

Assessment of the robustness of recent births in
estimating infant mortality using multi-country
Demographic Health Survey data

Malvern Munjoma (Mr)
University of Cape Town

A dissertation submitted to the faculty of Commerce of the
University of Cape Town in partial fulfilment of the requirements
for the degree of Master of Philosophy in Demography

October 2013

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

PLAGIARISM DECLARATION

This research is my original work, produced with supervisory assistance from my supervisor. I have used the Harvard convention for citation and referencing. Each contribution to this dissertation from the works of other people has been acknowledged, cited and referenced. In addition, this dissertation has not been submitted for any academic or examination purpose to any other university.

Signature..... Date: .../.../...

ABSTRACT

This dissertation investigates the robustness of recent births in estimating infant mortality rates from the proportion of deaths observed among births reported in a 24 month period. The Blacker Brass technique is applied to all births reported in the 24 month period and to most recent births in the 24 month period. The study uses birth history data from 76 Demographic and Health Surveys conducted in 16 countries across the developing world between 1986 and 2011. All births (and the deaths of those births) occurring in five 2-year periods before each survey were extracted to obtain five estimates of infant mortality using the Blacker-Brass and direct estimation methods from each dataset. This allows trends in infant mortality for the 10-year period before the survey to be compared and relative errors to be calculated. The results showed a decline in infant mortality in most datasets and are consistent with the United Nations and the World Health Organisation 2013 estimates. The relative errors did not indicate any systematic bias of the Blacker-Brass method applied to all births; however, further investigations showed that the method underestimated infant mortality in the period closest to the survey date in most datasets. Furthermore, the relative errors were positively correlated with the directly estimated level of infant mortality. There were, however, no significant differences in the relative errors across countries. Blacker-Brass estimates were also derived using last births reported in the 24-month period of each DHS dataset. Unlike the estimates which were derived from all births in the 24-month periods most of these estimates were below the approximate 95 per cent confidence intervals of the direct estimates. Therefore, the Blacker-Brass estimates, which used recent births only had more bias. DHS data on last births only, in each of the 24-month period were used to calculate infant mortality using the Blacker-Brass method as in census datasets. This evidence suggests that applying the Blacker-Brass method to census data underestimates infant mortality. We suggest a modification of census questions on recent births to include all births in the past 24 months and vital status of these children on census date to get more robust infant mortality rates using census data. We further propose a simple improvement to the method using all births in the 24 month period that reduces the error by almost 35% relative to the direct method estimates when compared to the original Blacker-Brass method. The revised method is validated using DHS data from four countries that were not part of the initial data used.

TABLE OF CONTENTS

PLAGIARISM DECLARATION	2
ABSTRACT	3
TABLE OF CONTENTS	4
LIST OF TABLES.....	6
LIST OF FIGURES	7
ACKNOWLEDGEMENTS	8
1. INTRODUCTION.....	9
1.1 Background and context.....	9
1.2 Statement of the problem	9
1.3 Objectives of the study	10
1.4 Justification for the study.....	10
1.5 Structure of the dissertation	13
2. LITERATURE REVIEW.....	14
2.1 Introduction	14
2.2 Child and infant mortality in developing countries.....	14
2.3 Direct methods of measuring infant mortality.....	15
2.4 Indirect methods of estimating infant mortality.....	17
3. DATA AND METHODS.....	25
3.1 Data sources.....	25
3.2 Methods	26
3.3 Data quality	32
3.4 Conclusion	39
4. RESULTS	42
4.1 Direct estimation and All Births Method results	42
4.2 Factors associated with the differences in the infant mortality measures	49
4.3 Direct estimation and Most Recent Births Method results.....	52
4.4 Derivation of correction factors to the All Births Method estimates ...	54

4.5	Derivation of correction factors to the Most Recent Births Method estimates	59
4.6	Out-of-sample validation	61
4.7	Discussion	62
5.	CONCLUSION	65
5.1	Introduction	65
5.2	Data quality	65
5.3	Estimates derived from recent births	66
5.4	Adjusting for mortality level – correction factors	66
5.5	Limitations of the study	67
5.6	Areas for further research	68
	REFERENCES	70
	APPENDICES	73
	Appendix A.....	73
	Appendix C.....	80

