained in greater detail in Chapter three. In brief, the aim of this strategy was to identify discrepancies between the reporting in either fertility schedule and the roster.

### **1.8 Theoretical Framework for the Study**

As discussed in great detail in *Chapter 2: Review of Literature* writers such as Tourangeau (1984) and Sudman *et al* (1996) have documented extensively on the four-step survey question and answer process. As mentioned before, much of the literature of measurement error in retrospective data has originated from its focus on the retrieval stage of the question-answering process. Mathiowetz *et al* (2002: 162) notes that the literature on the subject “classifies the lack of reporting as a retrieval failure on the part of the respondent comparing the characteristics of the events that are reported to those that are not reported”. Common to the majority of the literature is the issue of the length of the recall period. The basic premise is this: the greater the length of the recall period, the greater the expected bias due to respondent retrieval and reporting error.

The study has adopted the above as a predominant theoretical standpoint. I have expanded on this by linking it to the *Event-Placement Model* as developed by J.E Potter in 1977. In general, the model describes how a woman might misplace events during a period of time when responding to questions posed to her in a birth history questionnaire.

### **1.8.1 Background to the Event-Placement Model**

As noted by Potter (1977: 341) the model is based on two main propositions. They are as follows (Potter, 1977: 341):

- “The date of an event is recalled less accurately the longer ago the event took place” and;
- “If a birth history is elicited by asking questions about live births in the order in which they occurred, then the date a woman attaches to any event other than the first is influenced by the information she has already given about the previous event”

To explain the model, Potter (1977) uses the scenario where a respondent is asked to report on the date of each live birth. He acknowledges that in cases of fertility data collection, the respondent is also asked to report on the present age of surviving children. In addition, although the date-of-birth question may in most instances be asked first, which question of the two dominates field experience is not always clear.

Furthermore, Potter (1977) states that birth histories in a questionnaire also ask of the status of surviving children. These set of questions may appear first, followed by questions on deceased children, in some cases questions about abortions and of pregnancies not resulting in a live birth (miscarriages). Potter (1977:342) notes that that in conceptualizing the model it is of importance to note that “revisions or modifications of the birth history may be made, but such changes are presumed to be of secondary importance and are not represented in the model”.

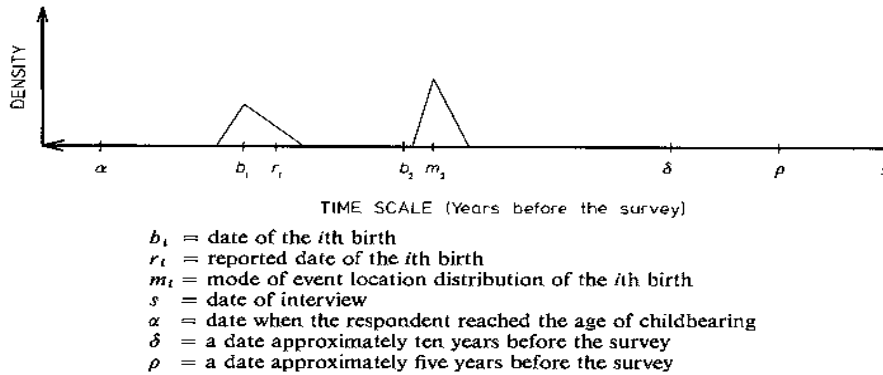
In summation, Potter (1977) notes that the Event-Misplacement Model should be equally applicable to questionnaires whether the first set of questions asks of the status of surviving children or of all live births.

### **1.8.2 Specifics of the Model**

Potter (1977) states that the first step that is required to construct the Event-Misplacement Model (Figure 1.1: Potter [1977:342]) is to define a time scale on which the point corresponding to the date of the interview is zero and the years before the survey are noted and measured positively (Potter, 1977). Secondly, an assumption is made where all the respondents’ live births are recalled stochastically along the proposed time scale (Potter, 1977). In other words,

respondents recall births in a random manner. The model also pre-supposes that no events are forgotten and that the location distribution of each event is triangular (Potter, 1977).

**Figure 1.1:**  
**Event Location Distributions for the Reported Dates of the First Two Births**



In questionnaires where the initial set of questions are about the first live births as opposed to subsequent live births, Potter (1977) mentions that the date of the first event will be remembered independently. That said, a disclaimer is also noted on the part of Potter (1977:342), where he notes that “if a question relating to the date of marriage is included prior to the birth history, this generalization might not hold. Conceptually, marriage could constitute the first event in the history just as easily as the first birth”. It is assumed in the model that the mode of the event location distribution corresponding to the first birth,  $m_1$ , is the true date of event  $b_1$  (Potter, 1977). The third parameter of distribution  $q$ , is assumed as being constant: thus if  $q_1 < 0.5$ , the event location distribution of the first birth will be skewed toward the date of the interview (Potter, 1977).

A problem that may be encountered when using the model is that if the second event is close to the date of the interview, it might be recalled quite inaccurately (Potter, 1977). According to the model, a narrow location distribution for a recent event could be displaced some distance from that of the true date (Potter, 1977). Simply put, it is reasonable to expect that a mother’s account of the age of the living child that is less than five years of age will be influenced by the child’s true age rather than by a recollection of the interval between adjacent events.

Potter (1977:342) adds that if a more reasonable set of assumptions are adopted, it might then occur that once the respondent has given a date for the earliest event, “subsequent events are placed on the time scale by increasingly accurate recollections of the intervals between adjacent



events, but only so long so as the events actually took place a considerable amount of years before the survey". On the other hand, events that occurred in the recent past are located independently of the dates that are attached to previous events (Potter, 1977). Between these interactions, there is middle ground where the date supplied for the previous events plays an influential role in the location of the event that follows (Potter, 1977). Nevertheless, the dependence (i.e. the mode of the event-location location corresponding to the second live birth,  $m_2$  is equal to  $r_1$  less the true length of the interval) is less than complete (Potter, 1977).

To allow for this, Potter (1977) accommodates for two additional parameters that are included in the model. These are  $\delta$  and  $p$  respectively. The former represents the correspondence to a point in the time scale approximately ten years before the survey and the later corresponds to a point that is located five years before the survey (Potter, 1977). The introduction of these parameters caters for the reduced linearity between  $\delta$  and  $p$  that is caused by dependence of the event location distribution on the date reported for the previous event (Potter, 1977).

Thus in summation, the Event Misplacement Model will be utilized as a frame of theoretical reference for the study. Support for the model has been documented by writers like (Srinivasan and Muthiah, 1987 and Hobcraft and Murphy, 1986). In using the model, the emphasis shall lie in the two main propositions of the model as outlined in Section 1.8.1.

### ***1.9 Limitations of the Study***

Due to the study being constructed around human subjects, ethical considerations were made. With respect to interviewing woman of reproductive ages (15-49), especially younger women (i.e. 15-18<sup>3</sup> years of age) asking of their past fertility had ethical implications. Given this ethical challenge, enumerators were able to proceed with the interview with verbal consent as given by the head of the household prior to the conduction of the interview with the women. Given that interviews with females (15-49 years old) were one-to-one and confidential, enumerators and the research team could learn of children that the head and other members of the household were unaware of.

---

<sup>3</sup> 18 is given as the highest age range for younger women as the South African Constitution defines a as a person under the age of 18 years as a 'child'. In addition 18 is also the age at which schooling (Grade 12) is completed. For the study, household members could learn as a result of the survey enumeration that their daughter (s) has had a child (children) unknown to them whilst still being a minor and at school.

Shortcomings from a literature perspective are that subject under investigation is often not given the recognition it deserves and thus few studies have been conducted *directly* on the subject. As a result some of the literature consulted is not of a recent nature. Hence in some instances the information is somewhat dated. Despite this, the findings of those studies could still prove true. Although we have progressed in time, the socio-economic and cultural institutions that govern societies and ultimately people's lives are still with us today. Thus it is likely that findings of the literature consulted could be valid at present day.

### **1.10 Organization of the Dissertation**

The first chapter serves as an anchor to the research study. The objective of such a structure is to provide the reader with an adequate understanding of the aims of the research. The next chapter aims to provide the reader with a theoretical background of the topic by reviewing existing literature on the subject. Chapter three presents the data and the methodology employed in the analysis and the interpretation of the study. Chapter four outlines the results of the study and provides a brief discussion of the main findings. Chapter five aims to provide an overview of the salient findings of the study, thereafter presenting implications thereof with respect to recommendations and future research.

**CHAPTER TWO**  
**Literature Review**

\*\*\*\*\*

**2.0 Introduction**

The following chapter presents a review of the literature on subjects related to birth histories and retrospective reporting. The review begins with an explanation of birth histories, followed by data irregularities, common response errors and its causes, the social, political and cultural context of reporting and concludes with the measurement of fertility in South Africa.

**2.1 Birth Histories: An Explanation**

Fertility is one of the core demographic processes and is critical in determining population dynamics. Thus, the methodologies that are utilized must cater for this great need. Since fertility is a subject that requires data on events that have occurred over time, censuses and surveys ask retrospective questions. According to Blacker and Brass (1979) retrospective surveys have been used widely in order to determine the key factors of population growth. In regions such as Latin America, Africa and Asia, where vital registration systems are either non-existent or seriously deficient, special inquisitions have to be made to gather information on fertility. Such a method used to elicit information regarding the fertility of women of reproductive ages is that of birth histories.

Birth histories are utilized to retrieve a complete chronological record of the respondent's fertility over time. As noted by Hobcraft and Murphy (1986) the biggest impetus towards the use of retrospective histories came with the implementation of the World Fertility Survey (WFS) in the 1970s and the Demographic and Health Surveys (DHS) in the 1980s. The utilization of birth histories is seen as a primary method of investigation for fertility studies conducted in developing countries and has become an "indispensable source of data studying fertility levels, trends and determinants" (Schoumaker, 2004: 1). Schoumaker (2004) adds that the birth histories are then reconstituted in order to calculate the classic indicators of fertility [i.e. age specific fertility rates (ASFR), the total fertility rate (TFR)] as well as determining fertility trends that can extend over a ten to fifteen year period. When used in conjunction with socio-economic data collected by fertility surveys, birth histories can also be applied for the purpose of explicative analyses of fertility behaviour (Schoumaker, 2004).

Potter (1977) illustrates the principle behind the method: A sample of women are asked about their reproductive history, their age at interview or date of birth, total number of live births ever had, the date of each birth, the survival status (living or deceased) of children ever born. These set of questions are not a generic prescription for the retrieval of fertility data, and thus allowances for variation are made in cases depending on the research objectives specific to the particular survey.

Advantages of this investigative technique are plentiful. Some of the foremost advantages, apart from the wealth of information retrieved, are that information is obtained at a reduced cost and in a timely manner (Beckett, Da Vanzo, Sastry, Panis and Peterson, 2001). Despite this, such a method as noted by Chidambaram and Pullum (1981) is riddled with errors with respect to misplaced events, misreporting, underreporting and omissions of births which often result in unsatisfactory solutions.

## **2.2 Irregularities in Retrospective Data**

Hobcraft and Murphy (1986), state that much of the strictly technical demographic work conducted in the past on birth histories placed great emphasis on the screening and correction of errors. As a result, problems of reporting errors encountered in retrospective data have not always been placed at the fore. In support of this Scott (1975) adds that in the past there have been numerous texts and studies on sampling error and the correction thereof. Consequently *direct* interrogations on survey response errors are lacking.

The receipt of so little attention to this issue is attributed to “the lack of procedure to model such error structures for a technical discussion of such issues” (Hobcraft and Murphy (1986: 4). Hobcraft and Murphy (1986) further argue that the preoccupation of *certain* demographers with the impact of reporting errors has distorted research priorities in the field having resulted in myriad evaluation reports placing emphasis on screening of reporting errors rather than on obtaining the best possible estimates for the core processes in demography.

From those select few studies that have placed importance on the *direct* investigation of data irregularities on birth histories, these studies have revealed that there are two key types of error (i.e. omission and timing in retrospective reporting). At times, they are often difficult to separate

or identify. Singh (1987) adds that since these errors are attributable to the fact that the information derived is from retrospective surveys, this information given by the respondent depends largely on the recall of events and dates in the past. Singh (1987) notes that these errors are exaggerated in populations where there are low levels of modernization. According to Singh (1987: 620) this implies that “knowledge of dates of vital events is of little relevance and where there is no cultural emphasis on knowledge of dates to counterbalance this”.

In addition, Hobcraft and Murphy (1986) report that some events might not be reported at all. Evidence from previous research also highlights that births in the more distant past, typically those occurring more than 20 years ago, are occasionally omitted (Hobcraft and Murphy, 1986). This is heightened even more so when the child in question died soon after birth or early in childhood (Hobcraft and Murphy, 1986). This may be due to cases where the first-born child who did not survive infancy is often omitted when reporting retrospectively in a survey or census (Coale and Banister, 1994). The literature also shows that these omissions are often sex selective, where more girls are featured than boys (Hobcraft and Murphy, 1986).

A third error identified in the literature by writers such as Blacker and Brass (1979) is that of faculty inclusions. Faculty inclusions occur when women report children who are adopted as well as children for whom they are acting or temporary parents (Blacker and Brass, 1979). These may include children of relatives, friends, and children born from her husband by another wife and so on (Blacker and Brass, 1979).

Hobcraft and Murphy (1986: 4) note that these omissions and additions are “potentially dangerous” for investigations of this nature. This comment is made with respect to the possibility of biased estimates of intervals that result from these errors. Furthermore, these errors and omissions are not randomly distributed among the sample of women (i.e. age, parities, level of education etc) under investigation. Consequently, deriving true differentials of these women becomes problematic (Hobcraft and Murphy, 1986).

### ***2.3 Process of Retrospective Reporting***

Researchers such as Torangeau (1984) and Strack and Martin (1987) have identified four sequential steps in the reporting process. Firstly, the respondent interprets the question and

secondly retrieves the information in order to respond to or construct the answer to the question. In the third step, the respondent formulates an answer based on the recalled information from the previous step. In the final step, the respondent places edit checks on the answer derived and decides on how to respond. These retrieval and reporting steps are followed for all types of questions (i.e. questions of past and current events as well as questions asking of opinion or attitude). These steps are seemingly logical and quite effortless. The issue then is where in this sequence does reporting become incorrect and unreliable?

When the matter of recall error is under the spotlight, researchers in the field pay attention to the second step in the reporting process, (i.e. memory retrieval). At this stage, there are errors in memory retrieval. As noted by Sudman, Bradburn and Schwartz (1996) events can be forgotten or rather omitted in three ways. Firstly, events may never actually be processed, encoded and stored in memory. It is also possible that if an event finds its way to storage, the finer details surrounding the event may not be recalled. Sudman *et al.* (1996) illustrates this by noting an example of a pregnancy. Although this is an exclusive and in some instances an episodic event, research conducted on the subject has demonstrated that whilst the pregnancy itself is rarely forgotten, finer details surrounding the pregnancy is often lost.

The second way in which errors are caused in the reporting process is due to retrieval failure. This occurs due to minimal or no recall (by way of talking and remembering etc) of an event that has passed. The last step states that events can be omitted through inaccurate reconstruction of a memory. This occurs when details of a similar event that has happened to another person who has shared their memory with the respondent may override the original memory of the respondent or leave a stronger memory trace on the respondent (Beckett *et al.*, 2001). As a result, the memory becomes distorted.

To re-iterate, Mathiowetz *et al.* (2002) notes that once the respondent has comprehended the question/s posed to them, he or she must then retrieve the relevant information that the question/s ask of. In addition the respondent then has to make a judgment on the information retrieved as to whether it answers the question and if so, the response is communicated. Of equal interest, is the communication of the response to the interviewer. Respondents may hold back on responding and answering truthfully to certain questions particularly questions that are

perceived as personal and sensitive in nature (Gaskell, Wright and O’Muircheartaigh, 2000). As a result, responses are manipulated and are not a true reflection of circumstances, sadly impacting on the results derived.

## **2.4 Evidence of Response Errors from the Literature**

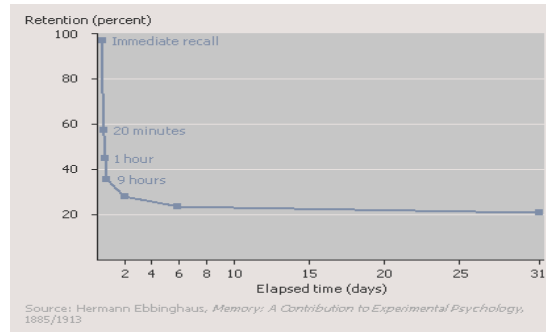
Research conducted on the subject of recall and reporting errors in retrospective surveys have revealed that there are several types of errors that are typical to retrospective surveys. The following discussion sheds some light on the most common of these.

### **2.4.1 Time and Recall Ability**

The first kind of response error is the association between time and recall ability. Beckett *et al.* (2001) makes the point that this is a rather complex association and is not always well understood. One of the foremost studies conducted in the field of memory recall was carried out as early as 1885 by German philosopher Hermann Ebbinghaus. He conducted the study by using himself as a subject. During his work Ebbinghaus identified a negatively accelerating ‘forgetting curve’. As noted in Beckett *et al.* (2001), Ebbinghaus discovered that the rates of forgetting were highest immediately after the event, then fell and leveled off over time. Figure 2.1 illustrates one of Ebbinghaus’s experiments in which he is the subject. In brief he memorized lists of nonsensical syllables, thereafter testing his memory of the syllables at intervals ranging from 20 minutes to 31 days. As illustrated in Figure 2.1, this curve shows that he remembered less than 40 percent of the items after nine hours, but that the rate of forgetting leveled off over time. Beckett *et al.* (2001) supports this claim by noting that the rate of decline in recall ability over time varies by the type of event, for example certain street names versus a personal event such as a birth.

**Figure: 2.1**

**Forgetting Curve (Ebbinghaus, 1885)**



Beckett *et al.* (2001) add that the reporting of errors and incompleteness increases with the length of recall. It is argued that events in the distant past are more likely to suffer from heaping<sup>4</sup>.

Instances of heaping are documented by writers such as Hill and Choi (2006). Hill and Choi (2006) in their research sought to investigate age patterns and trends of early and late neonatal mortality in developing countries. Admittedly this is not where our focus lies, but what does have a relevance to our research is the fact Hill and Choi (2006) retrieved information from their respondents utilizing full birth histories of women aged 15-49 years of age from the DHS.

For the study a heaping index<sup>5</sup> was calculated for regions of the world that showed significant heaping (i.e. where heaping indices were not close to 1). Examples of these regions were found in Africa and the Middle East (Hill and Choi, 2006). Some of the specific countries (DHS survey period noted in brackets) were Ethiopia (2000), Kenya (1998), Malawi (2000), Tanzania (1999), Nigeria (1999), Columbia (2000), Guatemala (1998), Egypt (2000), India (1998), Nepal (2001) and Philippines (1998) amongst others (Hill and Choi, 2006).

From their work, interesting observations are made with regard to the manner in which women report births and deaths. Hill and Choi (2006) suspected that their data was loaded with two

---

<sup>4</sup> Heaping is a specific form of response error that is typical of retrospective data. Here, respondents do not provide the actual estimate but rather a response (rounded) nearest to the actual answer. For example, the correct age of a child may be 22, but the respondent chooses to report the age as 20.

<sup>5</sup> The heaping index was calculated as the number of deaths at seven days divided by one-fifth of the deaths at ages 5 to 9 days. In the absence of error, the heaping index is close to 1.



types of data error. The first of which is the omission of children that die within the first few days of life and secondly heaping.

On the second error, upon the assessment of the quality of the data, heaping by women was highly prevalent. It was found that women tended to report the age at death of their babies at age 7 days when in fact the death occurred earlier than 7 days. The demographic implication of this is serious and highly distorting. According to Hill and Choi (2006:432) “a death at less than 7 days removes a death from the early neonatal category which adds it to the late neonatal category”. Thus, heaping by the women creates problems with the demographic classification of deaths, which in turn influences demographic trends. Consequently, the demographic picture reached in these instances is not entirely true.

Overall evaluation of WFS data from between the period of 1974 and 1982 revealed that reporting was more complete in Latin America and less complete in Africa and Asia. The worst cases of data reporting were noted for Bangladesh, Benin, Mauritania and Yemen AR (Singh, 1987). Common to these unsatisfactory cases, was that the date of last birth was generally reported with a higher frequency than dates of earlier births, thus bringing the problem of recall ability related to recent events (Singh, 1987).

Whilst Hill and Choi (2006) acknowledge that retrospective reporting can suffer from recall lapse; it was found that there was no systematic pattern of differences between the estimates by recall period in their study.

#### **2.4.2 Saliency**

A second type of response error common to experimental and empirical literature is the aspect of saliency that is attached to events. According to Mathiowetz *et al.* (2002:163) “Saliency is hypothesized to affect the strength of the memory trace and subsequently, the effort involved in retrieving the information from long-term memory. The stronger the trace, the lower the effort needed to locate and retrieve the information”. Similarly, Sudman and Bradburn (1973) identify salient events as those events that have continuing economic or social consequences for the respondent. It is postulated by myriad studies (Sudman *et al.*, 1996 and Hertrich, 1998), that events deemed highly salient to the respondent are better recalled. To add, through their

research work Cannell, Fisher and Bakker (1965) comment that those events judged to be important to an individual is more *completely* and *accurately* reported than other events.

To illustrate the concept of saliency, I will draw on a study conducted in the 1970s. The example to follow not only highlights the saliency attached to the reporting of deaths but also the sex selectiveness of that reporting. The following example draws on findings of a study conducted in several villages north of Pakistan<sup>6</sup>. The aim of the study was to determine mortality and health problems in relation to cultural factors with particular emphasis with regards to the treatment between the sexes. Births to the same mother were collected. In other words, respondents accounted for their siblings.

Ahmad (n.d) cited in Blacker and Brass (1979) noted that the reported mean numbers of male and female children decrease sharply after 40 years of age. This maybe attributed to the omission of dead siblings. It was found that for females (sisters) that are above the ages of 20, these were more often omitted than their brothers. This finding correlates with the confirmation by other studies that there is a common failure to report dead sisters than brothers in infancy. This is perhaps reflective of the male dominated ideologies that are prevalent in Northern Pakistan where the minimal attention paid to the death of a sister is illustrative of their social standing.

Another crucial point that emerged from this study was that the reporting of siblings is greatly influenced by who the main respondent (either male –head of the household or female) is. It was found that male respondents reported on their brothers better than their sisters. Thus according to Blacker and Brass (1979: 57) “the male head will be a factor predisposing towards greater omissions of siblings of his wife and other female dependents but women may have, in contrast, a more retentive memory than men of their dead sisters”.

Thus by reviewing the example, we see that female deaths are considered important for certain individuals, where this importance is influenced by social and cultural conditions. The example

---

<sup>6</sup> Preliminary results of the study were published in the paper by Blacker and Brass (1979). Attempts to track the final report were unsuccessful. For reference purposes I will refer to the work as *Blacker and Brass (1979)*.

points to the many complexities of retrospective reporting which are often to the detriment of achieving good data quality. .

### **2.4.3 Telescoping**

Voluminous amounts of literature consulted regarding retrospective birth histories also brought up the notion of 'telescoping'. According to Beckett *et al.* (2001), telescoping is where more events are recalled as having occurred in the most recent period and fewer in the more distant past. To expand, responses due to telescoping errors are typically classified as either backward telescoping or forward telescoping (Mathiowetz *et al.*, 2002). The former relates to the tendency of the respondents to report events as occurring earlier and the later refers to the tendency to report events more recently than they actually occurred. A third error under the subject of telescoping is recall decay (errors of omission). Recall decay is described as the inability of the respondent to recall relevant events occurring in the past.

According to Mathiowetz *et al.* (2002: 163) "Forward telescoping is believed to dominate recall errors when the reference period for the questions is of short duration, while recall decay is more likely to have a major effect when the reference period is of long duration. In addition to the length of the recall period, the relative salience of the event affects the likelihood of either telescoping or memory decay". Mathiowetz *et al.* (2002) explains that if an event/s that is unique and that has a profound impact on the life of the respondent, the likelihood of forgetting the event is minimal. However, the vividness of such event/s may cause the respondent to recall the event as having occurred sooner than it actually did (i.e. forward telescoping) (Mathiowetz *et al.* 2002).

Research done on telescoping (Rubin and Baddeley (1989), Huttenlocher, Hedges and Bradburn (1990)) suggests that there are three factors that are believed to contribute to this phenomenon. The first of which is that the longer ago the event took place, the probability of forgetting the event is higher. Secondly, errors in dating events are random or unbiased, but these increase as the time between the event and recall of the event lengthens. The third and final contributory factor towards telescoping is where there are intrusions of events that occur outside the reference period. For example, if a respondent has to give an account of visits made to an antenatal clinic in the past month, it is possible that they might include visits that occurred

outside the month, that is earlier or later visits. According to Beckett *et al.* (2001: 596 and 597) “the combination of the second and third factors means that events that occurred further back in time will be more likely to be remembered as having occurred within the recent past. This process leads to errors in reports of the timing of events and trends to produce an overestimate of the number of events in the recent reporting interval”.

The issue of telescoping was investigated in the First and Second Malaysian Family Life Surveys (MFLS-1 [1976-77] and MFLS-2 [1988-89]). The aim of both rounds of the surveys was designed in order to investigate the influences on fertility, infant feeding and survival with the sample being restricted to only ever-married women of childbearing age.

Beckett *et al.* (2001) notes that several researchers (Haaga, 1986, Sine and Peterson, 1993 and Smith and Thomas, 1997) have explored the presence of telescoping in the MFLS with mixed results. Whilst some studies (Smith and Thomas, 1997) showed telescoping in the data, research conducted by Haaga (1986) and Sine and Peterson (1993) note that the MFLS data is unusually accurate for retrospective reports of this nature. They attributed this to the requirement of Malaysian citizens to keep identification cards detailing their children’s date of birth.

## **2.5 Response Errors Caused by Survey Instruments**

### **2.5.1 Questionnaires as a Source of Measurement Error**

Not *all* errors are altogether caused by the respondent. According to Mathiowetz *et al.* (2002) questionnaires can also serve as a source of measurement error. They are of the opinion that the questions asked will convey to the respondent the significance of the study. Mathiowetz *et al.* (2002) argue that although this message is (hopefully) conveyed, there are several linguistic, structural and environmental factors that influence the interpretation of a question by the respondent. In other words, the specific wording of a question, its structure, the order in which it is placed and the manner in which it is presented plays an integral role to achieving a close to accurate response.

A common field error is the mis-translation of questionnaires from English to a specific language. Drawing from experiences in the Gambian Census of 1973, the translation from English to African vernacular languages posed a problem, as key questions became ‘seriously

garbled' (Blacker and Brass, 1979: 58). For example, in Gambia it is thought of as more polite to speak of *children* rather than *births*. In addition, in the two main local languages<sup>7</sup>, there is no equivalent word for the English word *date* (Blacker and Brass, 1979). Consequently, locals are unfamiliar with the concept of dates. In cases where the date needs to be asked (for example, asking the date of the last birth) enumerators asked the age of the last child born (Blacker and Brass, 1979). The seriousness of such errors committed on the part of the enumerator will be highlighted in a later example in the section 2.5.2. Though what can be said for now are that such errors leads to highly infected data.

A paper written by Bignami-Van Assche, Reniers and Weinreb (2003) sought to evaluate the quality of data collected as part of the Kenya and Malawi Diffusion and Ideational Change Projects. They specifically investigated interviewer effects, response unreliability and sample attritions and their possible interactions. In their study, Bignami-Van Assche *et al.* (2003) discovered that the reporting of total number of children seemed less reliable across both Malawi and Kenya. In Kenya, men appeared to report more consistently than women. The situation in Malawi was contrary to Kenya, in that women reported more consistently than Malawian men. Bignami-Van Assche *et al.* (2003) concluded that these inconsistencies observed maybe attributable to the difference in wording of the questionnaires. In the Kenya Diffusion and Ideational Change Project, the question posed referred to the number of deceased children whereas in the Malawi Diffusion and Ideational Change Project, respondents were asked how many of the reported number of children were still living. This finding hence points to how a questionnaire can influence and bias responses given by the informant.

As seen in the Gambian, Kenyan and Malawian experience, the above mentioned errors and contexts do not lend themselves to a simple matter of semantics but are more involved in their own right. Mathiowetz *et al.* (2002) has commented that the wording of questions is a major problem common to many surveys. Whilst, it is possible to standardize the *language* read by the respondent or the interviewer, this in turn does not imply the standardization of *meaning* that is attached to the questionnaire (Mathiowetz *et al.*, 2001). In the same vein, the standardization of the questionnaire language should be approached with caution. Blanket strategies such as these

---

<sup>7</sup> The two principle local languages are *Mandinka* and *Wolof*

can be at the peril of developing countries, particularly the sub-Saharan region of Africa. This comment is made in light of several studies findings regarding the varying literacy levels among and between developing nations of the world. Hence, the wording and language of questionnaires should be made appropriate for the intended survey respondents.

### ***2.5.2 Enumerators as a Source of Measurement Error***

Returning to the Gambian Census of 1973, tape recordings were a mandatory part of fieldwork. The aim of these tape recordings was to capture the enumerator-respondent encounter. Consequently, these recordings provided valuable insight into the causes and nature of reporting errors (Blacker and Brass, 1979). For example, analysis of the tapes found that enumerators failed to ask the scheduled questions (Blacker and Brass, 1979). According to Blacker and Brass (1979) about one third of the tape recorded interviews revealed that an important question like the date of the most recent birth was simply not asked by enumerators. Yet in these cases, information was still recorded on the questionnaires. The question that springs to mind is how was this possible?

It turns out that enumerators inferred from the information given from the eligible women based on the age of the youngest child in the household (Blacker and Brass, 1979). Taking into account the previously mentioned example of the 1973 Gambian Census, such inferences are extremely dangerous as a typical scenario that could have been a reality for most women is if the women in question had given birth to another child who may have died in infancy or is living elsewhere (Blacker and Brass, 1979). A logical or error check question (i.e. whether or not the last born child is still living) was added in the schedule of questions, but once again not asked by the enumerators in cases that spanned more than half of the interviews as percentage the recordings (Blacker and Brass, 1979).

### ***2.5.3 Timing Errors in Questionnaires***

Studies have shown that errors, for example errors attributable to timing, can also originate from the survey instrument and the dynamics of the interview itself (Hobcraft and Murphy, 1986). Typically, birth histories are collected by reporting and collecting data on the first birth followed by subsequent births. Reporting beginning with the first birth proves crucial to the collection and accuracy of the whole birth history.

Hobcraft and Murphy (1986) add that a common understanding is shared between the interviewer and the respondent. That is, the acknowledgement of a certain minimal interval must elapse between births. All is well until this point in the chain; the problem however begins when the dating of the first birth is incorrect. As a result, subsequent births will either be pushed forward or backward in time. This can lead to a 'bunching' of reported births for a period of five to fifteen years before the survey (Hobcraft and Murphy, 1986 and Potter, 1977).

In addition, Hill and Choi (2006) further highlight the problem of 'bunching' and shifting of deaths backwards for recent births that was prevalent in their study. According to Hill and Choi (2006), the effect of such instances is not clear on the net effects of child deaths reported but will most definitely affect analytical procedures.

## **2.6 Socio-Cultural and Political Context of Reporting**

The issue of 'context' is a crucial point to consider when engaging in discussions regarding retrospective reporting. Individuals are governed by the socio-economic and cultural institutions that preside over the societies that surround them. It is these influences that shape individuals decisions and actions, whether physical, verbal or mental, overt or covert. Contexts evolve as per the prevailing socio-cultural and political conditions, hence evolving individuals in the process. No doubt, it is these very multi factorial etiologies that play an important role in *what we say, how we say it and to whom we say it*. The discussion to follow brings to the fore, the socio-cultural and political context of retrospective reporting.

A study conducted by Hertrich (1998), among the Bwa society in Mali, although it had a slightly different focus but where the salient issues that emerged remain relevant, highlighted some of the issues mentioned above. The study aimed at investigating the reliability of responses of both women and men (husbands and wives) with regards to birth and marriage histories. Naturally, the former is where our interest lies. Although the irregularities were not checked against a vital statistics record but that of verbal autopsies of either partner, issues of omissions and misreporting of births and infant deaths by women are still applicable to the study.

Hertrich (1998) found that women do not always emerge as being the best and automatic source of information on birth histories. Reasons for this conclusion include instances where

men reported a higher number of cases (specifically foetal deaths) as opposed to women. Logic would prescribe that due to the intimacy of loss with these kinds of death; women would surely report and remember these instances. On the contrary, the study found that, culturally, foetal deaths were seen as a failure in terms of childbearing and so women choose to mask, neglect and blot the event out of their memory (Hertrich, 1998). The study also found that deaths in infancy were more frequently reported by women as opposed to their male counterparts. For the reporting of live births, underreporting by males was noted to be very slight and hence points to the cultural nuances of patriarchic domination and male virility in the Bwa society of Mali (Hertrich, 1998).

The evaluation of the demographic data of China suggests that birth statistics are fraught with severe underreporting (Merli, 1998). It is highlighted that in censuses, surveys and registration systems respondents often fail to record births and infant deaths. Whilst it is generally accepted that the under registration of births and infant deaths is due to women failing to recollect events from the distant past. The situation although inclusive of the former, is somewhat different in China.

Through research in the past, clues regarding omissions of births usually come from a scrutinizing look at sex ratios of reported births over time. Through various studies (Smith, 1994; Merli, 1998 and Merli and Raftery, 2000) that sought to investigate the underreporting of births and infant deaths in China, outcomes of these studies reveal that retrospective birth histories in China have been responsible for the rise in reported sex ratios<sup>8</sup>. Smith (1994) and Merli and Raftery (2000) further argues that this hardly disputed increase coincides with the implementation of the one-child policy. The study conducted by Merli (1998) noted that births and deaths are *intentionally* omitted from birth histories and other retrospective fertility surveys for reasons attributable to family planning policies like the one child policy. Reason being, this

---

<sup>8</sup> The sex ratio is the ratio of males to females in a given population that is expressed as the number of males for every 100 females. The sex ratio at birth in most countries is about 105 or 106 males per 100 females. After birth, sex ratios tend to vary because of different patterns of mortality and migration for males and females within the population (Population Reference Bureau, 2004).



system operated on a scheme of rewards and penalties, where incentives were great for those couples that reported neither births nor infant deaths.

Furthermore, Bang, Reddy and Deshmukh (2002) for their research used data from a study that was conducted in 13 sites in Maharashtra, India to argue that official statistics in India underestimate neonatal and infant mortality. During their investigation, Bang *et al.* (2002) discovered irregularities with the data collection procedures. It was suspected that respondents were paid for reports of deaths. Consequently, this meant an over-reporting of child deaths.

The above two examples of China and India who are incidentally the world's most populous countries and where levels of poverty and inequality between and among urban and rural areas are highly polarized (Pal and Ghosh (2007) and Ravallion and Chen (2004)) demonstrates how financial temptation can be wielded to distort the quality of data.

In the context of China, Coale and Banister (1994) commented that the elevated sex ratios are due to the collection of birth histories in a culture where the definition of a birth may exclude death shortly after delivery. By comparison, the death of an infant in the Western regions of the world (United States, United Kingdom and so on) is considered to be a *stillbirth* in China (Coale and Banister, 1994). In addition, the higher than normal sex ratio reveals that the Chinese tend not to report such children as ever being born. Coale and Banister (1994) comments that the birth of a child in China is only recognized and celebrated once the child has reached the age of a month or more This follows on from the Chinese tradition in which is child is not seen as 'fully human' during the first year of life (Coale and Banister, 1994). Others dismiss this cultural explanation by noting that it is simply a case of under registration due to the flaws in the registration process (Coale and Banister, 1994).

The preference for boys as opposed to girls is well documented in regions of the world such as East Asia, South Asia, Middle East and North Africa (Arnold, Choe and Roy, 1998). Social and economic advantages and disadvantages associated with either sex are noted by writers such as Bulatao, 1975; Vlasoff, 1979 and Pollak and Watkins, 1993. This crude 'cost-benefit analysis' of boy and girl children is most pervasive in the context of India, where girls are viewed as a liability. Due to the strong preference for boy children it affects the minimal attention and

inferior treatment paid to girl children. This statement is confirmed by the United Nations *Human Development Report* which recorded India in the bottom quarter of all countries with respect to the status accorded to women (Arnold *et al.*, 1998). The question that then arises is what does this have to do with reporting? The link is there.

According to Arnold *et al.* (1998) the preference for a son is believed to be one of the principal underlying causes of excess female mortality which is prevalent in early childhood. A considerable amount of these female deaths are unreported in surveys and censuses (Arnold *et al.* 1998). In support of this, Das Gupta and Bhat (1996) quote statistics that are reflective of underreporting of female child deaths in India. They note that more than one million additional girl children were missing between the period of 1981 and 1991 (Das Gupta and Bhat 1996). This estimated figure was over and above the already missing girls as a result of the skewed sex ratios in the Census of 1981 (Das Gupta and Bhat 1996).

In light of the fact that India is one of the world's highest populated countries, close to accurate data is needed to monitor and control the situation. However as seen in the example, the nature of reporting is heavily influenced by the socio-cultural norms that operate in parts of India. Whilst the intention to rectify is present, strategies to curb and control for population growth is marred by such practices. It is thus hoped that the study will perhaps shed some light on respondent characteristics associated with reporting errors, where these findings could serve as a step in the direction of improving the quality of data.

## **2.7 Respondent Socio-Economic Characteristics**

Beckett *et al.* (2001) have commented that empirical research on respondent characteristics that are associated with poor reporting has yielded inconsistent findings. Consequently, there has been limited evidence that has pointed to an association between the recall of events and respondent characteristics. To put this into perspective Marquis, Cannell and Laurent (1972) concluded from their research that there are no consistent patterns of underreporting by respondent characteristics (i.e. education, sex, family income or race/ ethnicity). On the other hand writers like Beckett *et al.* (2001); McGranahan (1976); Marckwardt (1973) and Ito (1963) to name but a few maintain that there is a link to respondent reporting and their socio-economic characteristics. The discussion to follow seeks to bring these arguments to the fore.

### **2.7.1 Education**

Through consultation with the literature I have found that studies like those carried out by Beckett *et al.* (2001) have found some correlation and others like Marquis *et al.* (1972) established no correlation between education and reporting. For example a study conducted by Hertzog and Rodgers (1989) was unable to ascertain whether there is a significant relationship between education and reporting. In the study, when the effect of education was controlled for a somewhat weak statistically significant relationship emerged between age and recall ability of respondents.

Taking into account the grey areas associated with education and retrospective reporting, we can assume the following for now. Education equips an individual with a range of intellectual and practical skills from they can draw on. Apart from the academics of education, the acquisition of education also provides an individual with various frameworks with which they see the world and act accordingly. Thus the link to reporting is a simple one. Education (even at the very basic level) allows a person to comprehend, understand and provide answers to questions posed to them. It is hoped they would answer these to the best of their ability. However, when taking into account errors like omissions and misreporting, the logical flow described above is not applicable. The situation becomes blurred and complex where such errors cannot be solely attributable to a lack of education.

For my study, it would be interesting to note the levels of education associated with accurate reporting. It is also quite possible that the association is negligible. Whatever the case may be, it is thus hoped the study helps in making the grey areas with regards to education and reporting a little clearer.

### **2.7.2 Ethnicity (Population Group)**

As noted by Sibanda and Zuberi (2005) race or ethnicity is a useful variable in determining relationships between sub groups of a population. These provide us with socio-economic and demographic information so as to allow researchers to make comparison and draw conclusions about a population. In some societies in which there is the existence of more than one racial or ethnic group, one group tends to dominate the others (Weeks, 1989).

The link between respondent reporting and ethnicity has been explored by Webster (1996). In the study, one of the many hypotheses tested investigated respondent ethnicity in relation to response quality. The investigation placed particular emphasis on the field relationship shared between Hispanic and Anglo respondents and interviewers. The aim was to determine if ethnicity has a significant effect on the quality of response given by the respondent in relation to their assigned interviewer. Webster (1996) found support in instances where respondents bias their response items according to their interviewer's culture (Hispanic or Anglo). Whilst the focus was different to this study, it does however highlight the point that ethnicity does have a role to play in retrospective reporting. It is also quite possible that this association may be negligible. That said, the fact however remains that there is an effect recorded and it needs to be addressed accordingly.

Specific to our study it would be interesting to know if there is a particular racial group that misreports more significantly than others or is there simply no relationship between racial group and misreporting? If a relationship does exist, what is misreported by those groups?

### ***2.7.3 Age of Respondent***

Evidence from the literature highlights that misreporting and omissions occur in older women than any other age group (Marckwardt 1973). Women of these ages, either due to misunderstandings or memory failures have a greater tendency to omit children who died in infancy, children who have grown and left the home and those children who were born to another husband (Blacker and Brass, 1979). In his study, McGranahan (1976) concluded that the characteristic of age does influence reporting.

As mentioned, the omission of births is common to retrospective surveys, even in prolific surveys such as the WFS. In the WFS conducted between 1974 and 1982 the omission of births was tested by examining the mean number of children ever born by age: It was hypothesized that in the absence of fertility increases in the past, as well as omission, the mean number of children born should continuously increase with age (Singh, 1987). This hypothesis proved true and was observed for all but some countries. For countries like Mauritania, Morocco, Bangladesh, Pakistan and Indonesia it was found that there were a lower mean number of live births for 45-49 year olds compared to the 40-44 age groups (Singh, 1987). This was suggestive

of the fact that the oldest cohort did not report *all* of live births or there was transference of more fecund 45-49 year olds into the 50-54 age group (Singh, 1987).

In addition, for instances where older women failed to report the deaths of their female children, sex ratios at birth by time periods were examined. This scenario proved true for countries like Mauritania, Sudan, Bangladesh and Pakistan (Singh, 1987).

Taking these findings into account, it would be remarkable if we are able to establish a link between reporting and age, especially older ages. If this is established strategies can be put into place to remedy issues surrounding bad data quality that are influenced by misreporting and omissions of events by women of older ages.

#### **2.7.4 Place of Residence**

For inquires regarding the investigation of reporting errors in developing countries, place of residence (urban and rural) is one of the most frequently used variables in the explanation of such errors.

Urban areas are viewed as commercial centers with concentrated wealth, power and a western lifestyle (Weeks, 1989). On the other hand, rural areas are characterized by poverty, lower education levels, technological underdevelopment, lack of services and a close link to traditional practices (Weeks, 1989).

Based on this divide, it has been suggested by innumerable studies that women in urban areas report better in their birth histories than women of rural areas. It is postulated that the respective socio-economic contexts and the advantages and disadvantages associated with either spatial location as described above have an influence on the quality of reporting.

One of the hypotheses in a study conducted by Marckwardt (1973) in Peru found that there was less accuracy in reporting by older women and by women who live in rural areas. This finding pointed to the differential educational opportunities in Peru and no doubt in other parts of the world. A second hypothesis of the study pointed to the occurrence where younger women were more reluctant to admit to having had a baby (Marckwardt, 1973). On the contrary, older

women from rural areas although not reluctant to admit to a previous birth tended to report the births closer in time than they actually occurred (Marckwardt, 1973).

From the review of literature, we see that issues surrounding retrospective birth histories such as the failure to report births and deaths are not *entirely* deliberate or intentional as there are contemporary social, political and cultural forces that interact and influence reporting and omissions of birth and deaths.

## **2.8 Measurement and Collection of Fertility Data**

In their document, *Principles and Recommendations for Population and Housing Censuses* the United Nations (2006) provide some guidelines as to the collection of fertility data and its measurement thereof. Keeping the objective of the review in mind, I shall only refer to those guidelines that have a bearing on my research, so that in the chapters to follow, it will be possible to assess as to whether the manner in which our data was collected conformed to the prescribed recommendations as set by the United Nations and if not, what were the implications in terms of the results derived.

The United Nations (2006) stresses the point that the universe for which the data must be retrieved from must consist of women, 15 years of age and over, regardless of their marital status. The only instance where this option may not be carried out is when and if cultural ideals and values override this approach. With respect to this cultural component, the collection of such data from a group of women who have never been married is deemed unfeasible. The United Nations (2006) also stipulate that the appropriate age limit for the collection of fertility data be capped at under the age of 50, hence allowing for more concentrated data collection efforts for women of these ages. In addition, the United Nations (2006) prescribes that the group from which fertility data was collected should be clearly described in the census report so as to minimize or avoid ambiguity in the analysis of the data. Furthermore, it is suggested that all efforts should be made to collect this information *directly* from the woman or mother (preferably a natural [biological] mother) in question. This is attributed to the increased likelihood that the person will recall her fertility history more accurately than any other member of the household (United Nations, 2006).

In order to ensure the completeness of coverage as well as to assist the respondent in recalling the number of children ever born, the United Nations (2006: 118) have recommended a sequence of questions to be included in the following order:

- (a) "Total number of sons ever born alive during the lifetime of the woman";
- (b) "Total number of sons living (surviving) at the time of the census";
- (c) "Total number of sons born alive who have died before the census date";
- (d) "Total number of daughters ever born alive during the lifetime of the woman";
- (e) "Total number of daughters living (surviving) at the time of the census"; and
- (f) "Total number of daughters born alive who have died before the census date".

For these questions, the United Nations (2006) stipulates that all children born alive during the lifetime of the woman up until the census date of enumeration should be included. Children born alive should include all live births irrespective whether the children were born (a) in or out of wedlock; (b) whether they were born in the preceding or current marriage; (c) whether they were born in a de facto union and (d) whether they were living or dead at the time of the census. More importantly, foetal deaths, stillborns and adopted children should be excluded.

The United Nations (2006) acknowledge that the data on children living is not always complete and reliable. This is due to inefficient vital registration systems. Thus to combat this, the United Nations (2006) realizes the importance and possible improvement to the quality of the data obtained. They argue that if more detailed questions regarding the current residence of children ever born are asked there will be a marked improvement in the coverage and the quality of the data. They make recommendations with respect to asking the following kinds of questions (United Nations, 2006: 119):

- (a) "Total number of sons living in the household";
- (b) "Total number of sons living elsewhere";
- (c) "Total number of sons born alive who have died before the census date";
- (d) "Total number of daughters living in the household";

- (e) "Total number of daughters living elsewhere"; and
- (f) "Total number of daughters born alive who have died before the census date".

The benefit of using these questions, apart from the more accurate reporting of children ever born, the questions are specified by sex and hence will be more aptly suited to an improved subsequent analysis.

Improvements to questions such as those discussed were implemented in the census schedules of countries like Kenya and Swaziland (Blacker and Brass, 1979). Improved quality in data has been documented through the inclusion of questions on the women's most recent live birth (Blacker and Brass, 1979). Instances of data quality improvement in this respect were observed in Kenya for the censuses conducted in 1962 and 1969 (Blacker and Brass, 1979). Another case in point is that of the Swaziland censuses of 1966 and 1976. The improvements noted in Swaziland, not only impacted on the overall data quality of fertility, but this improvement also greatly influenced the shape of the fertility distribution in a positive manner (Blacker and Brass, 1979). These findings once again point to the need for an investigation retrospective data quality issues.

## ***2.9 Collection of Fertility Data in South Africa***

In the discussion to follow I will reflect on the manner in which data was collected for fertility in South Africa. The aim of this will be to ascertain what questions have been used in the past and what implications of this has been noted in the literature in terms of the quality of the data.

The South African society is characterized by high levels of poverty and vast inequalities, stemming from the historical policy of apartheid (Reddy and Khan, 2006). Hence the implementation of a census in a context such as this is a challenge. Despite these complications, the administering of a population census serves to inform policy making and planning at the national and provincial level in addition to the accumulation of a national statistical database (Khan, 2007). When paired, the context of South Africa and the necessity to carry this undoubtedly expensive, but essential exercise is of great imperative.



As noted by Moultrie and Timaeus (2002), fertility in South Africa has been under-examined due to inadequate census and vital registration, with particular reference to the African population. Consequently, the absence of research work and the investigation into fertility trends in the past were driven mainly by the lack of (quality) data. On the political front, South Africa's international isolation had repercussions in terms of its exclusion from prolific surveys such as the WFS in the 1970s and the earlier rounds of the DHS (Moultrie and Timaeus, 2002).

Since the new political dispensation in 1994, South Africa has conducted two population censuses, with one in 1996 and another, five years later in 2001. It is undisputed that the census of 1996 opened an information gateway to the multitudes of social and economic facets this country possesses. As a result, aspects of South African demography was analyzed and investigated in ways that were not possible before (Moultrie and Timaeus, 2002).

As noted by Moultrie and Dorrington (2004), in a census it is inevitable that not all respondents will answer questions completely and correctly. There is bound to be some inaccuracies, whether on the part of the respondent or the enumerator. This, however minute it maybe, impacts on data quality. The discussion to follow shall examine the question schedules used in both the 1996 and the 2001 Census. The purpose is then to comparatively assess the quality of data and try to ascertain the problematic areas in the data.

### **2.9.1 The 1996 Census**

The question in the table below is the fertility schedule of questions for Census 1996. These were extracted directly from the Census 1996 questionnaire.

**Table 2.1:**  
**Census 1996 Fertility Questions**

- *How many children, if any, has the woman ever given birth to? (live births) (Please include her children, who are not living with her and those who have died.)*
- *How many of her children are still living?*
- *When was her FIRST child born? (live birth)*
- *For those mothers born after 10 October 1946 – less than 50 years of age). How many children (live births), if any, has she given birth to IN THE LAST 12 MONTHS? (since 10 October 1995)*

As seen in Table 2.1 the questions asked referred to children ever born and children living. With specific reference to the questions “How many children, if any, has the woman ever given birth to?” and “How many children (live births), if any, has she given birth to in the last 12 months?” Moultrie and Timaeus (2002), note that the responses to these questions were poor. It is argued that responses to the first of these questions were not obtained from a significant proportion of women of reproductive age. Moultrie and Timaeus (2002) add that from an examination of the data it seems as though many respondents either failed to understand the second question or it was simply a matter of responses being recorded inaccurately. Consequently, adjustments were made to the lifetime fertility data in census 1996 as the unadjusted data would have been unable to generate robust estimates of fertility.

Whilst a discussion of how these adjustments were made is beyond the scope of this study, the point however that needs to be driven is the error fraught data. It is thus seen how misunderstandings on the part of the respondent, enumerators and perhaps the survey instrument itself can lead to ambiguities and inaccuracies in the data. Perhaps these errors (although serious) can be slightly overlooked as this was the first census administered to count the whole of the new formed democratic South Africa. If this consideration was made, the true test of data quality as a result of experience gained in the 1996 Census would come in the census that followed, that is Census 2001.

### **2.9.2 The 2001 Census**

The questions that appear in the table below are the Fertility Schedule as used in Census 2001 extracted from the Census 2001 questionnaire.

**Table 2.2:  
Census 2001 Fertility Questions**

- *How many children, if any, has (the person) ever had, that were born alive?*  
How many of these were boys?  
How many of these were girls?  
Include ALL her children, i.e. those who are still living, whether or not they live in this household, and those who are dead. DO NOT COUNT STILLBIRTHS (children born dead).
- *STILL LIVING: If the person has ever given live birth:*  
If boys: How many boys are still alive?  
If girls: How girls are still alive?

- **LAST CHILD BORN.** *If (the person) has ever given live birth:*  
When was (the person's) last born child?  
What was the sex of that child?  
Is that child alive or dead?  
Write the day, month and year of the last live birth and dot the appropriate box of the sex. If multiple birth, indicate only the last child. Dot the appropriate box whether the child is still alive on Census night 9-10 October. **DO NOT COUNT STILLBIRTHS** (children born dead)

From Table 2.2 we see that the census asked three main questions from women of reproductive age. On the whole, the demographic data that was collected on mothers was of an acceptable quality. The same however, cannot be said of the data on lifetime fertility collected from the same group of women. Moultrie and Dorrington (2004) in their assessment of the overall quality of the fertility data comment that the questions on fertility were answered poorly by the women of childbearing ages in the Census 2001, with the extreme being, ignoring of questions on sex composition. According to Moultrie and Dorrington (2004: 6) "Questions regarding the sex composition of the numbers of children born and the numbers of children surviving, in aggregate more than half of all women of childbearing age had their answers to one or more of the following questions subjected to logical imputation<sup>9</sup> or hotdecking<sup>10</sup>". Put simply, the values of missing data was estimated and filled in so as to minimize bias.

To illustrate this concept, Moultrie and Dorrington (2004) note that high degrees of imputation and hot decking were applied to women of younger ages. This was probably due to enumerators leaving the fertility responses given by younger women of childbearing ages who had no children as blank as opposed to filling in a zero (Moultrie and Dorrington, 2004). Thus instead of being noted as 'parity zero' these women are noted as 'parity missing'.

On the issue of specific questions asked, when looking at the Census 1996 question schedule, one of the questions asks of births in the last 12 months. As mentioned previously, this question in the 1996 Census was badly answered by respondents. Building on from this, Moultrie and Dorrington (2004) believe that in the Census 2001 the second question on children still living

---

<sup>9</sup> Imputation is the assigning of a value to a field, either for non-response or for replacement of a recorded value determined to be inconsistent with a set of edits (United Nations, 2001).

<sup>10</sup> Hot decking is a method of imputation where donor records are taken from a current deck of sample data. Cold deck imputation refers to another method of imputation where donor records are extracted from past survey data (United Nations, 2001).

should include births that occurred 12 months preceding the census regardless of the vital status of the child. This point is made in light of the Census 2001, only being able to capture only about a half of the 1.1 million births estimated to have occurred over the year before the census (Moultrie and Dorrington, 2004). Efforts were made in the 2001 Census to rectify this by asking the date of the last born child. According to Moultrie and Dorrington (2004: 14), “more than one in four women of reproductive age (26.4 per cent) did not have a plausible response recorded to this question. Blank responses that should have had a date of birth imputed or hotdecked accounted for 8.9% and 9.8% of all responses respectively. A further 7% of women gave nonsensical responses, and had to have their responses imputed based on the responses for the household member ‘identified’ as being that woman’s child”.