LIST OF TABLES

Table 3.1: DHS datasets by country	26
Table 3.2: Algorithm for determining exposure to risk.....	31
Table 3.3: Comparison of proportion dead among all births and last births only in the five-24 month period by country.....	35
Table 4.1: Significance difference Test in relative errors by country	50
Table 4.2: Significance difference test in relative errors by level of mortality.....	51
Table 4.4: Significant differences in the relative errors by time periods before survey date..	52
Table 4.5: Correction factors for ABM estimates for each mortality level	59
Table 4.6: Correction factors for MRBM estimates for each mortality level	61

LIST OF FIGURES

Figure 2.1 Children born in the last 24 months-estimation factors for infant mortality.....	22
Figure 3.1: Lexis plane showing calculation of exposure to risk.....	30
Figure 3.2: Sex ratio at birth by time before survey date	37
Figure 3.3: Sex ratios of dead children by time before survey date.....	38
Figure 4.1: Trends of infant mortality rates using the Blacker-Brass and direct estimates by country	43
Figure 4.2 Relative errors for the Blacker-Brass method relative to the direct method by country	45
Figure 4.3: Relative errors for the Blacker-Brass method relative to the direct method by country and time before survey	47
Figure 4.4: Comparison of the Blacker-Brass and direct method by country, DHS data	48
Figure 4.5: Trends of infant mortality using the direct estimates and Blacker-Brass method (using last births) by country	53
Figure 4.6: Relationship between reported deaths in the past 24 months and infant mortality from direct estimation.....	57
Figure 4.7: Relationship between direct method estimates and mortality rates derived from adjusted reported deaths per 1000 live births	58
Figure 4.8: Comparison of the direct estimates and adjusted deaths in the past 24 months.	61
Figure 4.9: Comparison of the Blacker-Brass and direct estimates.....	62

ACKNOWLEDGEMENTS

I wish to recognise the consistent and invaluable support from my supervisor Professor. Moultrie. Professor Dorrington's comments in the initial stages of this dissertation are hereby saluted. I would also like to acknowledge the motivation and inspiration that I got from a colleague and brother in faith Takwanisa Machedze who has always been there to support and encourage me. The role of my two fellow classmates Clara and Boboh in offering ideas and constructive criticism to my work does not go without mention. It is with heartfelt and utmost gratitude that I thank everyone who played a role in this long challenging, however exciting journey of this dissertation. God will surely reward you.

1. INTRODUCTION

1.1 Background and context

Mortality data in developing countries is often insufficient due to the absence of adequate vital registration systems. There is a pressing need in developing countries to come up with accurate measures for child and infant mortality as one of the key health indicators in a population. Demographers have used direct and indirect methods to come up with reliable estimates for child and infant mortality. These methods have been developed over time to improve the accuracy of these estimates.

More specifically for infant mortality, since it is known that the probability of dying is higher in the earlier years of life; various scholars have investigated the robustness of reports of recent births and their vital status as a tool for measuring infant mortality. This dissertation reviews the methods that have been used to measure infant mortality in developing countries where there are no complete and reliable vital registration systems. More specifically, it focuses on one of the more recent methods for measuring infant mortality, that proposed by Blacker and Brass (2005). This method uses the survival status of the most recent births in the past 24 months from the census date and a correction factor to convert the proportion dead of these births into a measure of infant mortality. The accuracy of the method has been questioned by (Blacker and Brass, 2005) and (Hill, 2012b). Therefore this dissertation seeks to assess the accuracy of the method and also suggests ways of improving the accuracy of the estimates based on recent births in estimating infant mortality.

1.2 Statement of the problem

Hill (2012b) argues that the Blacker-Brass method does not produce reliable estimates of infant mortality. The author's basis for this assertion is based on the assumptions used in the method, which are discussed in greater detail later. Blacker and Brass (2005) also express concern about the reliability of the method, expressing concern about the effects of reporting errors, among other reasons. Other than the authors and Hill's objections on the reliability of the method, the method has not been applied and tested adequately. There is no convincing evidence to date for demographers to apply the method in estimating infant mortality rates (IMRs).

1.3 Objectives of the study

This dissertation seeks to measure the accuracy of recent births in obtaining infant mortality measures as proposed by the Blacker-Brass method. To accomplish this, the Blacker-Brass method is applied on all births and deaths reported in the 24 month period before the survey, all births method (ABM) and to the most recent births only in the 24 month period, most recent births method (MRBM). IMRs obtained from the method will be compared against estimates obtained from the direct method of estimating infant mortality using datasets from various countries. Further, the research seeks to adjust the factors used to convert the proportion of deaths of recent births into a measure of infant mortality.