In addition, in their investigation, Moultrie and Dorrington (2004) observed that in instances where responses were given, there was an obvious preference for certain dates among those women whose data was not imputed or subject to hot decking. It was found that the dates given were almost uniformly distributed with heaping noted especially around the first of the month as well as the Census day (Moultrie and Dorrington, 2004). It was also found that the data on the proportion of children ever born that are still alive indicated that women interpreted the question on *children ever born* as asking about *children still alive* (Moultrie and Dorrington, 2004). Consequently, the data on the number of children born to women of reproductive ages was seriously deficient (Moultrie and Dorrington, 2004).

These examples clearly illustrate the distortion in the data. As a result; these flaws have great ramifications for the estimation of fertility trends. Although the efforts such as asking the date of the last born child in Census 2001 is commendable, this however does not mean much if the end result is not reached (i.e. improve the quality of the fertility data).

In summation it is seen that both the Census 1996 and 2001 were subject to poor responses. A logical conclusion that can be drawn at this stage is that there is a clear indication that responses may have been due to inaccurate recording of responses by the enumerator or misinterpretation on the part of the respondent. Whatever the case may be, we see the inaccuracies regarding the usage of retrospective data. For example, from the Census 2001 we see the heaping or preference of certain dates as opposed to others. Of equal importance is the

misinterpretation of questions on the part of respondents. Perhaps this maybe attributed to the wording and language of the questionnaire as discussed earlier or inadequate enumerator training. Unfortunately, this manner of fertility data collection is the only viable option at this stage in light of the absence of efficient and complete vital registration systems in the country. The way forward should thus be a case of improving on existing processes rather than revolutionizing processes of fertility data collection

### **2.10 Summary**

In this chapter discussions revolved around the process of retrospective reporting as well the context in which reporting occurs and the consequences thereof in terms of the quality of data obtained. In addition the South African context regarding the collection of fertility data was explored where discussions around the two post 1994 censuses were presented.

By way of collating the findings from the literature as well as the discussion presented in the chapter, I wish to end with a standpoint adopted by Ito (1963) in his paper titled *An Analysis of Response Errors: A Case Study*. In the paper, Ito (1963) comments that whilst age and education amongst other factors like respondent occupation and the training of the interviewer are linked to response errors, in a probability sample (like our sample) respondent characteristics cannot be controlled by the survey and research design and thus response errors are a reality. Whilst this is slightly de-motivating, it is this very relationship between response errors and respondent characteristics that are valuable in developing a system or technique of correction procedures that can outline the way forward. According to Ito (1963:447), “perhaps a general theory of response errors awaits more intensive study of psychological factors contributing to recognizable patterns of response errors... perhaps a true understanding of the ‘cause’ of response error depends upon greater knowledge of the human mind”.

I find it appropriate to depart from this chapter with these key words in mind, as it is this thinking of Ito (1963) that serves as a driver to the chapters to follow.

## CHAPTER THREE

### Research Design and Data Analysis

\*\*\*\*\*

### 3.0 Introduction

Chapter three presents a brief history of South Africa, the context within which the research was conducted as well as the manner in which was carried out. The chapter discusses the questionnaires, the sample, data analysis procedures as well as the framework for interpreting the data.

### 3.1 Brief History and Socio-Demographic Profile of South Africa

**Figure 3.1: Map of South Africa (United Nations, 2007)**



Situated at the southern tip of Africa, edged by the Indian and Atlantic Oceans, South Africa is a country rich in history and heritage. From its history dotted with losses and triumphs, South Africa has eleven official languages, nine provinces, with a vast number of practicing religions and ethnic groups.

From a population of 40.5 million in the 1996 Census to 44.8 million in Census 2001, the South Africa’s population according to the 2009 mid-year population estimates is an approximate 49.32 million (Statistics South Africa, 2009). The most populous provinces are Gauteng and KwaZulu-Natal, each with a population of over 10 million people (Statistics South Africa, 2009). The least populous province is the Northern Cape (1.1 million) (Statistics South Africa, 2009). As noted by Statistics South Africa (2009), the Black/ African is the most populous population

(39.13 million), followed by the White population (4.47 million), Coloured population (4.43 million) and lastly the Indian/ Asian population (1.27 million). With regard to the sex composition of the population, just over 52% (25.45 million) is female (Statistics South Africa, 2009).

The 2009 mid-year estimates note that life expectancy is 53.5 years for males and 57.2 for females (Statistics South Africa, 2009). The infant mortality rate is estimated at 45.7 per 1000 live births (Statistics South Africa, 2009). The TFR for South Africa is 2.38 per woman (Statistics South Africa, 2009).

The overall HIV prevalence rate for the country is estimated at 10.6%. The total number of people living with HIV is approximately 5.21 million (Statistics South Africa, 2009). 17% of the adult population (15-49 years old) is HIV positive (Statistics South Africa, 2009). According to the 2009-mid year population estimates, other HIV related estimates include an approximate 1.91 million AIDS orphans, 354 000 new infections among adults aged 15-19 and 59 000 new infections among children (Statistics South Africa, 2009).

According to findings from the 2007 Community Survey, living conditions in the country have improved. For example, 71% of households live in a formal dwelling compared to that 64% during the 1996 Census (Statistics South Africa, 2007). Significant increases in electricity usage for lighting, cooking and heating were noted between the Census 1996 and the Community Survey (Statistics South Africa, 2007). In addition, 88% of households have access to piped water. Notable improvements have been made with regard to school attendance and educational attainment (Statistics South Africa, 2007). School attendance among 5-24 year olds have increased from 63% in 1996 to 74% in 2007 (Statistics South Africa, 2007). The percentage of people with no schooling has decreased from 19% in 1996 to 10% 2007 (Statistics South Africa, 2007). Persons with some secondary education have increased from 34% in 1996 to 40% in 2007 (Statistics South Africa, 2007).

Despite these positives and achievements, a substantial amount of work to is yet to be done. It is hoped that inter-governmental and inter-sectoral efforts will aid in addressing challenges,

such as teenage pregnancy, the threat of xenophobic violence, poverty, inequality, service delivery and unemployment to name a few, that continues to hamper South Africa's progress.

### **3.2 Context of the Research**

As discussed in Chapter two: Literature Review in *Section 2.9*, the undertaking of a census is by no means an easy task. It is of utmost importance that for these operatives to be carried out at an optimal level, the success of such an undertaking is incumbent on the adequate research on planning and the methodologies to be put in place in anticipation of the census period. Repercussions of these processes if not put into operation comprehensively run deep, with the most adverse impacting on the dependability and trustworthiness of the data and results derived.

Taking cognizance of these above mentioned points, as well as South Africa's history of racial segregation, its young democracy and its socio-demographic profile, the study realizes the importance in employing a cyclical research process, where the research findings are transformed into practice which then informs the research. In the context (such as that of this study) of measuring a critical demographic process such as fertility, this cyclical process needs to be viewed as a top level strategy as well as a key success indicator in gathering quality fertility data. From this we will be able to determine, the gaps and loopholes, and where they possibly originate from and more importantly, how can we deal with these challenges.

It is with this perspective, that the study was conceived. The following chapter discusses the data and the methodology employed for the research study.

### **3.3 Research Design**

The *Final Data Handover Report of Project: Census 2001 Data Processing* recommended Statistics South Africa should explore the viability of asking extensive questions on fertility for the next census (Statistics South Africa, 2003). The motivation for such an action was driven in light of the problems experienced with regard to the quality of fertility data in the previous census. Following on from this, in November 2005, a census content research study was conducted where an alternative schedule of questions for measuring fertility was tested.



Fertility data was collected for two samples of women. The first sample of women first completed the Census 2001 questions on fertility. Immediately after completion of the Census 2001 fertility questions, the women were asked to complete a second questionnaire (i.e. female questionnaire) to account for all live births and deaths. A second sample of women followed the same procedure, except that the alternative fertility schedule of questions was administered to them and *not* the Census 2001 fertility schedule. As with the first sample of women, the second sample of women were asked to complete a second questionnaire (i.e. female questionnaire) to complete a roster of live births and deaths.

This second questionnaire, known as the female questionnaire was administered after the completion of the respective fertility schedule questionnaires. The relevant version of the female questionnaire (hereon referred to as the 'roster' of all live births and deaths) was completed for all eligible females (i.e. women of reproductive age). The aim in this strategy was to identify discrepancies between the fertility schedule and the roster of births and deaths.

### **3.4 Questionnaire Design**

In total, five questionnaires were used for the survey. The first version of the questionnaire was administered to sample one of women, and version two was administered to sample two of women. They were as follows:

- Fertility schedule questionnaire (Version 1) - Census 2001 fertility schedule of questions
- Fertility schedule questionnaire (Version 2) – Alternative fertility schedule of questions
- Female questionnaire (Version 1) – Roster of live births and deaths
- Female questionnaire (Version 2) – Roster of live births and deaths
- Living Standard Measure (LSM) Questionnaire

Questionnaires were translated and printed into Afrikaans, English, IsiZulu, IsiXhosa, SeSotho, SeTswana and TsiVenda. To minimize confusion, the name and version of each questionnaire was marked clearly on the cover of the questionnaire. Detailed foci of the respective questionnaires can be viewed in **Appendix A**. **Appendix B, C, D**, can be consulted for the different questionnaires. **Appendix E** can be referred to for additional information on the LSM and the motivation for its use in the research.

### **3.5 Sampling Frame**

For the study two probability samples were drawn for each version of the questionnaires. Each of the two national samples extracted consisted of a randomly drawn dwelling unit<sup>11</sup> from each of the Primary Sampling Units (PSU) of the Master Sample based on Census 1996 boundaries. There were approximately 2 971 dwelling units drawn. Sampling was not drawn from the Census 2001 boundaries as this would have led to response fatigue.

All households within an appointed dwelling unit were sampled. All eligible women in the covered households were to be interviewed. The implication of this method of sampling was that all Enumerator Area<sup>12</sup> (EA) types and geographic types (urban and rural) were covered. Both samples did not overlap.

Data for the first sample was collected and processed for 2 581 households. For the second sample this figure was **2 685 households**. Figure 3.2 on the next page illustrates the percentage distribution of these households across the national provinces.

For the first sample of women, using the Census 2001 fertility schedule questionnaire, information was collected for 2 720 women and 4 591 children utilizing the Female Questionnaire. For the second sample of women using the alternative fertility schedule information was collected for 2 881 women and 4 949 children utilizing the Female Questionnaire.

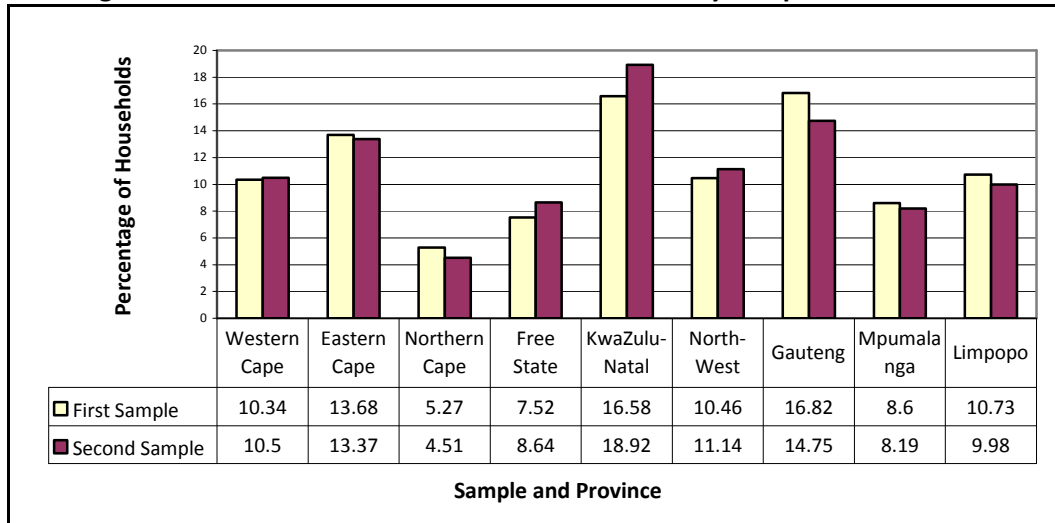
Figure 3.3 shows the percentage distribution of women used in analysis by sample across provinces.

---

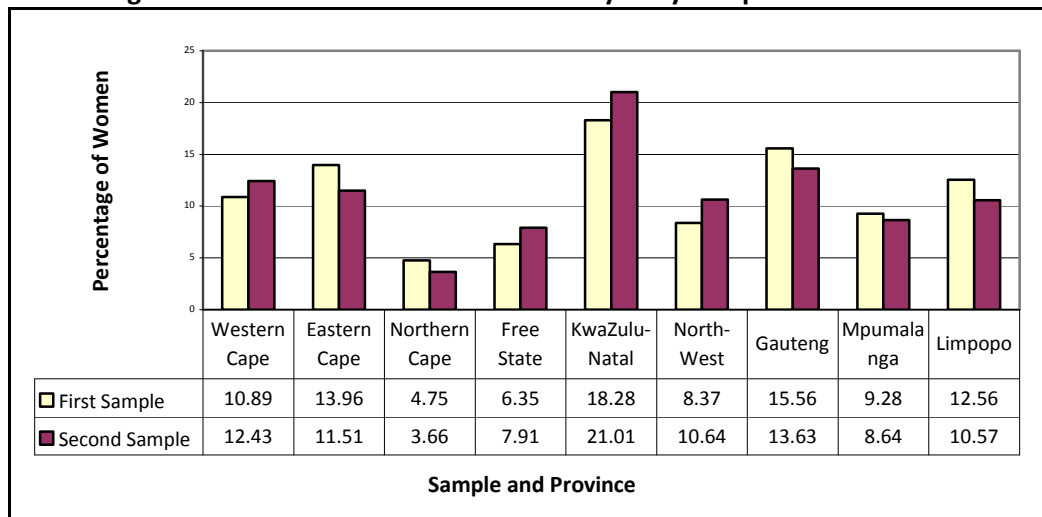
<sup>11</sup> Dwelling unit is a housing unit. A housing unit is a unit of accommodation for a household which may consist of one structure, or more than one structure, or part of a structure. A housing unit usually has a separate entrance from outside or from a common space, as in a block of flats (Statistics South Africa, 2004).

<sup>12</sup> Enumeration area is the smallest geographical unit (piece of land) into which the country is divided for census or survey enumeration. EAs typically contain between 100 and 250 households (Statistics South Africa, 2004).

**Figure 3.2:**  
**Percentage Distribution of Households for Data Collected by Sample and Across Provinces**



**Figure 3.3:**  
**Percentage Distribution of Women Used in Analysis by Sample and Across Provinces**



### 3.6 Data Analysis Procedures

In terms of the analytical procedure to be followed with respect to assessing the retrospective birth histories, firstly data sets were created which recorded respondent information for both samples of women (i.e. Census 2001 fertility schedule of questions and the alternative fertility schedule of questions) Thereafter, the study compared the data sets in terms of births and deaths recorded. In other words, inconsistencies are noted as when either of the two fertility schedules or their accompanying rosters record higher or lower proportions of women with births and deaths?

For the analysis, the study focused on women for whom at least one birth was recorded in the respective roster of births and deaths. By doing this, confusion around issues such as women with no children recorded, women who could not be interviewed and field non-compliance on the part of the enumerator with respect to the completion of roster of births and deaths for females with no children recorded and lost rosters are eliminated. As a result of this filter placed on the data, analysis for the first sample is limited to a total of 1 433 women and 1 504 women in the second sample.

Once this analytic foundation has been laid, the study explored the demographic and socio-economic characteristics of the sample. The aim here was to establish a profile of the respondents for whom higher or fewer births were recorded by the Census 2001 Schedule, the alternative fertility schedule and their accompanying rosters (i.e. an inconsistent record).

### **3.7 Method of Analysis**

#### **3.7.1 Grouping of Variables**

Variables that have been chosen for the analysis include: 'age', 'marital status', 'population group', 'level of education', 'province', 'geographical type' and 'living standard measure'. The rationale for choosing the above variables is simple: The above variables provide comprehensive information on individuals. On an individual basis, one can gauge their age, their relationship status, to which population group and province they belong and their level of educational attainment. On a broader level, their place of residence and LSM status is indicative of their socio-economic status.

For the analysis, categories such as 'age', 'marital status', 'level of education' and LSM have been collapsed into smaller groupings for analysis.

For the category of 'age', initially, five year age groups were extracted from both samples for analysis. These groups were 15-19, 20-24, 25-29, 30-34, 35-39, 40-44 and 45-49. These groups were used in providing a background analysis to both samples. Having decided these were too many groups for further analysis as well the low number of inconsistencies observed within these groups resulting in little statistical value and interpretation, the groups were further

collapsed into three main age groups. These were 15-24 (young women), 25-39 (middle aged women) and 40-49 (older women).

Four groups were created for 'marital status' The first group 'married' included individuals who were married in either a civil, religious, traditional, customary manner as well cohabiting individuals. The second group was 'never married' which included individuals who have never entered in a marital or cohabiting union. The last two groups were 'widow or widower' and 'divorced and separated'.

The variable 'level of education' was combined into four groups. The first group 'little or no schooling' included individuals who had no schooling and some primary school. 'some schooling' included those who completed primary school and some high school. The third category 'matric or higher' included women who had completed matric (Grade 12), completed a university or technikon degree or diploma. The fourth category of 'other or missing' included responses not specified on the questionnaire as well as missing values. This group was later omitted in the analysis due to its negligible size.

The LSM groupings of 1-10 into have been limited to three main groups (i.e. LSM 1-3, LSM 4-6 and LSM 7-10). LSM 1 to LSM 3 is considered a low LSM group (i.e. low socio-economic status). LSM 4 to LSM 6 is considered to be middle LSM group (i.e. average socio-economic status). LSM 7 to LSM 10 is considered to be a higher LSM group (i.e. higher material socio-economic status). Refer to **Appendix E** for additional information on the LSM.

The variable 'geographical type' and its division into 'urban and rural' was done by matching the primary sampling unit (PSU) number to the 2001 Census master sampling frame. Certain PSU numbers were found within urban and rural nodes according to the sampling frame and were thus classified as such.

### **3.7.2 Stages of Analysis**

Firstly, an analysis into descriptive statistics (i.e. frequencies and cross tabulations) was conducted for both samples. Analysis of the data was done via the statistical package SAS. In addition, the study also utilized secondary literature to provide explanations for the possible link

between the demographic and socio-economic profile on respondents who have irregularities in their retrospective birth histories.

*Part One* concentrated on establishing a demographic profile of women who report inconsistently.

*Part Two* examined the link between the socio-economic status of individuals and the aspects (i.e. sex selectiveness) of their birth histories that are misreported on.

### **3.8 Summary**

This chapter has dealt with the research design (how the research was conducted). The chapter discussed the composition of the sample and illustrated this graphically. Data analytical procedures (how the analysis will be done) were discussed where restrictions on the data and analytical techniques were presented.

**CHAPTER FOUR**  
**Results of the Study**

\*\*\*\*\*

**4.0 Introduction**

Chapter four presents the results of the study. In doing so, the results will be shown in tabular and graphical format. The chapter begins with an analysis of the background characteristics of both samples, moving onto the exploratory analysis which examines the demographic profile of women associated with irregularities in the reporting of their birth histories. Having established this profile, the chapter presents the link between response errors and the socio-economic status of respondents. The chapter concludes with presenting the link between the socio-economic status of individuals and the sex-selectiveness of reporting births and deaths.

**4.1 Analysis of the Background Characteristics for Sample One and Sample Two of Women**

Table 4.1 presents a demographic profile of both samples of women. The background characteristics that make up this profile are: (1) Age; (2) Population group; (3) Marital status; (4) Level of education; (5) Province; (6) Geographical type and (7) LSM Status. The aim of presenting such a profile prior to the main findings is firstly to equip the reader with a wider perspective on the sample as well as to provide a context in which the results can be viewed and scrutinized.

**Table 4.1:**  
**Background Characteristics of Women Aged 15-49 – Sample One and Two of Women**

Background Characteristic	Sample 1 (Census 2001 Fertility Schedule of Questions)		Sample 2 (Alternative Fertility Schedule of Questions)	
	N	%	N	%
<b>Age</b>				
15-19	75	5.23	71	4.72
20-24	196	13.67	206	13.70
25-29	238	16.61	266	17.69
30-34	212	14.79	269	18.61
35-39	254	17.73	254	16.89
40-44	250	17.45	244	16.22
45-49	208	14.51	193	12.83
Other/Missing	0	0	1	0.07
<b>Population Group</b>				
Black/African	1096	76.48	1169	77.73
Coloured	192	13.40	186	12.37
Indian/Asian	26	1.81	24	1.60
White	112	7.82	115	7.65
Other/ Missing	7	0.49	10	0.66

Background Characteristic	Sample 1 (Census 2001 Fertility Schedule of Questions)		Sample 2 (Alternative Fertility Schedule of Questions)	
	N	%	N	%
<b>Marital Status</b>				
Married	702	48.99	697	46.34
Never Married	616	42.99	685	45.55
Widow or Widower	60	4.19	60	3.99
Separated or Divorced	50	3.49	55	3.65
Other/Missing	5	0.35	7	0.46
<b>Level of Education</b>				
Little or No Schooling	291	20.31	277	18.41
Some Schooling	647	45.15	685	45.55
Matric or Higher	487	33.99	538	35.78
Other/ Missing	8	0.56	4	0.26
<b>Province</b>				
Western Cape	156	10.89	187	12.43
Eastern Cape	94	6.56	78	5.19
Northern Cape	174	12.14	150	9.97
Free State	75	5.23	90	5.98
KwaZulu-Natal	277	19.33	340	22.61
North West	121	8.44	165	10.97
Gauteng	223	15.56	205	13.63
Mpumalanga	133	9.28	130	8.64
Limpopo	180	12.56	159	10.57
<b>Geographical Type</b>				
Rural	692	48.29	750	49.87
Urban	741	51.71	754	50.13
<b>LSM Status</b>				
LSM 1-3	256	17.87	306	20.34
LSM 4-6	591	41.24	596	39.63
LSM 7-10	159	11.10	183	12.16
Other/ Missing	427	29.80	419	27.86
<b>Total</b>	<b>1433</b>	<b>100</b>	<b>1504</b>	<b>100</b>

In sample one of women,, the highest percentage of women in sample one was in the age group 35-39 (17.73%). For women in the second sample, the age group which recorded the highest percentage of women was 30-34 (18.61%). By comparison, sample one of women has a slightly higher percentage (5.23%) of young women (15-19) than sample two of women (4.72%). However, by holistic comparison it is seen that sample one is more characteristic of older women as opposed to women in sample two.

With regard to population group, the highest percentages were observed for 'Black/African' women in both samples (76.48% in sample one and 77.73% in sample two). The 'Indian/Asian'



population was the most unrepresented population in both samples with estimates of 1.81% and 1.60% noted for sample one and sample two respectively.

The majority of women in both samples were married. Sample one recorded 48.99% women who were married whereas sample two recorded 46.43% women who were married. Women who were never married also deserve a mention as their estimates were quite close to those of married women (i.e. 42.44% for sample one and 45.55% for sample two). The literature suggests that married women are not always the best source of such information (Hertrich (1998). Seeing that these estimates are fairly close, a point of interest in the analytical stage of the chapter would be to see if these two dichotomous marital statuses have an influence over the reporting of fertility data by women of these two samples.

In terms of level of education both samples have the majority of women who have *some schooling* (i.e. 45.15% for sample 1 and 45.55% for sample 2). Interestingly, sample two has lesser proportion of women who have little or no schooling and a higher proportion of women who are more educated (*'matric or higher'*) than sample one.

For provincial estimates, the highest percentages of women in both samples were observed in *KwaZulu-Natal*. These estimates were 19.33% and 22.61% for sample one and sample two respectively. This could be attributed to the sampling frame of the project. For sample one, the lowest percentage of women was noted in the *Free State* (5.23%) and in sample two, the lowest percentage of women was noted in the *Eastern Cape* (5.19%).

With regard to geographical location, both samples are more *urban* (just over half the sample) in nature where minute differences in percentages are observed for both urban and rural women. Seeing that these estimates, as with marital status, are very close, it would be interesting to note in the findings if there is a clear distinction between urban and rural women with regard to the reporting of fertility data (Marckwardt (1973).

With regard to LSM Status, women in both samples are predominantly from the middle income bracket (*LSM 4-6*). Of this income bracket, the higher of the two percentages are observed for women in sample one (41.24%) rather than sample two (39.63%). Furthermore, a higher

percentage (20.43%) is noted for women of sample two who belong to the lower income bracket (*LSM 1-3*) as opposed to sample one (17.87%).

For *LSM 7-10*, sample two has a higher proportion of women (12.16%). This is however closely matched by sample one (11.10%). These findings are worthy of mentioning especially since earlier it was mentioned that sample two had relatively younger women who were better educated than women in sample one. Despite this, caution should be exercised when interpreting results pertaining to *LSM*, due to the relatively small sample size across both samples.

Thus by reviewing both samples, it can be said the women of both samples are largely of middle age (although sample two records women who are relatively younger), are Black/African, married, reside predominantly in KwaZulu-Natal, with the majority of women residing in urban centers and are of the middle income category of *LSM 4-6*.

#### **4.2 Descriptive Exploratory Analysis**

The following section presents a demographic profile of women that are associated with inconsistent reporting of fertility.<sup>13</sup> Owing to the limited scope of the dissertation, only certain aspects of fertility and its reporting by women in both samples will be presented on. Variables such as '*children ever born*', '*children still alive and living in and not living in the household*' and '*children dead*' were taken into account. However other findings will be noted in brief where relevant. These variables were chosen to be analyzed as they are common to both fertility schedules of questions. *Appendix B to D* can be consulted for the questionnaires on fertility.

The analysis for this stage will be split into two parts. *Part one* will concentrate on establishing a demographic profile of women who report inconsistently. *Part two* examines the link between the socio-economic status of individuals and the aspects (i.e. sex selectiveness) of their birth histories that are misreported on.

---

<sup>13</sup> The analysis and results presented on sample one and sample two has been shaped by the manner in which the respective questionnaires were structured. In other words, there will be more results on women of sample two due to the inclusion of certain questions, hence it being called the *alternative* fertility schedule.

#### **4.2.1 Part One: Establishing a Demographic Profile of Women with Inconsistent Reporting**

With regard to the demographic profile, variables such as *'Population Group'* and *'Province'* will *not* be presented for a number of reasons. Owing to the sampling frame, a large percentage of respondents were from KwaZulu-Natal and were Black/African. The other provinces and population groups were severely underrepresented (with the exception of a few cases). Hence, if these variables are taken into account, statistically such a high percentage of respondents from one particular region and of one population group as well as its converse will create a false picture of the results of the study. Consequently, it is for this reason that *'Province'* and *'Population Group'* will *not* be considered for the descriptive analysis although provincial estimates will be reported on when considering national trends. Other background characteristics such as *'Age'*, *'Marital Status'*, *'Level of Education'*, *'Geographical Type'* and the *'Living Standard Measure'* will be considered for the analysis.

#### ***Analysis of Retrospective Reporting on Women with Live Births – Observed Irregularities in Birth Histories for Women in Sample One and Sample Two***

Table 4.2 shows the percentage of inconsistent records for the two samples of women by background characteristic. For sample one there are 37 inconsistent records, whereas for sample two there are 31 inconsistent records.

The highest percentage of inconsistent records for sample one was observed for woman aged between 40-49 years old (48.65%). For sample two of women, the highest percentages are observed for the categories 25-39 and 40-49 years of age. The lowest percentage of inconsistent records in both samples of women was recorded for the youngest age category (15-24). These estimates were 18.92% and 16.13% for sample one and sample two respectively. This finding points to the fact that older woman tend to misreport on their birth histories more frequently than their younger female counterparts.

For the category, *'Marital Status'* a difference is observed between the samples. For sample one, married women are singled out as having the highest percentage of inconsistent records for the reporting of children ever born (45.95%). For sample two, it is the women who were never married (45.16%). Perhaps the reason for this finding can be attributed to the

composition of sample two, i.e. sample two consists of younger woman than sample one and have not yet entered a marital union. Whilst this maybe true, what does need to be highlighted is the marginal difference (0.79%) between these estimates. This possibly points to the fact that marital status does not necessarily dictate whether misreporting will occur in between these groups.

**Table 4.2:  
Inconsistencies Observed for Women with Live Births by Background Characteristic**

Background Characteristic	Inconsistencies Observed between the Respective Fertility Schedule of Questions and its Roster	
	Sample 1: Census 2001 Fertility Schedule of Questions (N = 37)	Sample 2: Alternative Fertility Schedule of Questions (N = 31)
<b>Age</b>		
15-24	18.92	16.13
25-39	32.43	41.94
40-49	48.65	41.94
<b>Marital Status</b>		
Married	45.95	41.94
Never Married	40.54	45.16
Widow or Widower	8.11	9.68
Separated or Divorced	5.41	3.23
<b>Level of Education</b>		
Little or No Schooling	27.03	32.26
Some Schooling	48.65	58.06
Matric or Higher	24.32	9.68
<b>Geographical Type</b>		
Rural	56.76	51.61
Urban	43.24	48.39
<b>LSM Status</b>		
LSM 1-3	35.14	48.39
LSM 4-6	59.46	41.94
LSM 7-10	5.41	9.68

With regard to level of education, for both samples of women, the highest percentages of inconsistent records were observed for women who possessed some schooling (i.e. sample one, 48.65% and sample two, 58.06%). Although it has been established earlier that sample two consists of younger women, it is rather surprising to note that although they are younger, there is a greater percentage of these women with some educational attainment who report inconsistently than women of sample one (58.06% for sample two versus 48.65% for sample one).

For geographical location, for both samples, women who dwell in rural areas are observed to have been more inconsistent in their reporting. The estimates observed were over half of both samples (56.76% for sample one and 51.61% for sample two).

For the living standard measure, the lowest percentages of inconsistent reporting were noted for women in LSM 7-10 in both samples of women (5.41% for sample one and 9.86% for sample two). For sample one the highest percentage of women with problematic records is noted for LSM 4-6 (59.46%). For sample two, this group was LSM 1-3 (48.39%). From these findings, it can be deduced that women with a higher socio-economic status report more consistently. This can be influenced by their educational status.

Table 4.3 presents the results for inconsistent records for sample two of women for the variable '*Children Still Alive and Living In or Not Living In the Household*'. For either dichotomy there were 91 and 84 observed records respectively.<sup>14</sup>

From Table 4.3 it is observed that women of the age group 25-39 recorded the highest percent of inconsistent records for both variables (i.e. 48.35% for '*Children Still Alive and Living in the Household*' and 47.62% for '*Children Still Alive and Not Living with In the Household*'). Interestingly, for some of the remaining categories, for both samples, women with some educational attainment with an average socio-economic status are noted as having the most inconsistent records. Areas of difference is with respect to 'Marital Status' and 'Geographical Type', where women who were never married and living in rural areas reported more inconsistently for cases where children are still alive and are living in the household. For cases where the children are still alive and not living in the household, urban women who were never married reported most inconsistently. For both categories (i.e. children still alive and living in or out of the household) with respect to level of education, the highest percentage of irregular records was noted for women who possessed some schooling. For the former category a higher percentage was noted for women with little or no schooling (26.37%) as opposed to those women with higher levels of schooling (19.78%). The opposite was found for cases where children were still alive and not living in the household. Here, there was a lower percentage of

---

<sup>14</sup> There is are no results for sample one in this case, as the variable '*Children Still Alive and Living In or Not Living in the Household*' does not appear in the questionnaire (Census 2001 Fertility Schedule) administered to sample one of women.

women with little educational attainment (17.86%) and a higher percentage for women with higher levels of education (26.19%)

**Table 4.3:**  
**Inconsistencies Observed for Women with Children Still Alive and Living In and Not Living in Household**

Background Characteristic	Inconsistencies Observed for Children Still Alive and Living In and Not Living in Household – Sample 2	
	Children Living in Household (N =91 )	Children Not Living in Household (N = 84)
<b>Age</b>		
15-24	15.38	20.24
25-39	48.35	47.62
40-49	36.26	32.14
<b>Marital Status</b>		
Married	46.15	40.48
Never Married	47.25	52.38
Widow/Widower	4.40	3.57
Separated/Divorced	2.20	3.57
<b>Level of Education</b>		
Little/No Schooling	26.37	17.86
Some Schooling	53.85	55.95
Matric/Higher	19.78	26.19
<b>Geographical Type</b>		
Rural	52.75	47.62
Urban	47.25	52.38
<b>LSM Status</b>		
LSM 1-3	36.26	32.14
LSM 4-6	52.75	52.38
LSM 7-10	10.99	15.48

The most consistent records in both dichotomies were observed for young women (15-24), who have a high level of education, and are from the upper income category.

***Analysis of Retrospective Reporting on Children Dead***

Table 4.4 below highlights the inconsistencies observed for children dead for sample two of women. As this is an analysis of *observed inconsistencies* through a *direct* question on children dead, sample one is not considered for this part of the analysis as the Census 2001 Fertility Schedule of Questions administered to sample one of women did not have a direct question asking about the number of children dead. Refer to Appendix B to D for questions on fertility.

**Table 4.4:**  
**Inconsistencies Observed for Women with *Children Dead* by Background Characteristic**

Background Characteristic	Inconsistencies Observed for Children Dead between the Alternative Fertility Schedule of Questions and Roster
	Sample 2 (N = 29)
<b>Age</b>	
15-24	6.90
25-39	41.38
40-49	51.72
<b>Marital Status</b>	
Married	44.83
Never Married	41.38
Widow/Widower	6.90
Separated/Divorced	6.90
<b>Level of Education</b>	
Little/No Schooling	31.03
Some Schooling	55.17
Matric/Higher	13.79
<b>Geographical Type</b>	
Rural	48.28
Urban	51.72
<b>LSM Status</b>	
LSM 1-3	41.38
LSM 4-6	51.72
LSM 7-10	6.90

Of the 29 irregular records, Table 4.4 shows that older women (40-49) recorded the highest number of inconsistencies (51.72%). These inconsistencies occurred typically among married women (44.83%) with some educational attainment (55.17%). Women who reported inconsistently on their dead children were from the urban areas (51.72%) and were from the LSM 4-6 group (51.72%). Conversely, the profile of women who reported correctly on their dead children were of the youngest age category (15-24- 6.90%), attained a high level of education (13.79%) and were from the LSM 7-10 group (6.90%).

***Comparison of Observed Records between Sample One and Sample Two of Women***

Thus far it has been established through the descriptive analysis that the highest percentage of irregular reporting occurred among older women, women who are in a marital union (with the exception of a few cases), have some educational attainment, are from the rural area and are of

an average socio-economic status. Conversely, consistent reporting is associated with women who are young, possess a high level of education and are of the upper income group.

Given this, the question that remains is how does each sample fare when *like* is compared with *like*? That is, which sample of women (and possibly which fertility schedule), taking into account previous findings, performs better at reporting on their birth histories for the same variables on fertility. For the following graphical representations, both samples of women will be examined in three scenarios. These are explained as follows:

***Scenario One:*** *Percentage of matched records of women between the respective fertility schedule and roster.*

Scenario one examines the range (highest and lowest percentages) of matched records. A higher percentage is indicative of consistent retrospective reporting by women regarding their birth histories. For example, if the fertility schedule recorded a woman with three children ever born; for it to be consistent reporting, the roster also needs to pick up the same record, i.e. a woman with three children ever born. Conversely a low percentage of matched records of women yields the unfavorable situation of inconsistent reporting by either of the two samples of women.

***Scenario Two:*** *Percentage of records of women where the respective schedule recorded more births or deaths than the roster*

*Scenario two* presents inconsistencies in reporting of births and deaths. For this scenario, inconsistent reporting is noted when the respective schedule records more births or deaths than its roster. For example, if the fertility schedule records a woman with four children ever born and the roster records the same woman with two children ever born, it can be deduced that the schedule over reported the number of births.

***Scenario Three:*** *Percentage of records of women where the respective roster recorded more births or deaths than the schedule.*

In scenario three, records of women in the respective rosters are examined. If a roster has recorded a woman with three children ever born and the schedule notes the same woman as having only two children ever born, it can be said that the roster has over-reported the total number of children ever born to the woman. Likewise, for this record, the schedule has under-reported the total number of children ever born.



Hence, the above three scenarios point to how consistent and inconsistent reporting of birth histories occur. Of equal importance and to re-iterate, the objective of the analysis is not to determine which fertility schedule was more efficient at recording retrospective birth histories, but rather to determine factors associated with inconsistent reporting by two samples of women to whom different fertility questionnaires were administered.

**Comparison of Records for Live Births**

**Figure 4.1:**

**Comparison of Records for *Live Births* for Sample One and Sample Two of Women – National Trend**

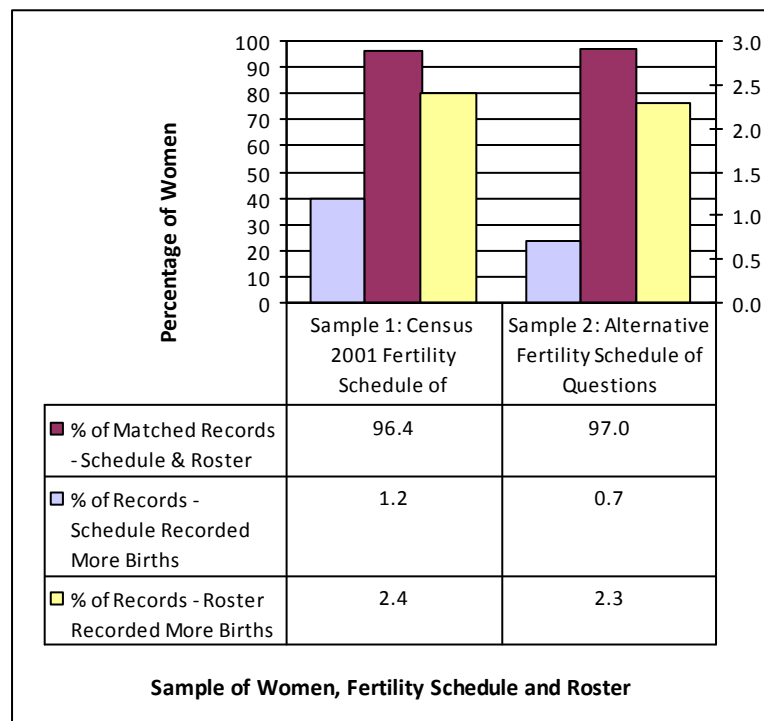


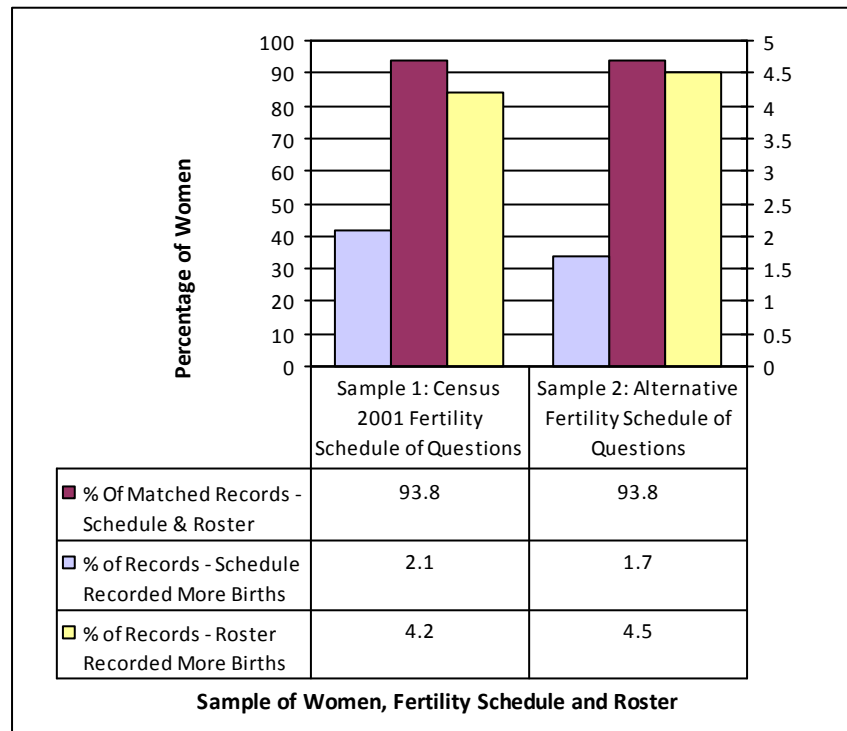
Figure 4.1 shows that at the national level, for sample one of women, there was 96.4% of matched records between the Census 2001 fertility schedule of questions and its roster. The estimate for sample two of women is marginally higher (97.0%). For cases where the respective schedule recorded more births, a higher percentage was noted for sample one (1.2%) than sample two (0.7%). There were no significant differences observed between the relative percentage of women of both samples for whom the roster recorded more births than the schedule (2.4% sample one and 2.3% for sample two).

From the above findings it can be said that sample two of women reported more consistently (i.e. a higher percentage of matched records between the schedule and the roster) than women

in sample one. This could also be linked to the composition of each sample as well as the types of questions asked in the alternative fertility schedule of questions, hence ensuring better quality of reporting.

**Comparison of Records for Boy Children Born Alive**

**Figure 4.2:  
Comparison of Records for Boy Children Born Alive for Sample One and Sample Two of Women – National Trend**



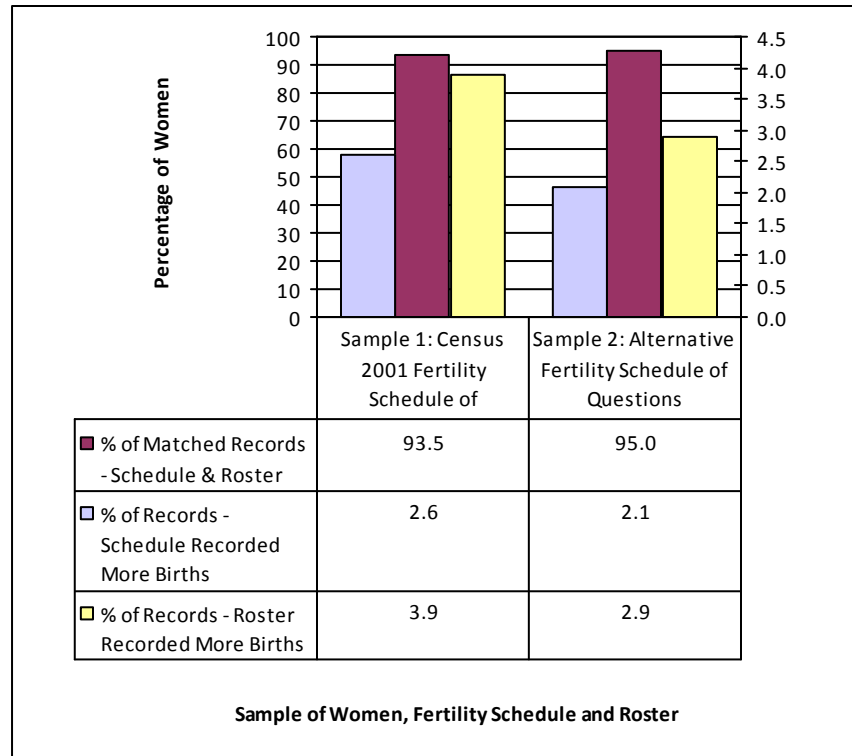
From Figure 4.2, it appears that at the national level, across all three categories of comparison there are no significant differences observed for women with regard to their reporting of boy children born alive. Despite this, there is a slight leniency towards the alternative fertility schedule of questions (i.e. recorded a lower percentage of inconsistent reporting in cases where the schedule recorded more births - 1.7% versus 2.1%).

Seven provinces fall in line with this national trend. These provinces are Western Cape, Eastern Cape, Northern Cape, KwaZulu-Natal, North West, Gauteng and Limpopo. These provinces are a mix of provinces (geographical location, income and population group). As a result of this majority (based on observed records only) it seems as though women of sample two reported

more consistently (higher percentages of matched records and a lower percentage where the schedule recorded more births than the roster) on their boy children born alive.

**Comparison of Records for Girl Children Born Alive**

**Figure 4.3:**  
**Comparison of Records for Girl Children Born Alive for Sample One and Sample Two of Women**  
**– National Trend**



In Figure 4.3 for the category of ‘% of Matched Records’ it is seen that women of sample two through the alternative fertility schedule captured 95% of matched records between the schedule. However for the second category, where the schedule recorded more births, a higher percentage of women reported incorrectly on their girl children born alive was noted for sample one (2.6%) than sample two (2.1%). With regard to the roster recording more births, the same pattern is observed. (i.e. the Census 2001 fertility schedule measured a relatively higher proportion of women (3.9%) from sample one than those captured by the alternative fertility schedule for sample two (2.9%)).

### **Comparison of Records for Children Still Alive**

**Figure 4.4:  
Comparison of Records for *Children Still Alive* for Sample One and Sample Two of Women –  
National Trend**

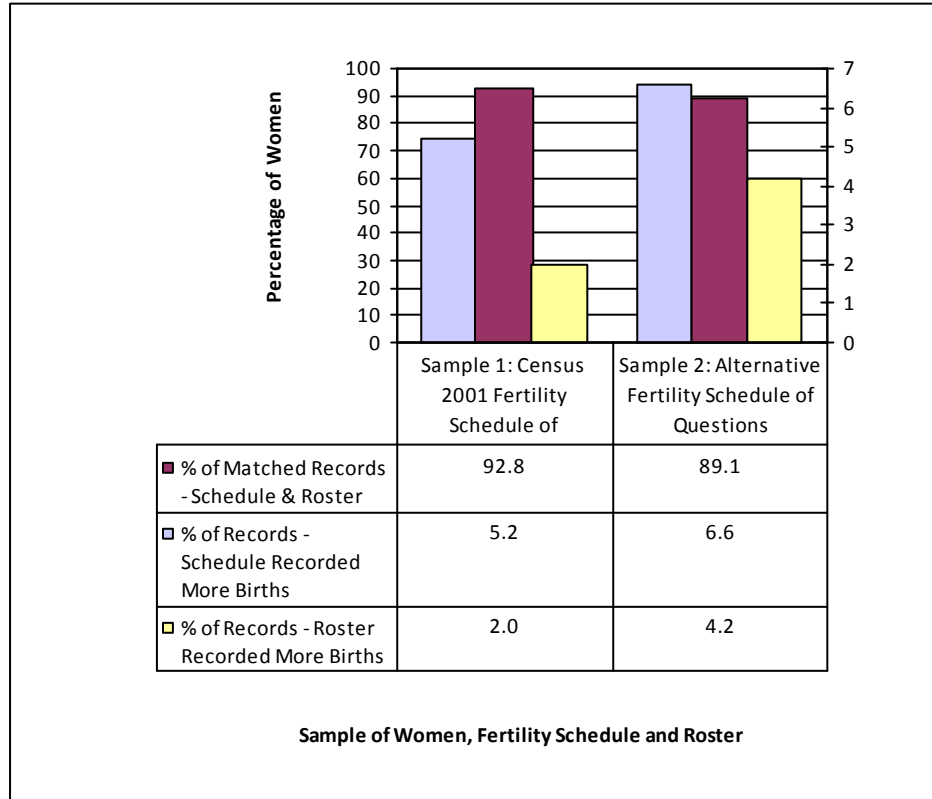


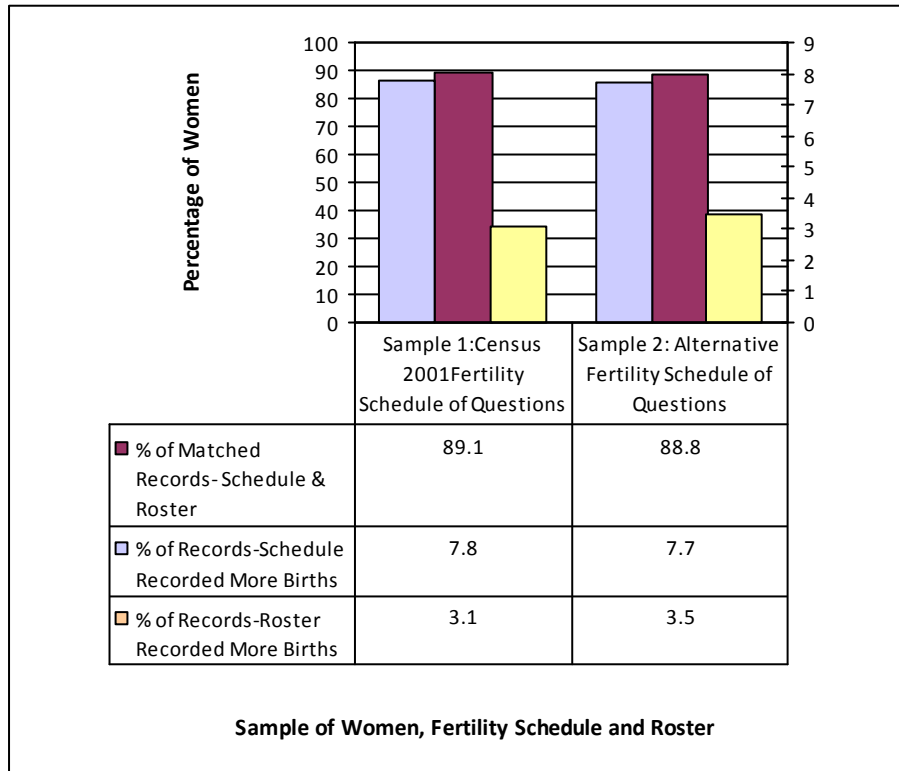
Figure 4.4 shows that women of sample two (89.1%) recorded a lower percentage of matched records than women of sample one (92.8%). The same trend (i.e. women of sample two recorded more inconsistencies) is also evident in the remaining categories where either the schedule or the roster recorded more births.

The failure of women of sample two to report on their fertility consistently could be linked to their demographic profile. On the other hand, perhaps their demographic profile had little or no bearing on the manner in which they reported. To support this argument, women of sample two were administered the alternative fertility schedule. As seen in *Appendix B to D* the questions on children still alive differed between the alternative schedule of fertility questions and the Census 2001 Fertility Schedule of Questions. Perhaps the ‘double barreled’ nature (i.e. asking if the children were still alive *and* whether or not the children reside in the household) of the question on children still alive in the alternative schedule of questions may have had an effect on the

manner in which women reported. Likewise, the phrasing of the same question in the Census 2001 Fertility Schedule of Questions (i.e. 'If the person has ever given birth to any live birth/births: If boys: How many are still alive? If girls: How many are still alive?') may have produced a similar effect on women of sample one

**Comparison of Records for Boy Children Still Alive**

**Figure 4.5:**  
**Comparison of Records for Boy Children Still Alive for Sample One and Sample Two of Women – National Trend**

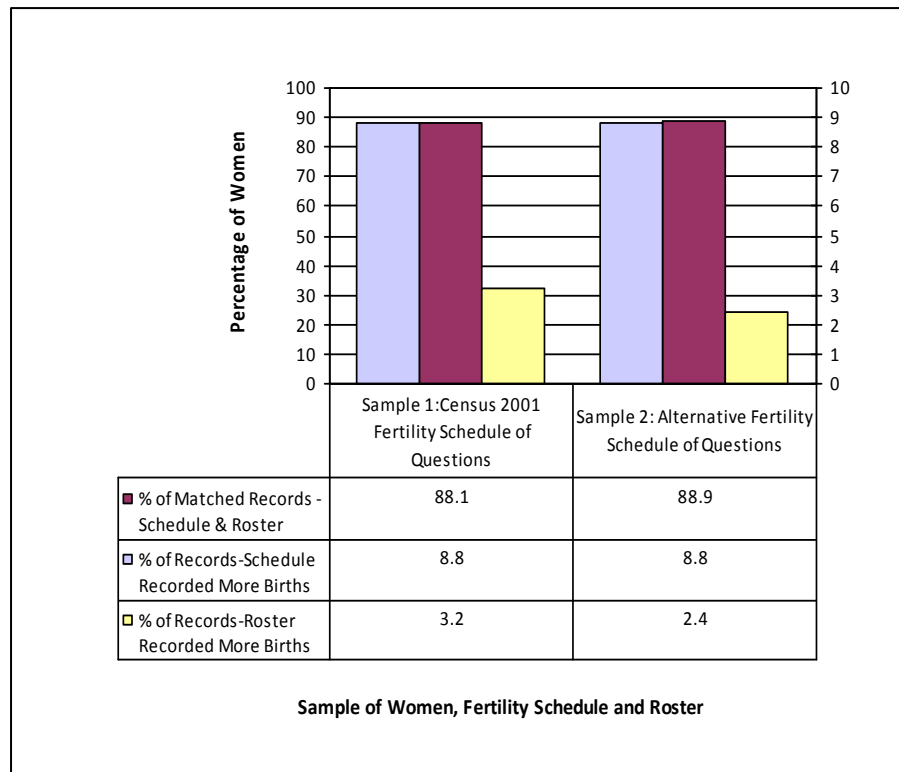


In Figure 4.5, the pattern observed nationally through all three categories of comparison is that no significant difference is observed for the percentage of women with boys still alive between the Census 2001 fertility schedule (sample one) and the alternative fertility schedule (sample two). For the first category of comparison, a higher percentage of women was recorded for sample one through the Census 2001 fertility schedule (89.1%) than for sample two through the alternative schedule (88.8%). In cases where the schedule recorded more births, a higher percentage was noted for sample one (7.8%). For cases where the roster recorded more births, a higher proportion was noted sample two (3.5%).

Furthermore, five of the nine provinces (Western Cape, Eastern Cape, KwaZulu-Natal, North West and Gauteng) favoured the alternative fertility schedule by recording a higher percentage of matched records. Thus, it appears that sample two of women have reported more accurately (i.e. less irregularities captured) on their birth histories than women of sample one.

### **Comparison of Records for Girl Children Still Alive**

**Figure 4.6:  
Comparison of Records for *Girl Children Still Alive* for Sample One and Sample Two of Women  
– National Trend**



In Figure 4.6, for the first category of comparison, although no significant difference is observed between the schedules, it is seen that sample two through the alternative fertility schedule (88.9%) recorded a higher proportion of matched records as opposed to sample one through the Census 2001 Fertility Schedule (88.1%).

For cases where the schedule recorded more births, we see the same proportion (8.8%) of women for both samples of women. For the third category (*'Roster Recorded More Births'*) higher proportions of women (hence inconsistent reporting) were noted for sample two (2.4%) rather than sample one (3.2%).