The following research questions will be addressed;

- What are the infant mortality estimates of the ABM?
- What are the infant mortality estimates using the direct method?
- Do the estimates from the two methods differ significantly?
- What are the factors associated with the differences in the estimates?
- What are the infant mortality estimates of the MRBM?
- What is the bias between the direct estimates and the Blacker-Brass estimates using last births only?
- What adjustments can be made to the more accurate method between the ABM and MRBM so that we have a more accurate approach using recent births which is more consistent with the direct method?
- What modifications if any would be required to the data conventionally collected in censuses for the method to be applied to the estimation of IMRs from census data?

1.4 Justification for the study

Developing countries have had a predicament of not having reliable data for estimating child and infant mortality. According to Chalapati, Bradshaw and Mathers (2003) only a third of countries in the world had a complete vital registration system that produced reliable data then. In a study conducted by Mathers, Ma Fat, Inoue *et al.* (2005), out of 46 African countries, only 6 countries, Mauritius, Seychelles, South Africa, Cape Verde, Sao Tome and Zimbabwe had usable data from vital statistics in 2005.

The absence of vital statistics has left most program implementers with no choice but to rely on censuses and other cross-sectional surveys for vital statistics.

Most program implementers now use research findings to roll out interventions which address specific problems. Strategies aimed at reducing child and infant mortality rates rely heavily on data from regular, reliable, studies to help implementers monitor and evaluate the performance of their programmes. This has created the need to utilise cross-sectional data collected during censuses and other major surveys and process it in a usable format.

The methods developed by demographers have been improved over time to improve the accuracy of mortality measures. More specifically, direct and indirect techniques to measure child mortality have been developed over the past several decades. An example can be given of life tables which do not incorporate the effect of HIV/AIDS resulting in significant underestimation of mortality. Some model life tables have adjusted for the effect of HIV/AIDS, such as the Actuarial Society of South Africa (ASSA) HIV/AIDS models (most recent version developed in 2009). However, the effect of treatment and HIV/AIDS which has significantly reduced HIV prevalence may have not been adequately considered yet.

There has been further refinement of models that take into consideration the effect of HIV/AIDS prevalence, treatment and other intervention factors such as male circumcision and PMTCT. Examples include Spectrum projections which use country specific assumptions used for demographic projections and Actuarial Society of South Africa HIV/AIDS (ASSA) models used for the South African population. However, these models use a set of assumptions that are, in some cases, outdated considering that fertility, mortality and migration rates change slowly in some countries but more dramatically in others. As a result, currently infant and child mortality can only be adequately measured using direct estimation methods and complete vital registration systems.

The availability of updated data poses another challenge especially in developing economies which cannot afford to run nationally representative surveys frequently. Most developing countries use census data which are available in usable format a few years after the data collection exercise which is usually conducted each 10 years to obtain child and infant mortality measures. FBHs from Demographic and Health Surveys (DHS) usually conducted after every 5 years have also been used to obtain infant mortality measures.

In general though, both indirect and direct methods of estimating child mortality have their own limitations. For instance, Ewbank (1982) gives an example of bias emanating from indirect methods measuring child mortality pointing out that young women, especially teenagers have a higher risk of losing their children compared to their older counterparts because they lack proper skills of raising children. Therefore the population age structure may have an effect on mortality levels. Blacker and Brass (2005) supported this notion, further classifying these “younger women” to be those below the age of 25. Due to this bias, there is a general recommendation to disregard estimates based on data from these women especially for the 15-19 year old mothers.

Another problem mentioned earlier on, stemming from these methods especially in Southern African countries is HIV/AIDS. Most of the methods assume that child and maternal mortality are independent and yet in practice, this assumption is heavily violated because in most cases if the mothers are dead, their children are more likely to be dead as a result of vertical transmission of HIV. Mutemaringa (2011) provided evidence to support this notion. However, some authors further clarify the effect of HIV/AIDS on mortality measures. For instance, Artzrouni and Zaba (2003) and Mahy and Zaba (2005) measured the extent to which HIV/AIDS may affect the estimates of child and infant mortality using birth histories, owing to the fact that women less than 25 years old would be too young to have died of AIDS after giving birth. A conclusion was reached that it was not worth correcting for this bias in direct estimates since the error only ranged between 5 – 7 per cent. However, the authors suggested a correction if the data used are for births which occurred more than 5 years before the survey. However, Mahy (2003) and Mahy and Zaba (2005) note that the assumptions made in calculating this correction were irrelevant in a setting of high HIV epidemic hence it cannot be relied on. Further research has been done in an attempt to correct for the bias instigated by HIV/AIDS in estimating mortality by Walker N, Hill K and F (2012). The authors noted that time before the survey, changes in the HIV/AIDS epidemic and other mortality causes not related to HIV/AIDS all have an effect on HIV/AIDS induced bias. Even though results from this paper managed to show the extent of the bias introduced by HIV/AIDS and a method to correct for the bias, substantial factors affecting the accuracy of the results such as the effect of anti-retro viral therapy were ignored.