### **Comparison of Records for Children Dead**

**Figure 4.7:  
Comparison of Records for *Children Dead* for Sample One and Sample Two of Women –  
National Trend**

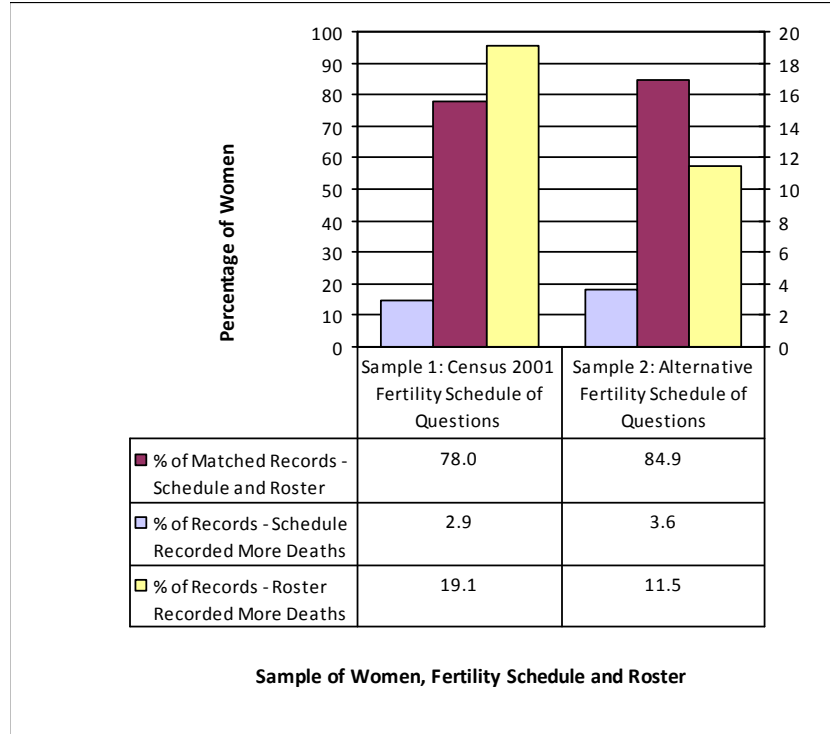


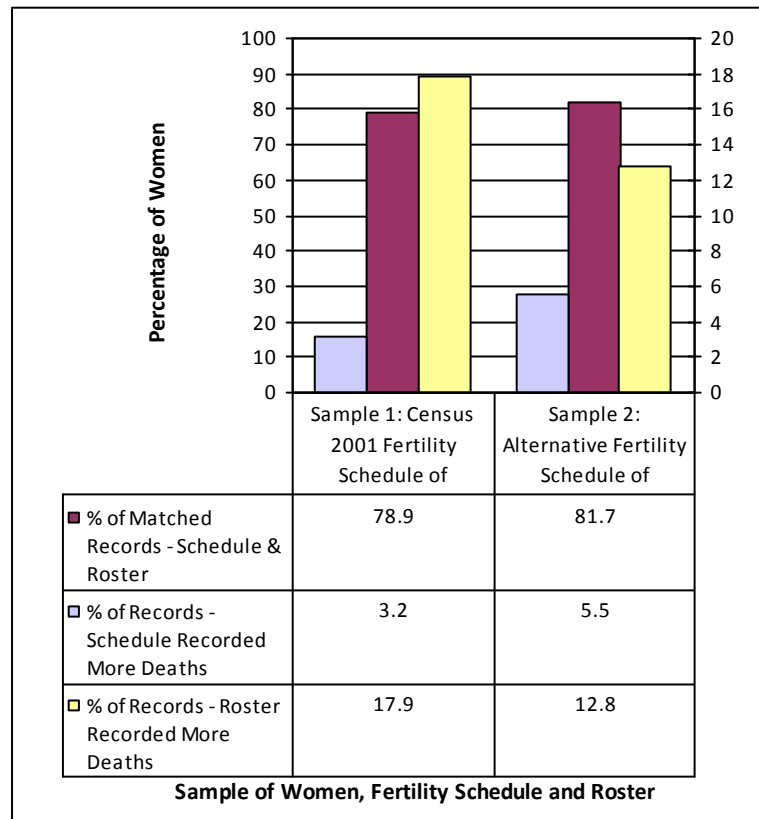
Figure 4.7 illustrates a comparative analysis of records for children dead for sample one and sample two of women at the national level. Sample two of women reported more consistently (higher percentage of matched records) than women of sample one (84.8% versus 78.0%). For cases where the schedule recorded more deaths than the roster, a higher percentage was noted for women of sample two than sample one (3.6% versus 2.9%). In the third case (*'% of records-roster recorded more deaths'*) women from sample one had a significantly higher percentage than sample two (19.1% versus 11.5%) hence indicative that women of sample one reported more inconsistently in this case.

From Figure 4.7 it can be deduced that women of sample one reported more inconsistently than their counterparts. This could be attributed to their background characteristics or due to the questions on birth histories in the Census 2001 Fertility Schedule of Questions and the manner they were phrased. There may have been some confusion during the administration of the

questionnaire, as was the case at the time of Census 2001 (i.e. misunderstanding of questions) (Moultrie and Dorrington, 2004).

### **Comparison of Records for Boy Children Dead**

**Figure 4.8:  
Comparison of Records for *Boy Children Dead* for Sample One and Sample Two of Women –  
National Trend**

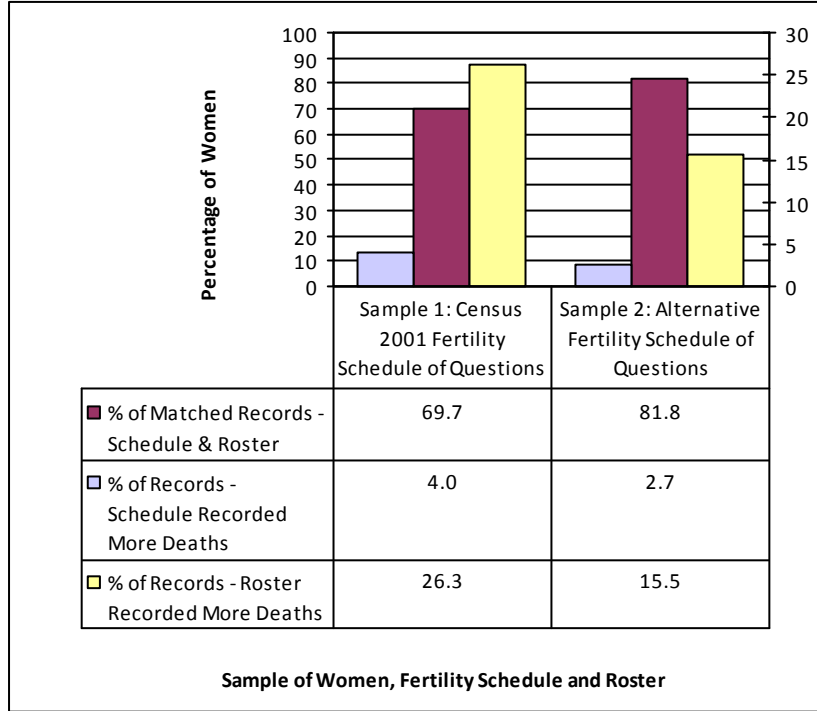


In Figure 4.8, a considerable difference is noted between the percentages of women in all three categories of comparison. Of the two samples, women in sample one recorded a lower percentage of matched records between the schedule and roster (78.9%). Although not significantly higher, sample two managed to record only 81.7% of consistent reports by respondents. Where the schedule was found to record more deaths, there were more records observed for women of sample two (5.5%) than women of sample one (3.2%). For the third category, where the roster recorded more deaths, sample two of women (12.8%) reported more consistently on their girl children dead as opposed to sample one (17.9%).



### **Comparison of Records for Girl Children Dead**

**Figure 4.9:  
Comparison of Records for *Girl Children Dead* for Sample One and Sample Two of Women –  
National Trend**



From Figure 4.9, we see that a significantly higher percentage of matched records are noted for sample two of women (81.8%) as opposed to women in sample one (69.7%). For the second category of comparison (*'Schedule Recorded More Deaths'*) a lower percentage of women was recorded for sample two (2.7%) than for sample one (4.0%). Where the roster recorded more deaths, a higher percentage was observed for sample one through the Census 2001 schedule (26.3%) than sample two through the alternative fertility schedule (15.5%).

By viewing the overall provincial estimates for dead girl children, the Western Cape, Eastern Cape, Northern Cape, KwaZulu-Natal, Gauteng and Limpopo all showed higher percentages of women with girl children dead for sample two of women than sample one. Thus it can be said that sample two of women reported more consistently than sample one.

#### **4.3 Summary of Part One**

*Part one* has shown that sample two of women has reported more consistently and accurately on their birth histories. This finding is supported by the high percentage of matched records

(births and deaths) between the respective fertility schedules and rosters. Furthermore, provincial trends also indicate that sample two of women have maintained a consistent level of reporting for the majority of the fertility differentials used. To re-iterate, sample two consist largely of younger women, are Black/African, married, reside predominantly in KwaZulu-Natal, with the majority of women residing in urban centers and are of the middle income category of LSM 4-6. At this stage it thus appears as though age, education, geographical location and LSM status has an impact on the manner in which respondents report on their fertility.

#### ***4.4 Part Two: Establishing a Link between the Socio-Economic Status of Individuals and Retrospective Reporting of Birth Histories***

The following section aims to highlight the extent of misreporting by examining the socio-economic status of women in both samples. The analysis thus far has established that older, married women with some educational attainment, of rural areas from either the middle and lower income categories tend to misreport more frequently than their converse counterparts.

The second part will examine and attempt to determine a link between the socio-economic status of an individual and aspects of inconsistent reporting. In other words, the results will look particularly at the sex-selectiveness (i.e. births and deaths) of reporting by women of sample one and two. For the study I have chosen the Living Standard Measure and Geographical Location to comprise the socio-economic status of individuals. The Living Standard Measure was chosen as a proxy for socio-economic status. By implication, the higher the LSM status, the higher the level of socio-economic status and vice versa<sup>15</sup>. Geographical location was chosen as an extension of socio-economic status by virtue of the advantages and disadvantages of each location and how these concomitant benefits of either location can influence reporting.

##### ***4.4.1 Theoretical Expectations for Socio-Economic Status of Individuals and Retrospective Reporting: Living Standard Measure***

With regard to retrospective reporting and its link to LSM status it is expected that the most consistent reporting (i.e. (1) a high percentage of matched records between the schedule and the roster and (2) the lowest possible percentages in cases where either the schedule or the roster recorded more births or deaths) is likely to be observed within the upper income group

---

<sup>15</sup> Refer to Appendix G on the LSM for further detail

(LSM 7-10). It is also expected that satisfactory levels of consistent reporting may extend to women in LSM 4-6 group. Women who belong to the middle and upper income LSM groups higher levels education, income and material welfare. It is hence postulated, that due to their composition, women of these groups are able to comprehend and understand (with little or no assistance from the enumerator) the questions posed to them on their birth histories. They are more likely to report on their birth history *correctly* and *consistently*.

The most inconsistent reporting (i.e. (1) low percentages of matched records between the schedule and the roster and (2) high percentages where either the schedule or roster recorded more births or deaths) is most likely to be observed in the low income group of LSM 1-3. Reason being, these groups are defined by lower income and education levels. From past research investigations (Beckett *et al.* (2001), it has been argued and established by some studies that errors in reporting typically originate from this profile of women. Due to their low educational status and lack of understanding of the questions being asked, women of this group are possibly unable to answer on their birth histories without the assistance of an enumerator. Misreporting in this case can occur through misunderstanding by the respondent due to the incorrect explanation by the enumerator. As a result, although the response is perceived to be correct, it is in fact incorrect and inconsistent.

#### **4.4.2 Theoretical Expectations for Socio-Economic Status of Individuals and Retrospective Reporting: Geographical Location**

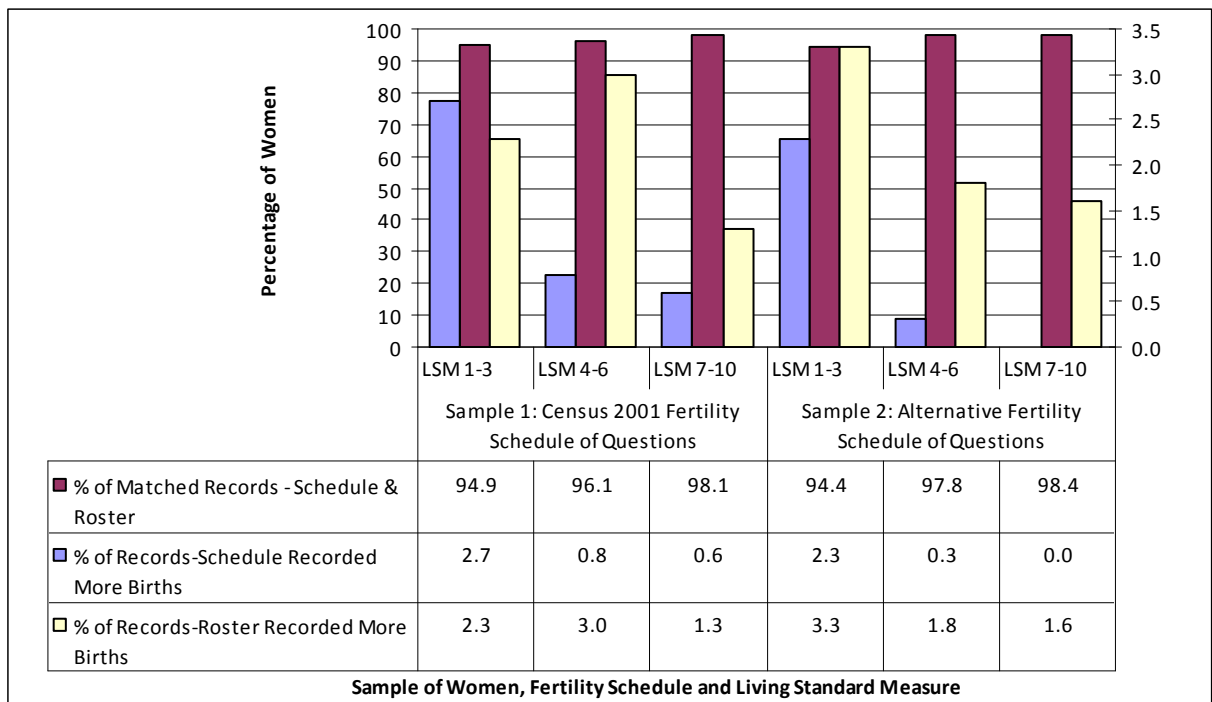
It is postulated that women in urban areas will report more consistently than their rural counterparts. Reason being, urban women benefit from the provision of health care infrastructure, services and access to health care information. As a result, women in urban area are more likely to register their births and deaths. Consequently, urban women are able to report on their birth histories more correctly. That said, there are cases where urban women do not register these vital events and thus can also report incorrectly on their fertility. This would be the exception rather than the rule.

Rural women find it difficult to visit such places given the time and distance involved, which may often eat into their daily income activities (Ndong, Gloyd and Gale, 1994). Traveling to urban areas for health care or to register a birth or death is a costly affair and women do not

necessarily have the funds (whether their own or from the household) to make these trips which needs to be done at regular intervals (Ndong et al. 1994). Furthermore, in cases where a birth or death has to be registered the complicated registration procedure, lack of knowledge and the perception of no benefit inhibits registration (Ndong et al. 1994). Without the necessary documentation, rural women (with the exception of some) are unable to report correctly on their birth history.

**Comparison of Records for Live Births and Socio-Economic Status for Sample One and Sample Two of Women**

**Figure 4.10:**  
**Comparison of Records for *Live Births* for Sample One and Sample Two of Women by Living Standard Measure**



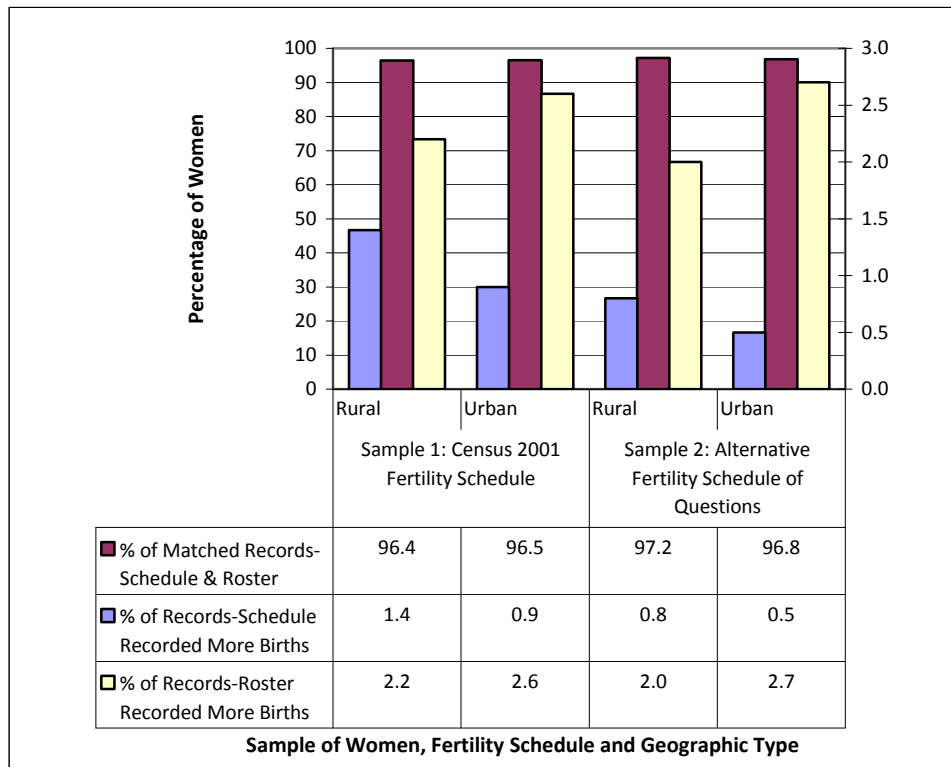
In Figure 4.10, the highest percentage of women in the category of '*% of Matched Records*' in both samples was noted in the LSM 7-10 group (i.e. 98.1% for sample one, 98.4% for sample two). For sample one of women we see that the highest percentage of women with live births for whom the schedule showed more births than the roster occurred amongst the LSM 1-3 group (2.7%). The same is observed for sample two (2.3%). With respect to the rosters recording more births than the schedule, a different pattern is noted. For sample one, LSM 4-6 group had

the highest percentage of women (3.0%), whilst for sample two, it was the LSM 1-3 group (3.3%).

In Figure 4.10, it is seen that the highest percentages of women for cases that recorded matched records occur in the LSM 7-10 group for both sample one (98.1%) and sample two (98.4%). This pattern (LSM 7-10 group records more consistently) is evident in the remaining categories (i.e. low percentages of where either the schedule or roster records more births).

These findings correlate with our initial expectations. It can thus be said that women who are of a higher socio-economic status tend to report more accurately on their fertility owing to their higher level of education and other inter-playing factors that influence their socio-economic status.

**Figure 4.11:**  
**Comparison of Records for Live Births for Sample One and Sample Two of Women**  
**By Geographical Type**

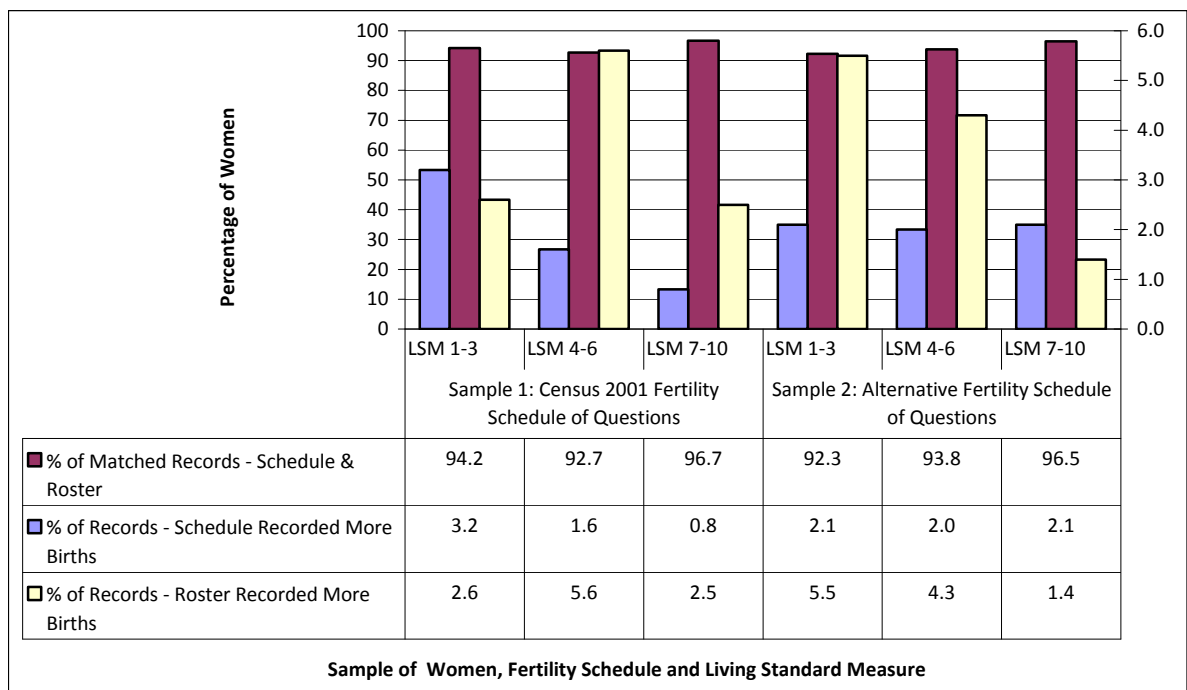


As seen in Figure 4.11, for the first category of comparison ('% of Matched Records') the most consistent reporting was found among the rural group (97.2%) for sample two. For sample one

in the Census 2001 Fertility Schedule there was not a significant difference between urban (96.4%) and rural women (96.5%). For sample one, we see that the highest percentage of women with live births for whom the schedule showed more births than the roster occurred amongst the rural group (1.4%). The same is observed for sample two (0.8%). With respect to the rosters recording more births than the schedule, urban women showed a relatively higher percentage (2.6%) in sample one **than** sample two (2.7%), hence indicating inconsistency in reporting.

An important point reflected in the graph is that a higher proportion of matched records of women with live births were observed for rural women as opposed to urban women. Although not consistent with initial expectations, such a finding is indeed welcome.

**Figure 4.12:**  
**Comparison of Records for *Boy Children Born Alive* for Sample One and Sample Two of Women by Living Standard Measure**

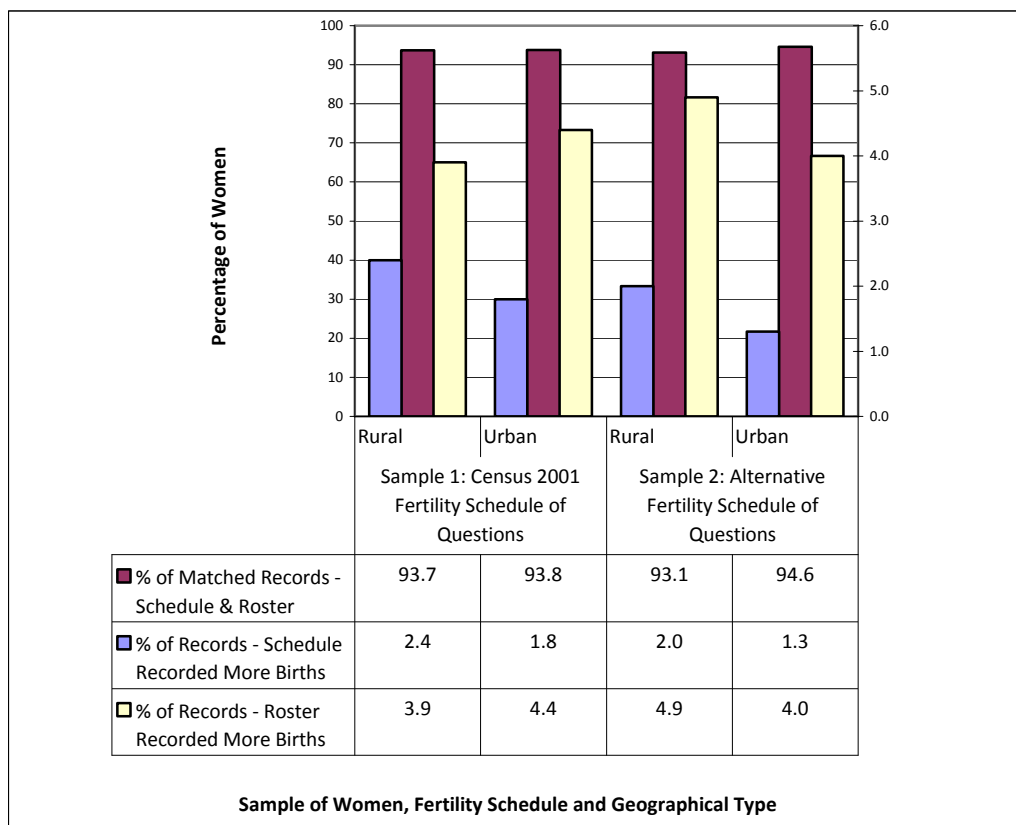


For the category '*% of Matched Records*' the most consistent reporting for sample one occurred in the LSM 7-10 group (96.7%). The same is noted for the women in sample two (96.5%). For cases where the schedule recorded more births, the highest proportion (i.e. inconsistent reporting) of women in sample one was measured for the LSM 1-3 group (3.2%). For sample two, the case is not clear as there is no significant difference in the proportions observed across

the LSM groups. For records where the roster recorded more births, for sample one and the Census 2001 fertility schedule, the LSM 4-6 group had the highest proportion of women (5.6%), whilst for the women of sample two and the alternative fertility schedule it was the LSM 1-3 group (5.5%).

It is encouraging to note that the most consistent reporting in both samples occurred in the LSM 7-10 group. Mixed results were observed for inconsistent reporting. Although inconsistencies were found in both the LSM 1-3 and LSM 4-6 group, it does however indicate that socio-economic status does have an impact on retrospective reporting.

**Figure 4.13:**  
**Comparison of Records for *Boy Children Born Alive* for Sample One and Sample Two of Women by Geographical Type**



In Figure 4.13, for sample one of women through the administration of the Census 2001 Fertility Schedule, the percentages of women recorded for the category '*% of Matched Records*' did not differ significantly amongst rural and urban women (93.7% for rural women and 93.8% for urban

women). Hence the level of reporting consistently or inconsistently is spread across these schedules and thus cannot be determined easily through observed records. For sample two, urban women showed a relatively higher proportion (94.6%) than rural women (93.1%).