As a result of these arguments surrounding the estimation of child and infant mortality together with the shortcomings of direct and indirect methods, it has become important for researchers to scrutinise the demographic estimation methods by further testing the current ones across diverse datasets, to identify errors and further refine them to come up with more accurate techniques. The use of the survival status of recent births to calculate child and infant mortality rates has become increasingly important. This is mainly because the estimates obtained are more recent and therefore provide latest information that can be used to for immediate and urgent policy reforms.

The Blacker-Brass method mentioned earlier can estimate IMRs for very recent periods. The method involves counting the number of deaths that occurred from births in a 24-month period obtaining a proportion of children dead. The authors then propose applying a conversion factor, constant across all mothers' age groups to convert the proportion into an infant mortality measure. The method was applied to a few datasets and according to Blacker and Brass (2005) the results were plausible and consistent with estimates calculated from other methods. However, the method has been criticised by the authors and Hill (2012b) and moreover applied to very few datasets.

1.5 Structure of the dissertation

The dissertation is organised in five chapters. Chapter 2 reviews literature on child and infant mortality in developing countries. A more detailed description of the direct and indirect methods used to measure child mortality is also presented. The following chapter (3) describes the data sources and methods used in the research. This chapter explains how the IMRs were calculated using the Blacker-Brass and direct methods. The quality of the DHS data used is also investigated with regards to potential sources of errors in deriving the IMR. Missing data, accuracy of births and deaths reported in the 24-month period before the survey, sex ratios and birth and death ratios are discussed. In chapter 4 results of the ABM and MRBM are presented. Comparison of these estimates with direct method estimates are done together with factors associated with the differences in relative errors of the IMRs. Correction factors are also applied to the ABM and MRBM based on the direct estimates. A comprehensive discussion of these findings concludes this chapter. The final chapter is a conclusion of all the key findings of this research. It also summarises the key findings and suggests areas of further research.

2. LITERATURE REVIEW

2.1 Introduction

This chapter reviews levels and trends in child and infant mortality rates in developing countries particularly those in sub-Saharan Africa in section 2.2 and the methods used to measure infant mortality rates when vital registration systems are not reliable are also presented. The direct methods used to measure infant mortality are described and the assumptions involved with using these methods are also presented in section 2.3. The reliability of these methods is also discussed with regards to where and how they have been applied. Further, this chapter presents indirect approaches in measuring infant mortality in section 2.4 and assumptions used in the methods and the reliability of these methods where they have been applied is described. The Blacker-Brass method used to measure infant mortality will be described in more detail and the contexts in which it was applied will also be reviewed to ascertain its usefulness and accuracy in sub-section 2.4.2.

2.2 Child and infant mortality in developing countries

The developing world has the highest child and infant mortality rates compared to the developed countries according to the Inter-agency Group for Child Mortality Estimation (IGME) (2013). IGME reports that out of every 1000 live births, infant deaths ranged between 20 to over a 100 between 2008 and 2011 in developing countries whereas in the developed countries, an average of only 4 deaths occur per 1000 live births.

This huge difference reflects the influence of different mortality determinants between the developed and developing countries. Most developing countries have had socio-economic problems which affected the health delivery system over the past decades according to Cutler, Angus and Adriana (2006). Further, the HIV/AIDS epidemic increased mortality rates to very high levels and this resulted in increased family dependence burden especially in families which experience adult deaths.

The differences in levels of child and infant mortality between the developed and developing countries has called for concentrated efforts in the developing world to avail correct and updated infant mortality measures to reduce mortality burden.

Various humanitarian organisations over the past few decades have rolled out diverse poverty alleviation and HIV/AIDS interventions aimed at reducing child mortality. In particular the United Nations Development Programme (UNDP) has set a target to reduce child mortality by two-thirds between 1990 and 2015. Most child deaths in the region have been caused by malaria, diarrhoea, low respiratory infections malnutrition and HIV/AIDS according to IHME (2013). Improvements in the roll-out of HIV/AIDS interventions such as behaviour change programs, Prevention of Mother to Child Transmission (PMTCT) services and availability of malarial drugs are among the key factors that have helped in reducing the mortality rates.

Before the introduction of modern health care interventions mortality was very high. Afterwards there has been a notable decline. Garenne and Gakusi (2006) published estimates of child and infant mortality rates in the sub-Saharan Africa region. In their findings, they noted that there has been significant mortality decline of about 1.8 per cent per year in the second half of the twentieth century. However, Garenne and Gakusi identified different country-specific patterns in the decline caused by the socio-economic and political circumstances in those countries. In some countries there has been a monotonic decline between 1950 and 2000 after which mortality increased again in the peak of the HIV/AIDS epidemic. These countries include Botswana, Ethiopia, Liberia, Malawi, among others.

These changing trends in child and infant mortality in developing countries due to the HIV/AIDS epidemic and the effect of interventions has made it difficult to obtain updated and reliable estimates in developing countries thereby inevitably relying on direct and indirect methods. Both methods use a specific set of assumptions that have been appraised and criticised by various scholars.

2.3 Direct methods of measuring infant mortality

Direct methods have been used mostly to obtain estimates of infant mortality. This method does not use life tables or data on fertility patterns and regarded as the most robust way of obtaining infant mortality measures. In most countries where vital registration is incomplete, IMRs have been calculated using FBHs from DHS datasets. Adetunji (1996) defines direct methods as those which use data from vital registration systems or derived directly from vital events reported in cross-sectional surveys. A description of the method, the assumptions it uses and where it has been applied is presented in this section.

2.3.1 Method description

Sullivan (1972) identified three approaches to direct estimation of child and infant mortality namely the vital statistics method, the true cohort life table approach, and the synthetic cohort life table approach. Rutstein and Rojas (2006) also reported the same techniques and give a further explanation of the methods as briefly described below.

2.3.1.1 *Vital statistics method*

This method requires data from a vital registration system. The infant mortality rate is obtained by dividing the number of deaths under the age of one by the births in that period Rutstein and Rojas (2006). However this approach yields accurate estimates only if the registration of births and deaths are complete, hence it is of very limited application to most developing countries which have incomplete registration databases. If incomplete data are used, there may be under or overestimation of mortality rates. If there is evidence to suggest that deaths and births are not registered to the same extent nationwide, then the estimates may be biased. For instance, if the registered births are under reported relative to the deaths, the mortality rates are overestimated and if they have higher coverage there is underestimation of the mortality rates.

2.3.1.2 *Cohort life table method*

To obtain the infant mortality measure (probability of dying before the age of one) using this method, the number of deaths to children under the age of 12 months are divided by the number of births in the same cohort. Here, all recent births are not used. Rather, the number of births which occurred before the age of one in the period 12 – 23 months before the survey is divided by deaths which occurred before first birthday. This is done so that there is equal exposure to risk of dying in one year among all births used in the calculation. Hence the probability of death obtained may apply to a more distant time period in this case one year before the survey date. Rutstein and Rojas (2006) stress the need for all children to which the rate applies to be fully exposed to the risk of dying within the specified time. Since the rates obtained are for a defined cohort, there is no precise time reference for the calculated under-five mortality rates. Another major weakness of the method is that if births for older mothers are used, the time reference of these estimates will become more outdated.

2.3.1.3 Synthetic cohort life table method

In this method, infant mortality is calculated in a two-stage process as described by Rutstein and Rojas (2006). The probabilities of survival of the children in each age group, (life table $p(x)$) are calculated using the age groups 0, 1-2, 3-5 and 6-11 months. To get the probability of dying before the age of one, survival probabilities are used as illustrated by the formula;

$${}_1q_0 = 1 - \prod_{i=0}^{11} p_i, \text{ for } i = \text{age in months, } p \text{ the survival probability and } {}_1q_0 \text{ is the}$$

probability of dying before turning one i.e. the infant mortality rate.