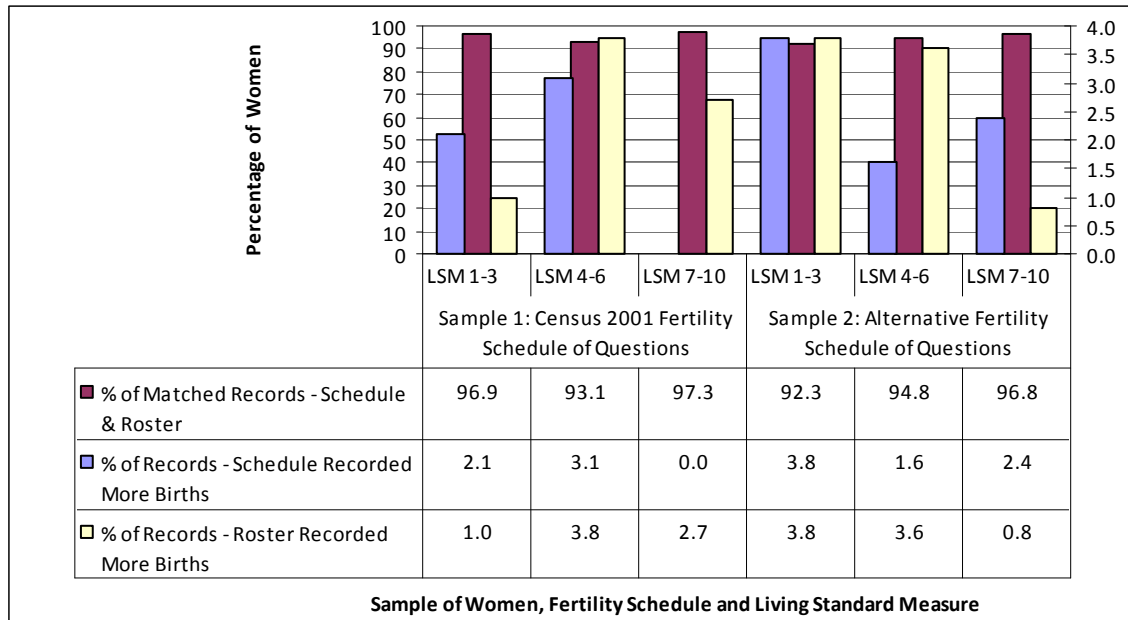
For cases where the schedule recorded more births, rural women in both samples showed a relatively higher proportion than their urban counterparts. For the third category of comparison ('Roster Recorded More Births') higher percentages of urban women (4.4%) was measured in sample one. For sample two, it was rural women (4.9%).

Whilst there may be no clear distinction as to whether rural or urban women report consistently in the first category of comparison, rural women are singled out as having a higher proportion for cases where the schedule recorded more births. Although urban women (4.4%) in sample one is observed to have reported more inconsistently in the recording of births, this estimate is lower than that of rural women of sample two (4.9%). It can thus be concluded that rural women of sample two based on this comparison have reported inconsistently on their birth histories.

In Figure 4.14 for the first category ('% of Matched Records') the highest proportion of women with consistent reporting on their girl children born alive occurred in sample one women in the LSM 7-10 group (97.3%).) The same was observed for sample two of women. For sample one, in the second category of comparison, women who reported inconsistently was from the middle income group (LSM 4-6 group - 3.1%). In sample two, women who reported inconsistently, were from the lower income group, LSM 1-3 (3.8%). For the third category ('Roster Recorded More Births') the highest proportion of women who reported inconsistently on their girl children born alive are observed for the LSM 4-6 group (3.8%). Whereas for sample two, it was women in the LSM 1-3 group (3.8%).



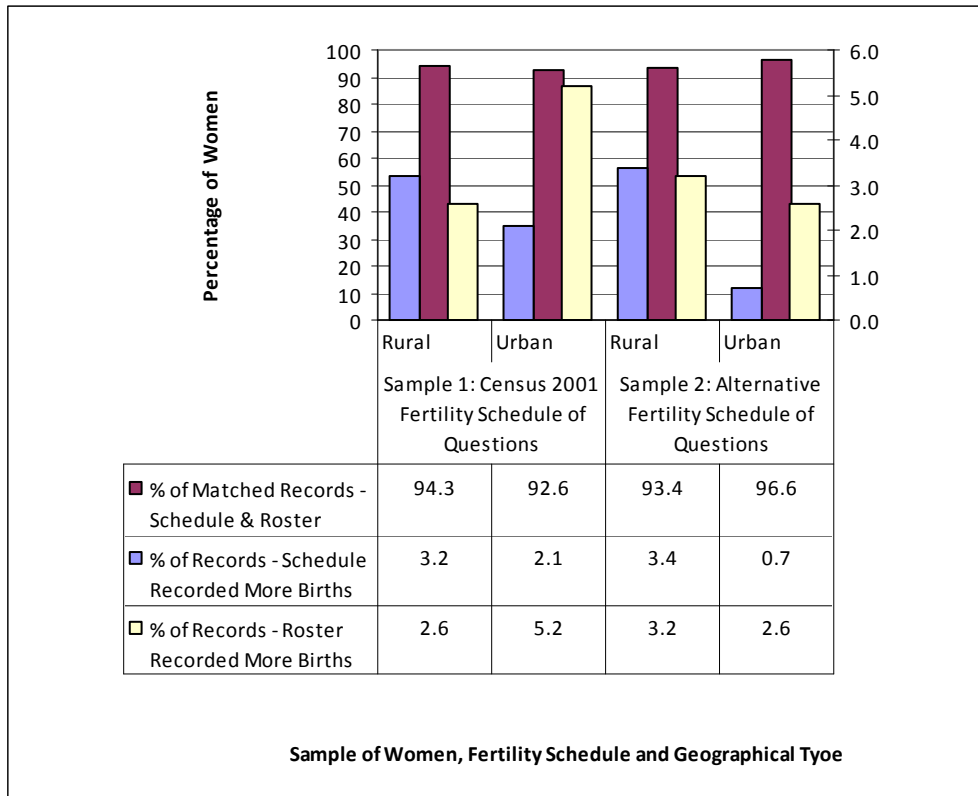
**Figure 4.14:**  
**Comparison of Records for *Girl Children Born Alive* for Sample One and Sample Two of Women by Living Standard Measure**



As with previous findings and expectations outlined earlier, Figure 4.14 has confirmed that women of LSM 7-10 tend to report more correctly on their fertility than women of the lower LSM groupings.

Figure 4.15 shows that rural woman in sample one reported more consistently (94.3%) than their urban counterparts (92.6%). For sample two, urban women (96.6%) reported more consistently than rural women in sample two (93.4%). With respect to the schedule recording more births, rural women in both samples showed a relatively higher proportion than urban women. In the third category of comparison urban women in sample one reported more inconsistently (5.2%) whilst rural women in sample two demonstrated the same pattern (3.2%).

**Figure 4.15:**  
**Comparison of Records for Girl Children Born Alive for Sample One and Sample Two of Women**  
**by Geographical Type**



**4.4.3 Comparison of Records for Dead Children and Socio-Economic Status for Sample One and Sample Two of Women**

With regard to the proportion of children dead, conventional demographic wisdom indicates that we would expect to see higher proportions of dead children in the LSM 1-3 group and possibly the LSM 4-6 group. Reason being, this is a poorer, less educated population, it is likely that this group would report a higher number of dead children. We should thus anticipate higher percentages of matched records in these groups.

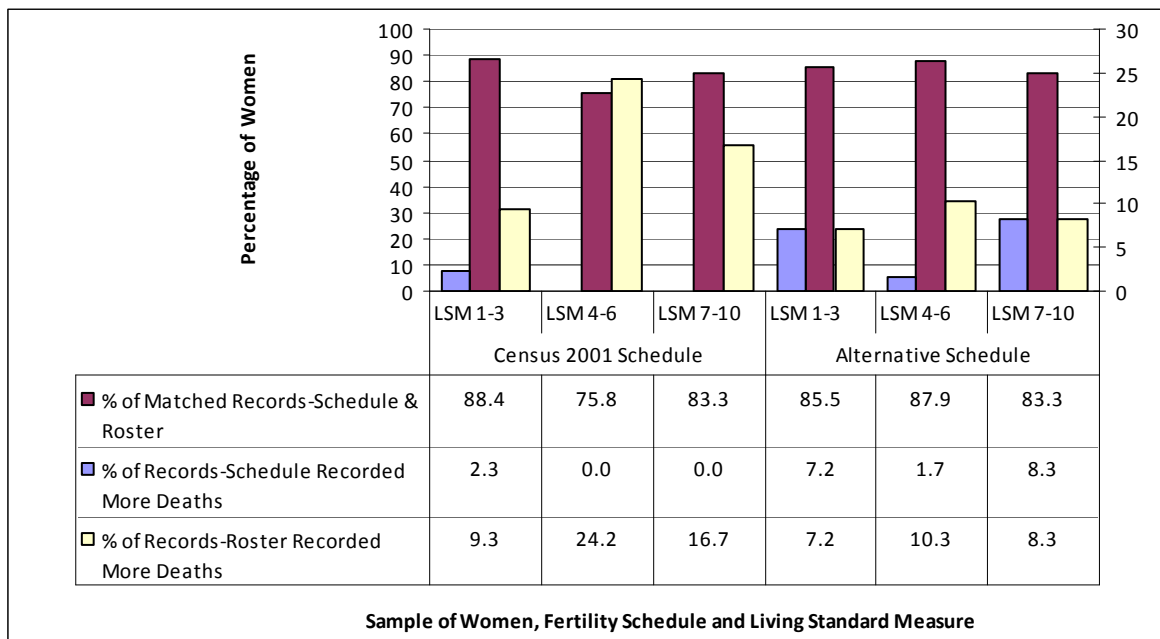
Consequently, it is expected that low proportions of women with dead children are to occur in the LSM 7-10 group. Reason being, this population is characterized by well-educated, higher income earners with greater access to medical and health care services and facilities. Thus the probability of women of this LSM reporting dead children is low.<sup>16</sup>

<sup>16</sup> The percentages in the higher LSM groups should be read with caution due to low observed counts and the small

Furthermore, it is expected that rural women will report more inconsistently on their dead children especially with regard to dead girl children. Literature has often alluded to rural women creating a deficit in the number of dead girl children as opposed to dead boy children (Blacker and Brass, 1979).

Figure 4.16 shows that for the first category of comparison for sample one, the highest proportion of women with dead children was recorded for the LSM 1-3 group (88.4%), whereas in sample two, the highest proportion of women was noted for the LSM 4-6 group (87.9%). With respect to the rosters recording more deaths than the schedule, we see that the LSM 4-6 group in both samples had the highest proportion of women (24.2% and 10.3% respectively).

**Figure 4.16:**  
**Comparison of Records for *Dead Children* for Sample One and Sample Two of Women by Living Standard Measure**



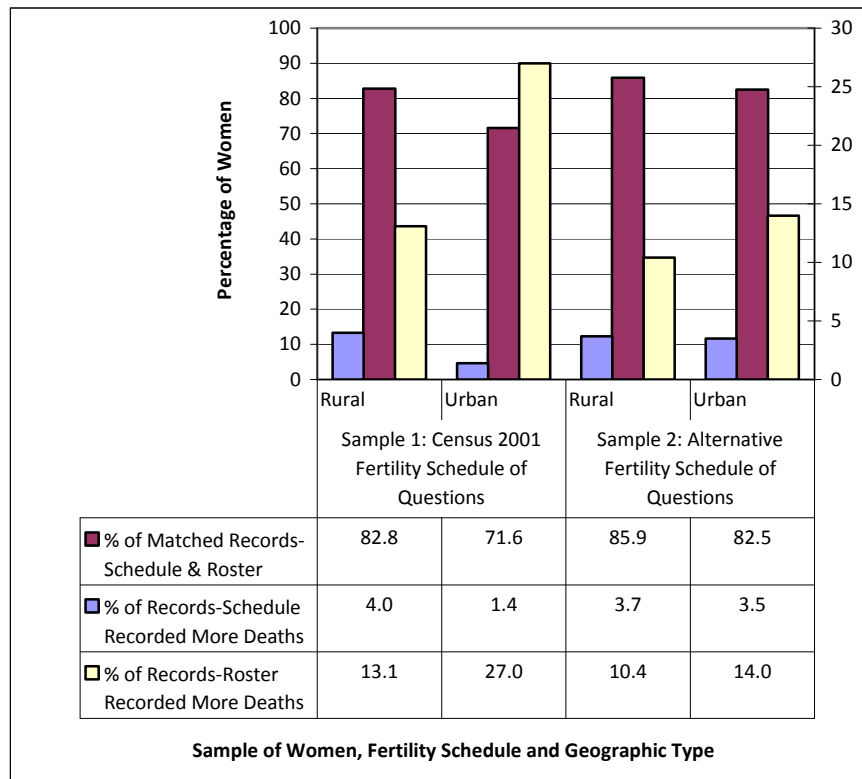
As seen in Figure 4.16 for sample one we see that there is indeed a higher percentage of women with matched records with dead children in the LSM 1-3 (88.4%) and thus confirms our assumptions made earlier regarding the proportion of women with dead children and lower LSM groups. The lowest proportion of women for sample two is noted in the LSM 7-10 group (83.3%). This supports the assumptions made above with regard to the composition of the LSM

sample size.

7-10 group and the proportion of women with dead children. Whilst this is the case, the percentages in the LSM 7-10 groups must be approached with caution as they have low observed counts which results in inflated percentages.

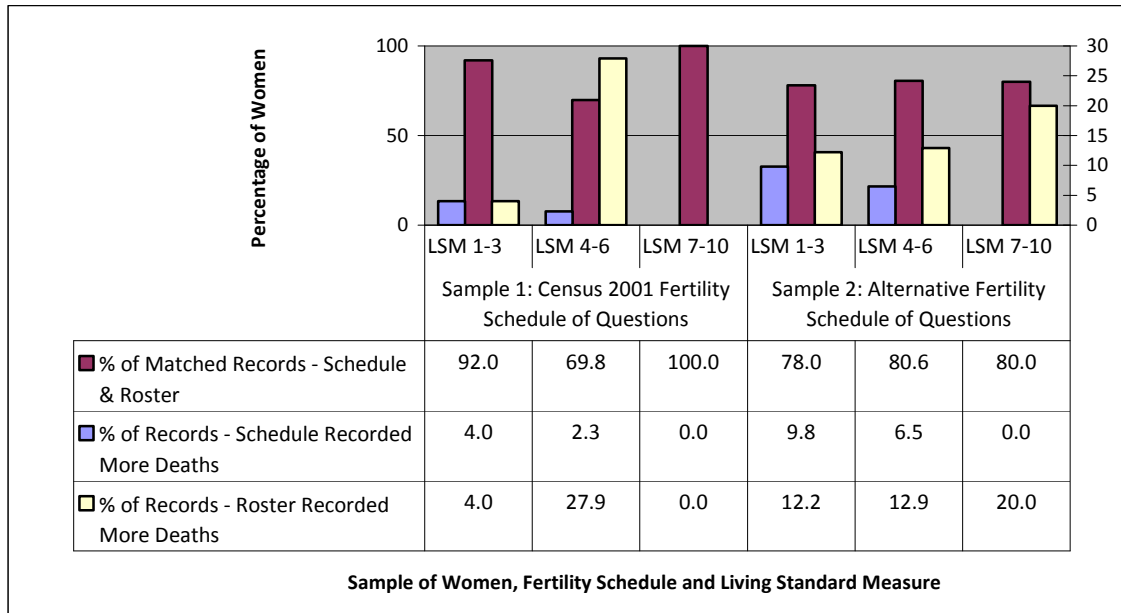
Figure 4.17 shows that for both samples, a higher percentage of matched records are observed for rural women (82.8% and 85.9% respectively). For the second category of comparison (*'Schedule recorded more deaths'*) a higher proportion of women with child deaths was noted for rural women (4.0%). No significant difference was observed between the rural and urban groups of women from the sample two (3.7% versus 3.5%). In terms of cases where the rosters recorded more deaths than the schedules, for sample one, urban women with dead children were twice that of rural women (27.0% and 13.1% respectively).

**Figure 4.17:**  
**Comparison of Records for *Dead Children* Sample One and Sample Two of Women by Geographical Type**



Thus Figure 4.17 provides empirical evidence for the theoretical considerations made earlier (i.e. higher proportions of women with matched records with dead children for rural women and lower proportions of women with dead children in urban areas).

**Figure 4.18:**  
**Comparison of Records for *Boy Children Dead* for Sample One and Sample Two of Women by Living Standard Measure**



The percentages observed in Figure 4.18 should be approached with caution due to the low relative counts of women in these LSM groups. This statement is made with direct reference to LSM 7-10 in both samples and will hence not be considered for analysis.

In Figure 4.18, it is seen that for sample two of women, there were 92.0% and 69.8% matched records for sample one for the group LSM 1-3 and LSM 4-6 respectively. In sample two the highest percent of matched records were from the LSM 4-6 group. In the second category of comparison (*'Schedule Recorded More Deaths'*), the highest percentages were recorded for women in the LSM 1-3 group for both samples of women (i.e. 4.0% and 9.8% respectively). Where the roster recorded more deaths, the highest proportion of women measured for women in sample one was the LSM 4-6 group (27.9%).

Due to low numbers of observed counts resulting in inflated percentages hence reflective of the small sample size, substantial conclusions cannot be drawn from these proportions due to this statistical inaccuracy.

**Figure 4.19:**  
**Comparison of Records for *Boy Children Dead* for Sample One and Sample Two of Women by Geographical Type**

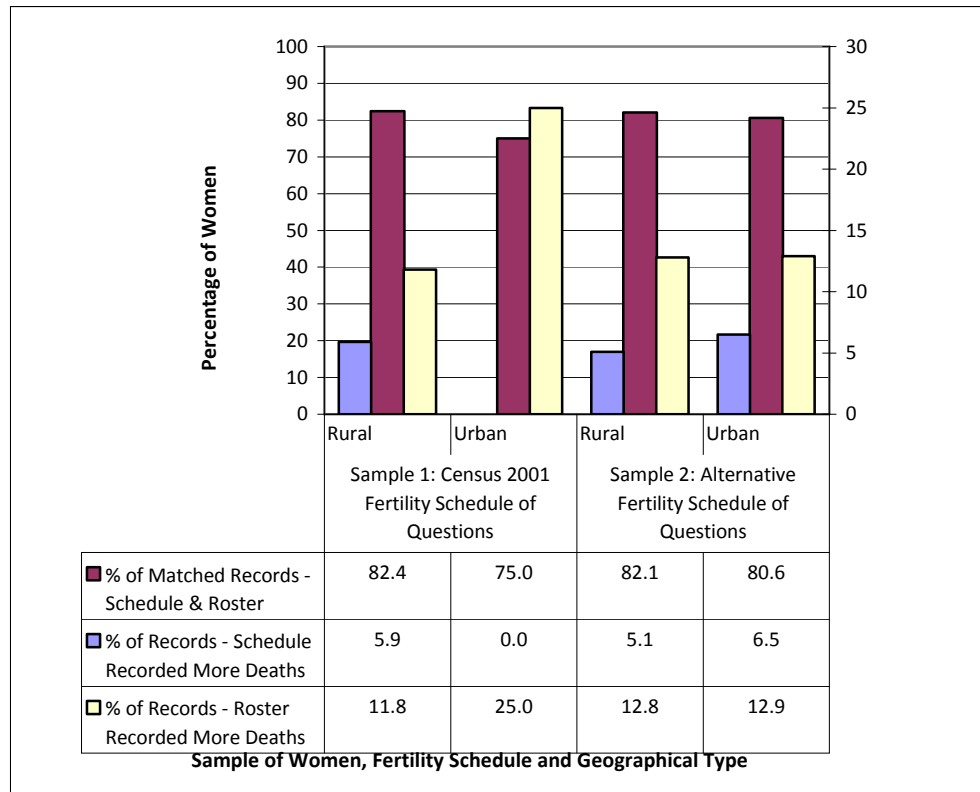


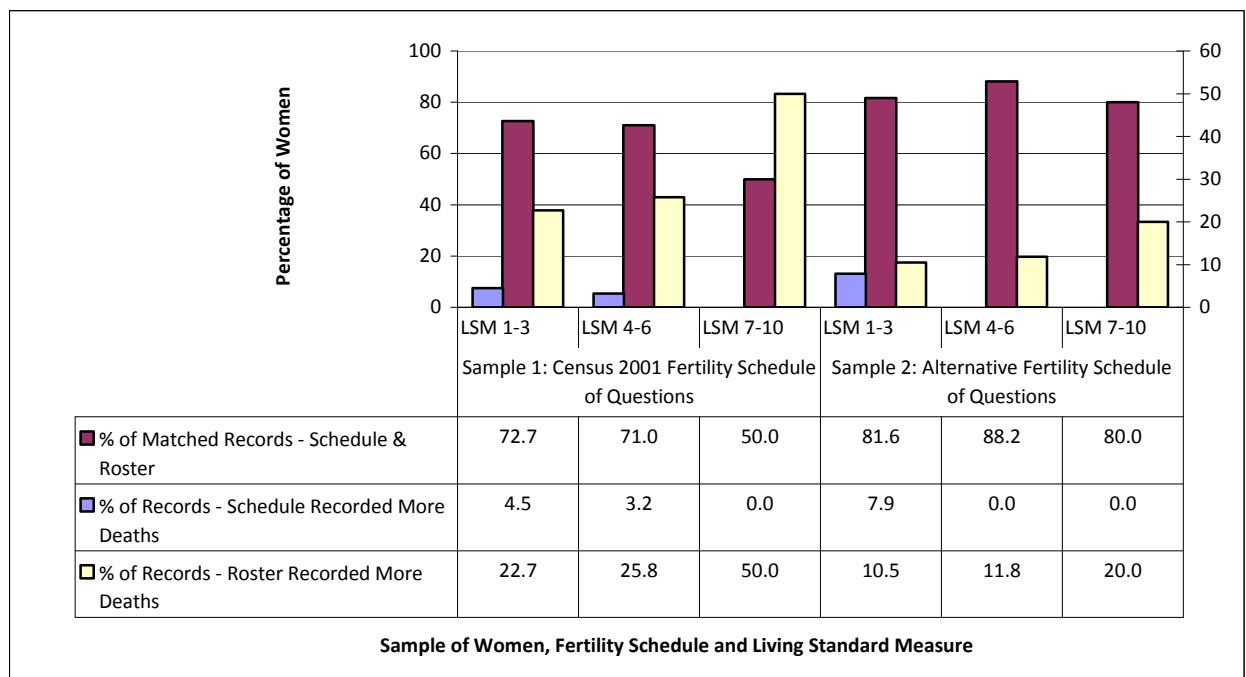
Figure 4.19 shows that with respect to cases where equal number of deaths was recorded through the schedules and the rosters, we see that for both samples the highest proportions of women with matched records occurred in rural areas (i.e. 82.4% for sample one and 82.1% for sample two). With regards to the category '*Schedule Recorded More Deaths*' in sample one, a higher percentage of women was noted in the rural area (5.9%), whereas in sample two a higher percentage was recorded for urban women (6.5%).

Where the rosters recorded more deaths than the schedule, for women in sample one a significant difference in proportions is noted between rural and urban women, with the higher

proportion being noted for urban women (25.0%). For sample two there is no noted difference between the proportions of rural (12.8%) and urban women (12.9%).

Figure 4.19 shows that consistent and inconsistent reporting by rural and urban women appears to be balanced. Hence it can be concluded that in this case, geographical location has played no part in determining the propensity to misreport on respondents fertility.

**Figure 4.20:**  
**Comparison of Records for *Girl Children Dead* for Sample One and Sample Two of Women by Living Standard Measure**

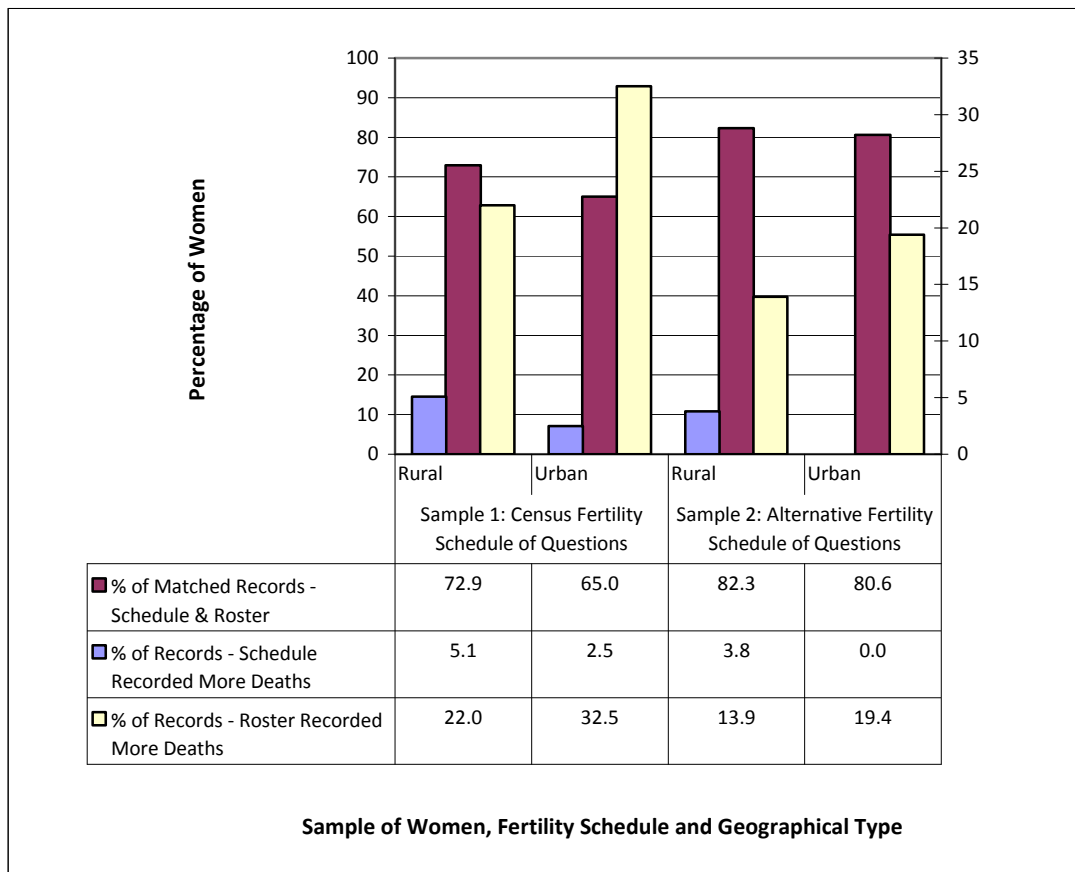


At the outset, percentages of women in certain groups are misrepresented. This should be approached with caution as it is a case of low observed counts rather than high rates of response. It is for this reason that LSM 7-10 will not be considered for analysis.

For sample one, for the category '*% of Matched Records*' the highest proportion of women was noted in the LSM 1-3 group (72.7%). For sample two, the highest percentage of women occurred in the LSM 4-6 group (88.2%). Where the schedule recorded more deaths, the highest percentages of women in both samples of women were observed for the LSM 1-3 group (i.e. 4.5% for sample one and 7.9% for sample two).

As expected, the highest proportions of matched records are from the low to middle income groups. Inconsistent reporting is also observed to have taken place within these groupings, thus solidifying the case that a lower socio-economic status is associated with a tendency to report incorrectly.

**Figure 4.21:**  
**Comparison of Records for *Girl Children Dead* for Sample One and Sample Two of Women by Geographical Type**



For both samples of women, the highest proportion of women measured for the first category of comparison was amongst rural women (i.e. 72.9% for sample one and 82.3% for sample two). In both samples, for cases where the schedule recorded more deaths, higher proportions of women were noted for rural women as opposed to their urban counterparts (i.e. 5.1% for sample one and 3.8% for sample two). For the third category '*Roster more deaths*' a higher percentage of women occurred in the urban group in sample one (32.5%) whereas in sample two, the highest percentage was for urban group of women (19.4%).



It was earlier outlined that urban women would report better than rural women. However, Figure 4.21 illustrates, that rural women for both samples recorded the highest percentage of matched records. This is a welcome finding as the literature often comments on the underreporting of girl children in rural areas.<sup>17</sup> Although this study was able to disprove this notion, Figure 4.21 also shows that rural women also misreported in cases where the schedule recorded more deaths. A surprise finding is that urban women misreported significantly in cases where the roster recorded more deaths than the schedule. These mixed findings point to that fact that the influence of geographical location and reporting remains unclear.

#### ***4.5 The Link between Socio-Economic Status and the Sex-Selectiveness of Reporting***

When comparing Figure 4.12 and Figure 4.14 we observe that there are fairly percentages of matched records between the schedule and the roster of girl children born alive than boy children born alive for both samples across all three LSM groups, with the highest occurring in the LSM 7-10 group. Interestingly, the LSM 7-10 group, in sample one for girl children born alive recorded the highest percentage of matched records. This is an encouraging finding as the deficit of girl children commonly commented on during enumeration in surveys or censuses can be challenged. Furthermore, in sample one as noted from Figure 4.12 and Figure 4.14 are the levels of inconsistent reporting (i.e. higher percentages) of boy children born alive than girl children born alive across the second ('Schedule recorded more births') and third ('Roster recorded more births) categories of comparison.

With regard to reporting of boy and girl children born alive by geographical type (Figure 4.13 and Figure 4.15), for the category of '% of matched records' there was relative consistency in the reporting of these variables by both samples of women. Inconsistencies for sample one, were evenly spread for both urban and rural women, across the remaining categories of comparison. For sample two, there was a lower percentage of matched records for boy children born alive (rural-93.1% and urban-94.6%) than for cases that examined girl children born alive (rural-93.4% and urban 96.6%). In other words, rural women in particular recorded lower percentages of matched records of boy children born alive. It thus appears that the reporting on

girl children born alive by urban was more accurate and consistent and more comprehensive than that of boy children.

For dead boy and girl children (i.e. Figure 4.18 and 4.20), there were higher percentages for sample one (and a significant difference) of matched records for boy children as opposed to dead girl children across all LSM groupings. A notable finding in this instance was for women in sample one for dead boy children, LSM group 1-3 (92.0%). The reporting of dead girl children was disappointing across all three LSM groupings for sample one. Sample two in Figure 4.20 has recorded a higher percentage of matched records for dead girl children than for sample two of Figure 4.18 for dead boy children across all the LSM groups. The highest percentage of matched records occurred in the LSM 4-6 group (88.2%). For both cases, there is more inconsistent reporting across the categories, 'Schedule recorded more births' and 'Roster recorded more births' dead boy children for sample two in Figure 4.18 Conversely, records for dead girl children were inconsistent for sample one in Figure 4.20.

Although Figure 4.21 illustrates, that rural women for both samples recorded the highest percentage of matched records than urban women, but when compared to Figure 4.19 there is a general shortfall of reporting of dead girl children. For the remaining two categories "Schedule recorded more deaths' and 'Roster recorded more deaths', rural women in sample one of Figure 4.20 were more inconsistent in their reporting of dead girl children than dead boy children (sample one, Figure 4.19). Urban women both figures have misreported considerably, with more misreporting occurring for dead girl children.

The findings have shown that rural women reported better than that their urban counterparts on dead girl children. However, when compared to that of dead boy children the situation is no different as stipulated in the literature. From innumerable literature studies, it is gathered that there is a tendency to report differently on the sexes of dead children in rural areas. The affected of the two sexes is usually girl children. This is sadly the case in this study as well. The question to ask is whether this failure to report is intentional or not?

#### **4.6 Summary**

An attempt was made to illicit a model that would predict the likelihood of reporting consistently or inconsistently by the two samples of women- but to no avail. The small sample size and perhaps the low percentage of inconsistent records across both samples prevented a model with statistically significant estimates to be developed. Although the failure of being unable to produce a regression model has narrowed the scope of the study, the tables and figures in the previous section has shown that there is some association between certain demographics of women, their socio-economic status and the likelihood to report correctly and consistently on their reproductive history.

Chapter four presented through a descriptive exploratory analysis, a demographic profile of women associated with inconsistent and inaccurate reporting. The chapter also examined various fertility variables (*'children ever born'*, *'children still alive and living in and not living in the household'* and *'children dead'*) in tandem with the demographic profiles of sample one and sample two. An investigation into establishing a link between the socio-economic status of individuals and retrospective reporting was also conducted and presented. The sex-selectiveness of reporting on birth histories was also examined. Having presented these results, the next chapter discusses these findings in further detail.

**CHAPTER FIVE**  
**Discussion of Results**

\*\*\*\*\*

**5.0 Introduction**

This chapter discusses the results of the study. The discussion will examine the salient findings that emerged in the study. By summarizing these results, the chapter aims to tease out possible reasons for the results and to seek support for the findings from the literature. To re-iterate, the study sought to investigate:

- For women whose birth histories have irregularities, what are their individual characteristics?
- Is there a link between socio-economic status of individuals and retrospective birth reporting? That is, who are the individuals who are likely to omit and misreport births?

Is there a link between socio-economic status of individuals and the aspects of the birth that are misreported? For example, the sex selectiveness of births and deaths that is misreported or omitted? The discussion to follow provides answers to these important questions.

**5.1 Findings of the Study**

**5.1.1 Establishing a Demographic Profile of Women with Inconsistent Reporting**

The exploratory descriptive analysis for both samples of women has established that the most inconsistent reporting was observed for older women (40-49 years old), who are more likely to be married, have a satisfactory level of educational attainment, whose place of residence is in rural areas and are of an average socio-economic status. Conversely, more accurate and consistent reporting is associated with women who are younger (15-24 years old), possess a high level of education, and are of the upper income group. It thus appears from the findings that age, marital status, education, geographical location and LSM status has an influence on the manner in which respondents report on their fertility.

For the characteristic of age, the research study has found that older women, typically 40-49 years of age are associated with less accurate and inconsistent reporting. Comparisons of the schedule and the roster show that women of older age groups tended to omit or under report on their birth histories. This was singled out when the total number of live births and deaths

reported on in either the schedule or roster did not tally (i.e. if a woman reported three live births in the schedule, the roster only recorded 2 live births).

It is postulated in the literature (Marckwardt, 1973; McGranahan, 1976) that older women due to their age, misunderstandings and failure of memory are more likely to omit children that have died in infancy and children who have left the home (Blacker and Brass, 1979). Indeed our research has found a link between the age of respondents and retrospective reporting.

Another reason for this finding and its link to older women is that older women may also include children that are not their own *biological* children (i.e. children of another family member or children born to another husband) (Blacker and Brass, 1979). The latter is known as faculty inclusions. In the context of South Africa, the definition of a 'household' is blurred and complicated. Statistics South Africa (2004:8) defines a household as "a group of persons who live together and provide themselves jointly with food and/or other essentials for living, or a single person who lives alone". This is supported by an investigation by Statistics South Africa (2008) into the suitability of the Census 2001 mortality schedule of questions for the census in 2011. Through focus group discussions, participants noted that a household included family (i.e. parents and children, but is also inclusive of other relatives who reside with the family) (Statistics South Africa, 2008). Other notions expressed was that a household included a group of people who live together in a house or under one roof, where this definition can also extend to members of the family working elsewhere, domestic workers and paying guests as part of the household (Statistics South Africa, 2008).

Due to the very broad definition of the concept of household, it is not difficult to see why faculty inclusions by older women during enumeration, and possibly in this research study, may occur in such a context. Whilst we acknowledge that age does have an influence on reporting, can we simply attribute the fault in reporting to the age of a respondent?

In this study, the following question was asked to sample two of women: "If the person has ever given birth to any live child/ children: How many of these children are still alive *and* living with her in this household/ dwelling/compound? How many of these are boys? How many of these are girls?"

No doubt, the above question is highly loaded and will appear so to most individuals. In Chapter four, the 'double barrelled' nature (i.e. asking if the children were still alive *and* whether or not the children reside in the household) may have had an effect on the manner in which women reported. The complexity of the question may have confused respondents in the manner they reported on their birth histories. That said, error could not only lie with the respondent, but with the research tool and the enumeration process. This is supported by studies conducted by Mathiowetz *et al.* (2002), Blacker and Brass (1979). These findings are also consistent with research on the Malawian Diffusion and Ideational Change project conducted by Bignami-Van Assche *et al.* (2003), where the wording of the questions greatly affected the manner in which respondents report on their birth histories.

Coupled with an ineffective research tool, enumerators themselves can also be a source of measurement error. With regard to the Gambian census of 1973, Blacker and Brass (1979) reported on procedures that were not followed by enumerators during the administration of the questionnaire (i.e. failure to pose [correct] questions). Whilst it cannot be concluded that enumerators are a source of error in this piece of research, we can say that it is possible that enumerators did not explain loaded questions such as the above and concepts such as 'household' to respondents clearly. This may have attributed to why women (older) reported incorrectly.

The effect of marital status and retrospective reporting was not always clear in most cases. However, most often that not, married women in both samples were found to have reported more inconsistently than their counterparts. This finding is supported by Hertrich (1998) where, it was found that married women underreported on the number of foetal deaths as opposed to their husband's. This was attributed to cultural notions surrounding foetal deaths in their society. For our investigation, whilst not conclusive, it is probable that married women did not report correctly on their reproductive history for cultural, social or personal reasons.

Whilst writers such as Beckett *et al.* (2001) have found no significant relationship between education and reporting, studies conducted by Marquis *et al.* (1972) has found an association between the two variables. To add, our study has illustrated that woman with some educational

attainment (i.e. respondents have completed primary school and some secondary school) are associated with inconsistent reporting. Although we could not prove a statistically significant relationship between education and reporting, the descriptive exploratory analysis is convincing enough.

Place of residence is also cited in the literature as a precursor to accurate or inaccurate reporting. This is supported by Marckwardt (1973) whose research in Peru concluded that older women in rural areas reported less accurately on their fertility. Our research has proved this to be true in most cases. A further discussion on the place of residence follows in the next section.

### ***5.1.2 Establishing a Link between the Socio-Economic Status of Individuals and Retrospective Reporting of Birth Histories***

The second part of the analysis explored whether there is an association between the socio-economic status of an individual and aspects of inconsistent reporting. The socio-economic status of an individual was a collective of their LSM status and Geographical Location. The LSM was chosen as a proxy for socio-economic status. By implication, the higher the LSM status, the higher the level of socio-economic status and vice versa. Geographical location was chosen as the second factor that contributed to the socio-economic status of a person as a result of the benefits and drawbacks offered by either geographical location.

The analysis involved the examination of both samples of women, their socio-economic status and the following fertility differentials: *women with live births, women with boy children born alive, women with girl children born alive, women with children dead, women with boy children dead and women with girl children dead.*

The analysis revealed through most cases (with the exception of a few) that there is a link between the socio-economic status of an individual and propensity to report correctly or incorrectly. It was found that women of the higher income groups (LSM 7-10) reported more accurately on the births and deaths of their children. This was the case for women with live births, women with boy children born alive, women with girl children born alive and women with dead girl children. Cases that were not mentioned either had low observed counts or it was unclear to determine where the most inconsistent reporting took place. It is however

encouraging to note that some association (despite low observed counts) was evident. These findings could be attributed to a higher educational status that is accompanied with a higher LSM status. This hence solidifies the link that a higher socio-economic status is associated to more accurate and consistent reporting.

With regard to place of residence, the results for this instance were mixed. In other words, there was no unanimous outcome as to whether urban or rural women reported more consistently. This does not however mean that there is no association. Upon examination, individual fertility differentials show that rural women reported more inconsistently in cases related to boy children born alive. An unexpected finding of rural women reporting more consistently than urban women was noted for records on women with girl children born alive and women with dead girl children. This unexpected and positive finding is discussed further in the next section.

### ***5.1.3 Sex-Selectiveness in the Reporting of Births and Deaths***

The literature often cites a gross underreporting of births and deaths in rural areas (Blacker and Brass, 1979; Coale and Banister, 1994; Arnold, *et al.* 1998; Das Gupta and Bhat, 1996). Consequently, during survey or census enumeration in mainly rural areas, these deaths are not reported, resulting in skewed sex ratios. The failure to report these deaths are also fuelled by social and cultural factors (Coale and Banister, 1994).

Thus the unexpected finding of rural women reporting more consistently for cases on girl children born alive and dead girl children is encouraging and welcomed. Our findings have seemingly disproved some of the findings in the literature.

At the same time, whilst the above finding is positive, when the percentages of dead girl children is compared to that of dead boy children, we still see a deficit of dead girl children being reported by women of both samples. Several writers (Blacker and Brass (1979), Coale and Banister (1994), Arnold, *et al.* (1998) and Das Gupta and Bhat (1996) and Hobcraft and Murphy (1986)), have argued that due to the strong preference for a boy child in rural areas, this desire is often cited as one of the major underlying causes of excess female mortality in infancy. Subsequently, most of these deaths are omitted during censuses and surveys also resulting in distorted sex ratios.



In summation, whilst a higher socio-economic status is associated more accurate and consistent reporting, the link between place of residence and its influence on reporting appears to be a weak relationship. This can be a result of a small sample size and low observed counts. Holistically, the link between socio-economic status and reporting is present in some cases, though more rigorous investigations are needed to fully establish a more credible relationship.

The results of the study with respect to the sex-selectiveness of reporting of births and deaths is bitter-sweet. On the one hand, the study was able to disprove some of the literature in a positive light (i.e. more consistent reporting on girl children born alive and dead girl children by rural women) and on the other hand, the findings are consistent with the literature when it concerned the reporting of dead girl children in comparison with dead boy children.

## **5.2 Other Explanations**

That said, is it possible for other factors to have influenced such reporting? Factors that are not limited to geographical boundaries, levels of socio-economic status and types of women?

The literature identifies the concept of 'saliency' that could explain these inconsistencies across all barriers. Writers ( Mathiowetz *et al.*(2002; Hertrich, 1998; Sudman *et al.*1996; Sudman and Bradburn, 1973; Cannell, Fisher and Bakker, 1965) commented that events judged to be important to an individual are more *completely* and *accurately* reported than other events. It is possible that respondents who reported more consistently than others may have viewed and considered their reproductive histories and its vital events as important and hence reported on it without hesitation.

On the other hand, vital events could have been important to an individual but was not reported on and thus omitted due to cultural, social or personal reasons. Whilst this cannot be proved conclusively in our study, it is a possibility.

Furthermore, Potter's 1977 'Event Placement Model' as discussed in chapter one, has two main propositions to the model. These are as follows (Potter, 1977:341).

The date of an event is recalled less accurately the longer ago the event took place” and;

- “If a birth history is elicited by asking questions about live births in the order in which they occurred, then the date a woman attaches to any event other than the first is influenced by the information she has already given about the previous event”

The first point relates to recall ability and time. There is a possibility especially with that of older women in both samples; some events (births and deaths) were recalled less accurately owing to the time that has lapsed between the event and the date of enumeration. This may have resulted in omissions and misreporting. Whilst it cannot be shown as to which records have been influenced by this proposition, we can safely assume its possibility.

For the second proposition, simply put; if the birth history begins with the first live birth, the details of the subsequent birth are influenced by the details of the previous birth and will thus be noted as such. Thus in the case of inconsistencies, these often occur at the beginning of the birth history and thus irregularities will appear throughout the reporting. This is once again impacted on by time and recall ability. It is quite probable that women in both samples who do not have registration documents for their children or who are unable to remember exact details required during enumeration reported are more influenced by time and recall ability and hence reported inconsistently and less accurately. Perhaps this second proposition can offer some explanation for these irregularities found.

### **5.3 Summary**

This chapter has discussed the findings of the research and has sought to explain them through documented research in the literature. Most of the findings of the research have been consistent with the literature and such justifications were noted. Some findings of the study have also disproved some of the theoretical standpoints. These new findings are to be seen in a positive light as they will contribute to future research in the field of data quality. The chapter also presented other explanations for observed irregularities present in birth histories. Whilst these explanations cannot be proved to have resulted in these misreporting, they certainly could have had an influence. Taking these discussions into account, the next and final chapter presents the implications of these findings and opportunities for future research.

## CHAPTER SIX Conclusion

\*\*\*\*\*

This dissertation set out to establish a demographic profile of women associated with inconsistent reporting. It found that inconsistent and less accurate reporting is found among women who are older (40-49 years old), of the middle income category, most likely to be married, have some educational attainment, whose place of residence is rural.

The research also aimed at establishing an association between the socio-economic statuses of individuals and reporting. The socio-economic status was a collective indicator that comprised of LSM status and place of residence. The findings have shown that there is some association between the socio-economic status of an individual and the manner in which they report. It was found that women who were of a higher socio-economic status reported more consistently and accurately than women who belong to the middle to lower income groups. With regard to place of residence, there is some support for the association in some cases; however more research efforts are needed to fully establish this link.

The study also examined the aspects (i.e. sex-selectiveness of births and deaths) that are reported on and its link to the socio-economic statuses of individuals. In brief, women of LSM 7-10 reported more accurately and consistently on girl children born alive as opposed to boy children born alive. The same trend was observed for urban women. Rural women of LSM 4-6 are noted to have reported more reliably and more correctly on their dead girl children. Thus some associations were made with regard to the sex-selectiveness of reporting and the socio-economic status of respondents.

Waksberg and Pritzker (1969:1143) note that a major impact on the accuracy of collected data lay in three scenarios: "enumerators' systematic errors or the respondents' lack of knowledge of the answers or unwillingness to report the correct information". These are evident in all types of research and may have influenced the results of my research.

Taking these findings into account, what are its implications and the way forward? As shown in the study: given that a certain respondent profile is associated with inconsistent reporting, the

question is: how do we still strive to achieve fair levels of data quality amidst this profile of respondents? Furthermore, how can we combat the effect of enumerators as a source of measurement error? I believe that this impact on the measurement of fertility and other core topics can be addressed and minimized in the following ways.

The issue of training of census staff, enumerators and data capturers is a not a new one. Despite this, problems cited during enumeration and data capturing often goes back to the same point. Whilst it is acknowledged that such training does occur during surveys and censuses (a case in point is the training provided by Statistics South Africa for their ongoing surveys and censuses) more vigorous efforts need to be implemented.

It is recommended that specified training be organized for the census team according to their function and duties. Overlapping census functions should be incorporated so that the link and effect on individual work streams can be seen by those involved. Consistency and uniformity should be maintained for the training. This can be achieved through establishing a core training group that can oversee and maintain the consistency of census procedures (i.e. terms and concepts and the standardized understanding of these concepts).

In training fieldworkers and other census staff, it is necessary to make them aware of their subject matter (in cases of specific surveys), roles and responsibilities. Enumerators need to be informed of their critical part in the chain in ensuring the success of a census or survey. This can done through using past surveys as examples of how unsatisfactory enumeration compromised the results of the survey or conversely how good data collected in the field contributed to the success of the survey.

Another strategy would be to include a select few into the entire process from data collection to the dissemination of the results. It will aid fieldworkers in understanding the census or survey process as a chain of many links rather than separate procedures. This will also help reinforce their roles as enumerators and see how their contribution fits in the bigger picture.

During training sessions, incentive based training should be implemented. Incentives need not necessarily be cash based, but rather geared towards fostering a motivating and persevering

attitude amongst trainees. An example of this could be the development of a 'points based performance scorecard' that will reward individuals or teams, based on their performance at the training. These points can be gathered through routine and spot tests that record the progress of trainees and help monitor team performance. A reward scheme for top performers can be devised to further motivation.

Several errors occur prior to the data being captured. Continuous and meticulous quality control should be implemented at all stages to minimize these errors. Ideally the enumerator should place the first quality checks, followed by peer enumerators and lastly the field supervisor. Whilst this is a strategy that is currently employed, more efforts need to be done to ensure strict compliance by fieldworks and field supervisors.

The alternative schedule of fertility questions used in this study, developed through the recommendations by the United Nations 2006 document titled '*Principles and Recommendations for Population and Housing Censuses*' served as a test for the measurement of fertility. By implication, its use has shown satisfactory levels of data. Such useful exercises help to plan and assess the suitability of intended census questions. More efforts of this nature need to be employed to ensuring good, quality data as well as minimizing margins of error.

The first step would be to acknowledge, accept and prepare for these barriers. As survey methodologists, we need to understand and not underestimate its influence on survey responses. Perhaps, further research into sensitive topics needs to be carried out in order to understand the effect of culture and society.

For example, Statistics South Africa (2005) has followed this method and has subsequently identified an array of topics that were defined as 'sensitive' to respondents. Fertility was ranked as the fourth most sensitive topic out of twenty core census topics. This valuable information can aid the research team on how to approach such subjects without being insensitive to respondents.

Prior to enumeration, census or survey fieldworkers need to be educated on culturally appropriate behaviour and etiquette. This can be achieved by interacting with local leaders of the

research site to determine what is termed as 'acceptable behaviour' within the socio-cultural climate. Such efforts will assist in achieving a cooperative and productive interview. This can also hopefully create an environment of comfort and reduce respondents' unwillingness to report and misreport on their data.

Language barriers can be addressed by working with language experts and interpreters during training sessions. These individuals can assist in providing alternative and simply understood words and phrases in cases where respondents find difficulty in understanding a particular question/s.

Although publicity campaigns are a common feature prior to surveys and censuses, I feel it necessary to mention these key points, in light of the results of my study and with Census 2011 on the horizon. Publicity campaigns should emphasize the importance of respondent participation. The assurance of confidentiality between the enumerator and the respondent should receive vast amounts of attention in these campaigns. After all, it is these issues of confidentiality, privacy and sensitivity that affect respondent reporting. It is recommended that these campaigns interact with individuals at the grass root level. Specialised publicity campaigns can assist with these interactions and hence reach out to individuals with false perceptions and general lack of knowledge and understanding of surveys and censuses.

Several aspects of this research study could not be investigated as it fell out of the scope of the study. These gaps could provide impetus to new research that could contribute to investigations into data quality.

Due to limitations such as restrictions placed on the sample size, low observed counts of data irregularities the wording of the questionnaires and limited statistical applications of the issues such as heaping, bunching and telescoping of data could not be explored. Perhaps if another study on fertility were to be conducted, these and other matters could be investigated and discussed if the following suggestions are put into practice:

- Fertility data should be collected from all women aged 15 and above. It is suggested that the upper limit of 49 years of age be increased to include older women up to 65 years

old. It is assumed, that the decision to include more women will result in more data on reproductive histories. The inclusion of older women, hence an increased sample size, is hoped to bring out more inconsistent reporting and perhaps prove a statistically significant relationship between age and reporting. An increased sample size will also perhaps aid in establishing a more convincing case for the relationship between the socio-economic statuses of individuals and reporting.

- Perhaps for the execution of the above point, a survey designed for the sole purpose of eliciting data on full reproductive histories needs to be taken into consideration. Given the importance of fertility, the lack of reliable vital registration data in South Africa and poor data quality in the past, this is indeed a plausible option. If this option is implemented, this could open up a plethora of analyses thus making significant contributions to field of fertility and data quality.

This chapter has shown that the research objectives of the dissertation have been met. Given these findings, the chapter has outlined several general recommendations to deal with data quality issues relating to fertility and other core topics. These include the proper training of enumerators and data capturers, quality control during data collection, testing of questionnaires dealing with social, cultural and language barriers and the reinforcement of publicity campaigns. Gaps in the research were identified and also suggested as opportunities for future research.

Given the importance of good, quality data especially on the core demographic process of fertility and how data of this nature assists in ensuring the development of policies and plans for a country, I would like to end my report with a quote by the prolific Nobel Prize winning writer and economist, Professor Amartya Sen: “even if data limitations may quite often force us to make practical compromises, conceptual clarity requires that we do not smugly elevate such a compromise to a position of unquestioned significance” (Sen, 1984: 88).

## Appendix A: Questionnaires Administered in the Survey

<p><b>Fertility Schedule Questionnaire (Version 1 and Version 2)</b></p>	<ul style="list-style-type: none"> <li>▪ The questionnaire was administered to a responsible adult of the household, preferably the head of the household.</li> <li>▪ Section A of the questionnaire required general information for persons who usually live in the household as well as visitors who stayed the previous night.</li> <li>▪ Section B of Version 1 of the questionnaire (the fertility schedule used in Census 2001) was completed for females aged 15 to 49 years old.</li> <li>▪ Version 2 of the questionnaire was completed for an alternative fertility schedule that consisted of four questions with tally checks.</li> <li>▪ These sections were completed as far as possible in consultation with the relevant women.</li> </ul>
<p><b>Female Questionnaire (Version 1 and version 2) – Also referred to as the roster of live births and deaths</b></p>	<ul style="list-style-type: none"> <li>▪ Version 1 of this questionnaire had to be used with Version 1 of the Fertility Schedule Questionnaire,</li> <li>▪ Version 2 had to be used with Version 2 of the Fertility Schedule Questionnaire.</li> <li>▪ The questionnaire had to be administered to a female (15-49 years old) for whose fertility information was recorded in the relevant Fertility Schedule Questionnaire. The questionnaire was completed through an interview.</li> <li>▪ For Section A of the questionnaire a full roster of all live births for the female in question was completed. Responses on the Fertility Schedule Questionnaire were not changed if any discrepancies are noted.</li> <li>▪ For Section B the fieldworker compared the responses recorded in the Fertility Schedule Questionnaire with the birth history recorded in Section A.</li> <li>▪ Discrepancies were discussed with the respondent and based on this discussion the fieldworker completed the following information in Section C: <ul style="list-style-type: none"> <li>- If any, discrepancies were noted</li> <li>- All the reasons for the discrepancies noted</li> </ul> </li> </ul>
<p><b>Living Standard Measure Questionnaire (LSM)</b></p>	<ul style="list-style-type: none"> <li>▪ For this questionnaire, the respondent of the Fertility Schedule Questionnaire (responsible adult, preferably the head of the household) provided responses to this questionnaire.</li> <li>▪ The questionnaire collected data required for the calculation of a LSM according to the methodology of the All Media and Products Survey (AMPS)</li> </ul>



**Appendix B: Fertility Questionnaire Version One Administered to Sample One of Women (Census 2001 Fertility Schedule of Questions)**

**Record Information for all females 15-49 years old. Must be completed in consultation with the relevant woman.**

- *How many children, if any, has the person ever had, that were born alive?:*  
How many of these were boys?  
How many of these were girls?  
Include ALL her children, i.e. those who are still living, whether or not they live in this household, and those who are dead. DO NOT COUNT STILLBIRTHS (children born dead).
- *STILL LIVING: If the person has ever given live birth:*  
If boys: How many boys are still alive?  
If girls: How girls are still alive?
- *LAST CHILD BORN. If (the person) has ever given live birth:*  
When was (the person's) last born child?  
What was the sex of that child?  
Is that child alive or dead?  
Write the day, month and year of the last live birth and dot the appropriate box of the sex. If multiple birth, indicate only the last child. Dot the appropriate box whether the child is still alive on Census night 9-10 October. DO NOT COUNT STILLBIRTHS (children born dead)

**Appendix C: Fertility Questionnaire Version Two Administered to Sample Two of Women (Alternative Schedule Fertility Questions)**

**Record Information for all females 15-49 years old. Must be completed in consultation with the relevant woman.**

- *How many children, if any, has the person ever given birth to that were born alive (i.e. excluding stillbirths) even if they died soon after birth?*

How many of these were boys?

How many of these were girls?

Include ALL her children, i.e. those who are still living, whether or not they live in this household, and those who are dead. DO NOT COUNT STILLBIRTHS (children born dead).

- If the person has ever given birth to any live child/ children:

How many of these children are still alive and living with her in this household/ dwelling/compound?

How many of these are boys?

How many of these are girls?

- If the person has ever given birth to any live child/ children:

How many of these children are still alive and alive and now living elsewhere?

How many of these were boys?

How many of these were girls?

- If the person has ever given birth to any live child/ children:

How many of these children are now dead?

What was the sex of that child?

How many of these were boys?

How many of these were girls?

**Appendix D: Female Questionnaire Version One and Version Two – Also referred to as the Roster of Live Births and Deaths**

**NOTE: The respondent for this questionnaire should be a female for whom information in a Fertility Schedule Questionnaire was recorded and who is 15 years or older. This questionnaire MUST BE COMPLETED THROUGH AN INTERVIEW.**

**SECTION A: BIRTH HISTORY - RECORD ALL LIVE BIRTHS**

Complete the birth history in the following table. Do not alter the responses on the Fertility Schedule Questionnaire if any discrepancies are noted.

Birth Order	What name was given to your ...	Record single or multiple birth status.  1 = Single 2 = Multiple	Sex of the Child  1 = Male 2 = Female	On what day, month and year was (Name) born?		Is (Name) still alive?  1 = Yes 2 = No	If dead: In what day, month and year did (Name) die?		If alive: Is (Name) living in this household/dwelling/compound?  1 = Yes, 2 = No		
				DD	MM		YY	DD		MM	YY
Last (Most recent born) <input type="text"/>		<input type="checkbox"/> 1 <input type="checkbox"/> 2	<input type="checkbox"/> 1 <input type="checkbox"/> 2	DD	MM	YY	<input type="checkbox"/> 1 <input type="checkbox"/> 2	DD	MM	YY	<input type="checkbox"/> 1 <input type="checkbox"/> 2
Last but 1 <input type="text"/>		<input type="checkbox"/> 1 <input type="checkbox"/> 2	<input type="checkbox"/> 1 <input type="checkbox"/> 2	DD	MM	YY	<input type="checkbox"/> 1 <input type="checkbox"/> 2	DD	MM	YY	<input type="checkbox"/> 1 <input type="checkbox"/> 2
Last but 2 <input type="text"/>		<input type="checkbox"/> 1 <input type="checkbox"/> 2	<input type="checkbox"/> 1 <input type="checkbox"/> 2	DD	MM	YY	<input type="checkbox"/> 1 <input type="checkbox"/> 2	DD	MM	YY	<input type="checkbox"/> 1 <input type="checkbox"/> 2
Last but 3... (up to 10th birth) <input type="text"/>		<input type="checkbox"/> 1 <input type="checkbox"/> 2	<input type="checkbox"/> 1 <input type="checkbox"/> 2	DD	MM	YYYY	<input type="checkbox"/> 1 <input type="checkbox"/> 2	DD	MM	YYYY	<input type="checkbox"/> 1 <input type="checkbox"/> 2

**SECTION B: BIRTH HISTORY CONTROL SHEET**

Complete the following table through reference to the relevant entries in the Fertility Schedule Questionnaire and the roster in Section A. DO NOT CHANGE ANY OF THE ENTRIES if the totals do not tally.

Fertility Schedule Questionnaire (Version 1)			Roster of all alive births (Section B)		
Number of children ever born alive			Number of children ever born alive		
Boys			Boys		
Girls			Girls		
Number of children ever born alive who are still living			Number of children ever born alive who are still living		
Boys			Boys		
Girls			Girls		

**SECTION C: NATURE AND REASONS OF DISCREPANCIES BETWEEN RESPONSES TO THE FERTILITY SCHEDULE AND THE BIRTH HISTORY**

The interviewer must compare the responses recorded in the Fertility Schedule Questionnaire with the birth history recorded in Section A. Any discrepancies must be discussed with the respondent and based on this discussion the interviewer must complete the following questions.

<p><b>1</b> Look at the Birth History Control Sheet (Section B). Are any discrepancies noted between the responses recorded in the Fertility Schedule Questionnaire and the birth history recorded in Section A?</p> <p>1 = Yes 2 = No</p>	<p><input type="checkbox"/> 1 <input type="checkbox"/> 2</p>				
<p><b>2</b> If any discrepancies were noted: Please indicate ALL the reasons for the discrepancies noted.</p> <p>1 = Woman was not available for consultation 2 = Child/children to whom the woman has not given birth, but living with her, was listed as her own in the fertility schedule 3 = Child/children to whom the woman has not given birth and not living with her, was listed as her own in the fertility schedule 4 = Child/children forgotten when the fertility schedule was completed as he/she/they do(es) not live with mother 5 = Dead child/children forgotten when the fertility schedule was completed 6 = Gender of a child/children was incorrectly recorded in the fertility schedule 7 = Other (Specify)</p>	<p><input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7</p>				
<table border="1" style="width: 100%; height: 100%;"> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> </table>					

## Appendix E: Motivation for Usage of the Living Standard Measure in the Research Study

According to Statistics South Africa (2005) social class (due to class-specific perceptions and attitudes) should have a significant relationship with measurement phenomena. The Living Standard Measure (LSM), developed by the South African Advertising Research Foundation (SAARF), which is instrumental in its usage in market research, can serve as a proxy for material welfare.

The LSM measure which is based on a regression model is calculated by “finding the sum of the following for which the corresponding criterion is satisfied” (Statistics South Africa, 2005:6). In other words, respondents will respond either yes or no to each of the following commodities. Once having gone the list, the score is tallied.

Criterion	Constant
Hot running water	0,158200
Fridge/freezer	0,152515
Microwave oven	0,126829
Flush toilet in house or on plot	0,142228
VCR in household	0,134488
Vacuum cleaner/floor polisher	0,135318
Washing machine	0,138930
Computer at home	0,132148
Electric stove	0,163219
TV set(s)	0,133830
Tumble dryer	0,117338
Telkom telephone	0,097140
Hi-fi or music centre	0,105378
Built-in kitchen sink	0,165505
Home security service	0,091632
Deep freezer	0,093849
Water in home or on stand	0,127671
M-Net and/or DSTV	0,126068
Dishwasher	0,119925
Electricity	0,128613
Sewing machine	0,090320
Live in Gauteng	0,056788
Live in Western Cape	0,079999
1 or more motor vehicles	0,155217
No domestic worker	-0,222360
No cellphone in household	-0,175180

Home is a traditional hut	-0,201080
None or only one radio	-0,158250
Living in a non-urban area outside of Gauteng or Western Cape	-0,093220

To the tallied score a constant of 1,340410 is added to this sum (Statistics South Africa, 2005). The LSM is obtained by referring to the following table. The range of the LSM is identified by the total obtained after the constant has been added (Statistics South Africa, 2005).

Calculated total	Living Standard Measure
0,00000 – 0,72100	LSM 1
0,72101 – 1,05300	LSM 2
1,05301 – 1,35600	LSM 3
1,35601 – 1,72600	LSM 4
1,72601 – 2,12700	LSM 5
2,12701 – 2,68500	LSM 6
2,68501 – 3,01000	LSM 7
3,01001 – 3,32400	LSM 8
3,32401 – 3,65000	LSM 9
3,65001+	LSM 10

The LSM groupings of 1-10 into have been limited to three main groups (i.e. LSM 1-3, LSM 4-6 and LSM 7-10). LSM 1 to LSM 3 is considered a low LSM group (i.e. low material welfare). LSM 4 to LSM 6 is considered to be middle LSM group (i.e. average material welfare). LSM 7 to LSM 10 is considered to be a higher LSM group (i.e. higher material welfare).

## References

- Arnold, F., Bulatao, R.A., Buripakdi, C., Chung, B.J., Fawcett, J.T., Iritani, T., Lee, S.J and Wu, T., (1975). "The Value of Children: A Cross-National Study, Introduction and Comparative Analysis" Vol 1(1). East-West Centre: Honolulu.
- Arnold, F., Choe, M, K., and Roy, T. K., (1998). "Son Preference, the Family-Building Process and Child Mortality in India" *Population Studies*, Vol 53(2), pp. 301-315.
- Arokiasamy, P., (2004). "Regional Patterns of Sex Bias and Excess Female Child Mortality in India" *Population*, Vol 59 (6), pp.833-863.
- Auriat, N., (1993). "My Wife Knows Best: A Comparison of Event Dating Accuracy Between the Wife, the Husband, the Couple, and the Belgium Population Register", *The Public Opinion Quarterly*, Vol 57(2), pp.165-190.
- Bang, A., Reddy, M, H., and Deshmukh, M, D., (2002). "Child Mortality in Maharashtra" *Economic and Political Weekly*, Vol 37 (49), pp. 4947-4965.
- Beckett, M., Da Vanzo, J., Sastry, N., Panis, C., and Peterson, C., (2001). "The Quality of Retrospective Data: An Examination of Long Term Recall in a Developing Country" *The Journal of Human Resources*, Vol 36 (3), pp.593-625.
- Bignami-Van Assche, S., Reniers, G., and Weinreb, A, A., (2003). "An Assessment of the KDICP and MDICP Data Quality: Interviewer Effects, Question Reliability and Sample Attrition Consistency in Survey Response in Rural Malawi" *Demographic Research*, Special Collection 1, Research Article 2.
- Bignami-Van Assche, S., (2003). "Are We Measuring What We Need to Measure? Individual Consistency in Survey Response in Rural Malawi" *Demographic Research*, Special Collection 1, Research Article 3.
- Blacker, J., and Brass, W., (1979). "Experience of Retrospective Demographic Enquiries to Determine Vital Rates" *The Recall Method in Social Surveys*. Studies in Education (New Series) 9. Moss L., and Goldstein, H., (eds) University of London Institute of Education, London, pp. 49-61.
- Björkenstam, C., and Brundell., (2000). "Recall Effects in Retrospective Statistical Individual Swedish Surveys" Stockholm University. Downloaded from <http://www.scb.se/Grupp/Omscb/Dokument/Bjorkenstam.pdf>. Stockholm University.
- Bulatao, R.A., (1981). "Values and Disvalues of Children in Successive Childbearing Decisions" *Demography*, Vol 18(1), pp. 1-25.
- Cannell, C., Fisher, G., and Bakker, T., (1965). "Reporting of Hospitalization in the Health Interview Survey" *Vital and Health Statistics*, Series 2 No.6. U.S. Public Health Service: Washington D.C.



- Chidambaram, V. C., and Pullum, T. W., (1981). "Estimating fertility Trends from Retrospective Birth Histories: Sensitivity to Imputation of Missing Dates" *Population Studies*, Vol 35 (2), pp.307-320.
- Coale, A. J., and Li, S., (1991). "The Effect of Age Misreporting in China on the Calculation of Mortality Rates at Very High Ages" *Demography*, Vol 28(2), pp. 293-301.
- Coale, A.J., and Banister, J., (1994). "Five Decades of Missing Females in China" *Demography*, Vol 31 (3), pp. 459-479.
- Dakin, R. D., and Tennant, D., (1968). "Consistency of Response by Event-Recall Intervals and Characteristics of Respondents" *The Sociological Quarterly*, Vol 9(1) pp. 73-84.
- Das Gupta, M., and Bhat, P.N.M., (1996). "Intensified Gender Bias in India: A Consequence of Fertility Decline". Paper presented at the Annual Meeting of the Population Association of America, New Orleans, 9-11 May 1996.
- Department of Health., (2002). "South Africa Demographic and Health Survey 1998, Full Report" Department of Health, Medical Research Council and Measure DHS: Pretoria.
- Dillman, D.A., and Smyth, J. D., (2007). "Visual Design, Order Effects, and Respondent Characteristics in a Self Administered Survey" *Survey Research Methods*, Vol 1 (3), pp. 121-138.
- Ebbinghaus, H., (1885-1913). "Forgetting Curve" *Memory A Contribution to Experimental Psychology*. Downloaded from [http://encarta.msn.com/media\\_461547609\\_76157803\\_-1\\_1/Forgetting\\_Curve.html](http://encarta.msn.com/media_461547609_76157803_-1_1/Forgetting_Curve.html) on 22 January 2008.
- Gaskell, G.D., Wright, D. B., and O'Muircheartaigh, C.A., (2000). "Telescoping of Landmark Events: Implications for Survey Research" *The Public Opinion Quarterly*, Vol 64(1), pp.77-89.
- Gibson, J., and Kim, B., (2007). "Measurement Error in Long Term Retrospective Recall Surveys of Earnings" Working Paper in Economics-3/07, Department of Economics, University of Waikato: New Zealand.
- Haaga, J. C., (1986). "The Accuracy of Retrospective Data from the Malaysian Family Life Survey", Report N-2157AID. Santa Monica: RAND.
- Hertrich, V., (1998). "Are Men's and Women's Answers to be Equally Trusted? A Dual Collection of Birth and Marriage Histories in a Population in Mali" *Population: An English Selection*, Vol 10 (2), pp. 303-318.
- Herzog, A. R., and Rodgers, W.L., (1989). "Age Differences in Memory Performance and Memory Ratings in a Sample Survey" *Psychology of Aging*, Vol 4 (1), pp. 173-182.
- Hill, K., Choi, Y., and Timæus, I.M., (2005). "Unconventional Approaches to Mortality Estimation" *Demographic Research*, Vol 13(12), pp 281-300.

- Hill, K., and Choi, Y., (2006). "Neonatal mortality in the Developing World" *Demographic Research*, Vol14 (18), pp. 429-452.
- Hirschowitz, R., (n.d). "Methodology for Census 2001 South Africa: Applying the Lessons Learnt from Census '96" Statistics South Africa: Pretoria – Downloaded from [http://www.statssa.gov.za/assd\\_dec2007/WebsiteReports/SouthAfrica2.pdf](http://www.statssa.gov.za/assd_dec2007/WebsiteReports/SouthAfrica2.pdf) on 18 December 2007.
- Hobcraft, J, and Murphy, M., (1986). "Demographic Event History Analysis: A Selective Review" *Population Index*, Vol 52 (1), pp.3-27.
- Huttenlocher, J. L., Hedges, V., and Bradburn, N.M., (1990). "Reports of Elapsed Time: Bounding and Rounding Processes in Estimation" *Journal of Experimental Psychology, Learning, Memory and Cognition*, Vol 16(1), pp.196-213.
- International Statistical Institute., (1973). "The World Fertility Survey: An International Programme of Fertility Research" *International Statistical Review*, Vol 41(3), pp. 303-314.
- Ito, R., (1963) "An Analysis of Response Errors: A Case Study" *The Journal of Business*, Vol 36 (4), pp. 440-447.
- Jahn, A., Crampin, A.C., Glynn, J.R., Mwinuka, V., Mwaiyeghele, E., Mwafilaso, J., Branson, K., McGrath, N., Fine, P.E.M., and Zaba, B., (2007). "Evaluation of a Village-Informant Driven Demographic Surveillance System" *Demographic Research*, Vol 16(8), pp 219-248.
- Khan, N., (2007). "Research to Inform Census Methodologies: Testing a Disability Schedule for Census 2011". Paper presented at the 2<sup>nd</sup> Annual Population Association of Southern Africa conference held at the University of the North West, Mafikeng, North-West Province, 26-28 September 2007.
- Lawrence, L., Wu, S.P.M., and Long, D.A., (2001). "Comparing Data Quality of Fertility and First Sexual Intercourse Histories" the *Journal of Human Resources*, Vol 36(3), pp. 520-555.
- Marcwardt, A.M., (1973). "Evaluation of an Experimental Short Interview Form Designed to Collect Fertility Data: The Case of Peru" *Demography*, Vol 10 (4), pp. 639-657.
- Marquis, K.H., Cannell, C.F., and Laurent, A., (1972). "Reporting Health in Household Interviews: Effects of Reinforcement, Question Length and Reinterviews" Rockville Md: National Center for Health Statistics. (Vital and Health Statistics, Series 2: Data Evaluation and Methods Research, No 45) (DHEW Publication No. (HSM) 72-1028.
- Mason, K.O., and Cope, L.G., (1987). "Sources of Age and Date-of-Birth Misreporting in the 1900 U.S. Census" *Demography*, Vol 24 (4), pp. 563-573.
- Mathiowetz, N.A., Brown, C., and Bound., (2001). "Measurement Error in Surveys of the Low-

- Income Population” National Academy Press: Washington DC. Downloaded from <http://aspe.hhs.gov/hsp/welf-res-data-issues02/pdf/06.pdf> on 22 January 2008
- McFalls, Jnr, J.A., (2007). “Population: A Lively Introduction, 5<sup>th</sup> Edition” *Population Bulletin*, Vol 62 (1), pp 1-34. Population Reference Bureau: Washington DC.
- McGranahan, D.A., (1976). “Correcting for Informant Bias” *American Sociological Review*, Vol 41 (1), pp. 176-179.
- Merli, M.G., (1998). “Underreporting of Births and Infant Deaths in rural China: Evidence from Field Research in One County of Northern China” *The China Quarterly*, No 155, pp. 637-655.
- Merli, M.G., and Raftery, A.E., (2000). “Are Births Underreported in Rural China? Manipulation of Statistical Records in Response to China’s Population Policies” *Demography*, Vol 37 (1), pp. 109-126.
- Ministry of Health, Ministry of Planning and Development Cooperation and World Health Organization- Iraq. (2005). “Report of the In Depth Analysis-Iraq Child and Maternal Mortality, 1999” Downloaded from [http://www.emro.who.int/iraq/pdf/ICMMS Analysis.pdf](http://www.emro.who.int/iraq/pdf/ICMMS_Analysis.pdf) on 3 December 2007. Ministry of Health, Ministry of Planning and Development Cooperation and World Health Organization: Iraq.
- Moultrie, T.A., and Timæus, I.M., (2002). “Trends in South African Fertility between 1970 and 1998-An Analysis of the 1996 Census and the 1998 Demographic and Health Survey”. Burden of Disease Research Unit, Medical Research Council.
- Moultrie, T.A., and Timæus, I.M., (2003). “The South African Fertility Decline: Evidence from Two Censuses and a Demographic and Health Survey” *Population Studies*, Vol 57 (3), pp. 265-283.
- Moultrie, T. A., and Dorrington, R., (2004). “Estimation of Fertility from the 2001 South African Census Data” CARE Monograph No 9, Centre for Actuarial Research: University of Cape Town.
- Oppenheim-Mason, K., and Cope, L.G., (1987). “Sources of Age and Date-of-Birth Misreporting in the 1900 U.S Census” *Demography*, Vol 24(4), pp.563-573.
- Pal, P., and Ghosh, J., (2007) “Inequality in India: A Survey of Recent Trends” Working Paper No.45. United Nations Department of Economic and Social Affairs. Downloaded from [http://www.un.org/esa/desa/papers/2007/wp45\\_2007.pdf](http://www.un.org/esa/desa/papers/2007/wp45_2007.pdf) 11 Nov 09 on 11 November 2009.
- Pollak, R.A., and Watkins, S.C., (1993). “Cultural and Economic Approaches to Fertility: Proper Marriage or Mesalliance”, *Population and Development Review*, Vol 19(1), pp. 467-496.
- Potter, J.E., (1977). “Problems in Using Birth History-Analysis to Estimate Trends in

- Fertility" *Population Studies*, Vol 31 (2), pp. 335-364.
- Poulain, M., Riandey, B., and Firdion, J., (1992). "Data from a Life History Survey and from the Belgian Population Register: A Comparison" *Population*, Vol 4(1), pp.77-96.
- Ravallion, M., and Chen, S., (2004). "China's (Uneven) Progress Against Poverty". World Bank Policy Research Working Paper 3408. World Bank: Washington
- Reddy, V, and Khan, N., (2006). "Required Physical Infrastructure to Attain the Vision of The National System Innovation, Secondary School Education Component: 'Hands on Minds On'." Study commissioned by the National Advisory Council for Innovation. Human Sciences Research Council: Pretoria
- Rubin, D.C., and Baddeley, A.D., (1989). "Telescoping is Not Time Compression: A Model of Dating of Autobiographical Events" *Memory and Cognition*, Vol 17(1), pp.653-661.
- Sabagh, G., and Scott, C., (1967). "A Comparison of Different Survey Techniques for Obtaining Vital Data in a Developing Country" *Demography*, Vol 4 (2), pp. 759-772.
- Schoumaker, B., (2004). "A Person-Period Approach to Analyzing Birth Histories" – Downloaded from <http://www.cairn.info/revue-population-english-2004-5-page-689.htm> on 3 December 2007. Institute of Demography: University Catholique de Louvain.
- Scott, C., (1975). Reviewed Work: "Recall Relapse in Demographic Enquiries" by Ranjan Kumar Som *Population Studies*, Vol 29 (2), pp 328-332.
- Sen, A., (1984). "The Living Standard" *Oxford Economics Papers, New Series*, Vol 36(1), pp.74-90.
- Shyrock, S.H., and Siegal J.S., (1976). "Studies in Population: The Methods and Materials of Demography". Condensed Edition by Stockwell, E.G. Academic Press: Florida.
- Sine, J and Peterson, C, E., (1993). "The Second Malaysian Family Life Survey: Quality of Retrospective Data for the MFLS-2 New Sample", Report MR-110-NICHD. Santa Monica: RAND.
- Singh, S., (1987). "Evaluation of Data Quality" *The World Fertility Survey: An Assessment* Cleland J., and Scott, C., (eds) The International Statistical Institute. Oxford University Press, New York, pp. 618-643.
- Sirken, M.G., and Sabagh, G., (1968). "Evaluation of Birth Statistics Derived Retrospectively from Fertility Histories Reported in a National Population Survey: United States, 1945-64" *Demography*, Vol 5 (1), pp. 485-503.
- Smith, H. L., (1994). "Nonreporting of Births or Nonreporting of Pregnancies? Some Evidence From Four Rural Counties in North China" *Demography*, Vol 31 (3), pp. 481-486.
- Smith, J.P., and Thomas, D., (1997). "Migration in Retrospect: Remembrances of Things Past",

- Report DRU-1628-NICHD. Santa Monica: RAND.
- Smith, J.P., and Thomas, D., (2002). "Remembrances of Things Past: Test Re-Test Reliability of Retrospective Migration Histories" Paper CCPR-002-02- Online Working Paper Series. California Centre for Population Research, University of California: Los Angeles: RAND
- Srinivasan, K, and Mutiah, A.C., (1987). "Fertility Estimation from Retrospective Surveys: Biases Attributable to Pregnancy-Related Movement of Mothers" *Demography*, Vol 24 (2), pp. 271-278.
- Statistics South Africa., (1996). "1996 Census Questionnaire" Statistics South Africa: Pretoria.
- Statistics South Africa., (2001). "2001 Census Questionnaire" Statistics South Africa: Pretoria.
- Statistics South Africa., (2003). "Project: Census 2001 Data Processing-Final Data Handover Report" Version 1. Statistics South Africa: Pretoria.
- Statistics South Africa., (2004a). "Census 2001: Primary Tables South Africa Census '96 and 2001 Compared", Report No 03-02-24. Statistics South Africa: Pretoria.
- Statistics South Africa., (2004b). "Census 2001: Concepts and Definitions", Report No 03-02-26 (2001) Version 2. Statistics South Africa: Pretoria.
- Statistics South Africa., (2005). " Report on Census Publicity Research Study – February/ March 2005". Statistics South Africa: Pretoria.
- Statistics South Africa., (2006). "Census 2011 - Strategic Plan – Research and Methodology" Statistics South Africa: Pretoria.
- Statistics South Africa., (2007). "Census 2011 Strategic Plan – Final Draft" Statistics South Africa: Pretoria.
- Statistics South Africa., (2007). "Community Survey, 2007-Revised Version" Statistical Release P0301. Downloaded from <http://www.statssa.gov.za/publications/P0301/P0301.pdf> on 21 November 2009.
- Statistics South Africa., (2008). "Testing the Mortality Schedule for Census 2011: Report on Focus Groups – November 2007", Statistics South Africa: Pretoria
- Statistics South Africa., (2009). "2009 Mid-Year Population Estimates". Statistical Release P0302 Statistics South Africa: Pretoria.
- Strack, F., and Martin, L.L., (1987). "Thinking, Judging and Communicating: A Process Account of Context Effects in Attitude Surveys" *Social Information Processing and Survey Methodology*. Hippler, H.J., Schwarz, N., and Sudman, S., (eds) pp. 123-148. Springer -Verlag: New York.

- Sudman, S., and Bradburn, N., (1973). "Effects of Time and Memory Factors on Response in Surveys" *Journal of the American Statistical Association*, Vol 68(1), pp. 805-815.
- Sudman, S., Bradburn, N.M., and Schwarz, N., (1996). "Thinking About Answers: The Application Of Cognitive Processes to Survey Methodology". CA: Jossey-Bass Publishers: San Francisco
- Summers, G.F., (1969). "Towards a Paradigm for Respondent Bias in Survey Research" *The Sociological Quarterly*, Vol 10(1), pp.113-121.
- Thompson, C.P., Skowronski, J.J., and Lee, D. J., (1988). "Telescoping in Dating Naturally Occurring Events" *Memory and Cognition*, Vol 16(1), pp. 461-468.
- Tourangeau, R., (1984). "Cognitive Science and Survey Methods: A Cognitive perspective" *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*. Jabine, T., Straf, M., Tanur, J., and Tourangeau, R., (eds). National Academy Press: Washington, D.C.
- United Nations., (2001) "Handbook on Population and Housing Census Editing" Studies in Methods, Series F No 82. Department of Economic and Social Affairs, Statistics Division. New York
- United Nations., (2006). "Principles and Recommendations for Population and Housing Censuses: Revision Two" United Nations: New York.
- United Nations., (2007). "Map of South Africa". Map No 3768, Revision Six. Department of Peacekeeping Operations, Cartographic Section, United Nations: New York. Downloaded from <http://www.un.org/Depts/Cartographic/map/profile/southafr.pdf> on 24 November 2009.
- Vlasoff, M., (1979). "Labour Demand and Economic Utility of Children: A Case of Rural India" *Population Studies*, Vol 33(1), pp. 415-428.
- Webster, C., (1996). "Hispanic and Anglo Interviewer and Respondent Ethnicity and Gender: The Impact on Survey Response Quality" *Journal of Marketing Research*, Vol 33(1), pp. 62-72.
- Weeks, J.R., (1989). "Population: An Introduction to Concepts and Issues" 4<sup>th</sup> Edition. Wadsworth Publishing Company: United States.