This method relies on the assumption that if the mortality rates have been constant over lengthy period of time depending on whether it is infant or under-five mortality (longer for under-five mortality and shorter for infant mortality), the age-specific death rates obtained from a survey will be an unbiased estimator of the age specific death rates for a real cohort of children. The validity of this assumption is based on the fact that if mortality has been constant, the bias caused by omitting both births and deaths will not be large enough to cause a significant difference.

2.4 Indirect methods of estimating infant mortality

The initial idea of these methods was developed by Brass, Coale, Demeny *et al.* (1968) using data on all children born and those who are alive on the date of the survey. Data used in the indirect estimation of mortality are usually from censuses and a lot of assumptions are made in deriving the estimates Rutstein and Rojas (2006). Indirect methods are so named because they do not directly use information on age and timing of deaths. An appropriate life table is used to convert the proportion of children dead into a mortality measure. The choice of the appropriate life table depends largely on the prevailing age pattern, mortality level and other demographic factors in a population. These methods are therefore regarded to be less reliable compared to direct methods because a lot of assumptions are made when determining the distribution of births and deaths in converting proportion of dead children into a mortality measure. The reliability of indirect methods was also noted by Fernando (1985) in his research on Feeney's method of measuring infant mortality using simulations. Fernando concluded that the indirect method is appropriate and accurate if the age pattern of mortality approximates the Brass type of mortality age pattern. Therefore, appropriate life tables that suit the required mortality level and pattern are required.

Commonly used indirect methods such as the Children-Ever Born: Children Surviving (CEB: CS) method, previous birth technique and the Blacker-Brass methods are described.

2.4.1.1 Children Ever-Born: Children Surviving method

The method was formulated by Brass, Coale, Demeny *et al.* (1968) and developed by Coale and Trussell (1977). Women of child bearing age, usually 15-49, are asked about the number of children ever born in their life time. Of these children, each woman is asked the number of those surviving at the date of the interview. An average number of children born per woman are calculated for each five year age group. A suitable life table is chosen to convert this proportion dead into a mortality measure. Hill (2012b) mentions that one way of determining which life table to use is to plot the child mortality rates against the infant mortality rates using data from full birth histories. The optimal model choice is chosen from one which best fits these plots. Hill (2012b) also suggests the use of child mortality patterns from neighbouring countries in cases where full birth histories are not available.

Data used for estimating mortality using this method are available in three distinct formats namely; age of mother, time since first birth of the mother and duration of mother's marriage. Depending on the way the data are presented, different procedures are applied to convert the proportion dead into a mortality measure once the appropriate life table is chosen. Each life table is associated with unique coefficients required for this conversion. The next stage involves allocating these mortality rates to the appropriate time period to which they apply. This process also depends on the selected life table and on how the data are presented as described earlier. The final stage, as described by Hill (2012b), involves converting the probabilities of dying (${}_xq_0$) into a mortality level parameter of a relational logit model life table. The parameter is then used to calculate the probability of dying before the age of one, for infant mortality. The CEB: CS method uses a set of assumptions whose reliability has been disputed by various scholars. The first assumption used is that the age-specific fertility and mortality levels have been constant for a significant period of time. The mortality of children is also assumed to be independent of their mothers. This assumption is heavily violated particularly in populations where the HIV/AIDS prevalence is very high. This is because there is a high correlation of death between infected mothers and their children due to vertical HIV transmission, Hill (2012a). Another assumption made is that the mortality rates are constant across all age groups of mothers.

There is however, significant evidence from most DHS datasets that shows that mortality is higher in younger mothers especially those in the 15-19 age group. The 15-19 year olds age group is usually omitted for this reason.

Due to violation of most of these assumptions, there is a general recommendation to triangulate estimates obtained by this method with other methods.

Attempts have been made to adjust for the effect of HIV/AIDS in the method. Ward and Zaba (2008) carried out a study to check for the effect of HIV/AIDS on the method in estimating childhood mortality. The results of the study revealed that if the prevailing prevalence was between 5 – 10 per cent, the resulting under-five mortality would be biased by at least 5 per cent. Ward and Zaba also suggested a technique of adjusting the effect of HIV/AIDS using sero-prevalence and assuming that the epidemic and the population are stable. Mutemaringa (2011) also estimated the extent of the bias induced by HIV/AIDS for infant and mortality rates derived using the children ever born and children surviving method and results of the study showed a 15 per cent bias for women aged between 25-39 in six countries and in two of the countries an overall bias of 30 per cent was noted. Darikwa and Dorrington (2013) further developed Ward and Zaba's method by calculating child mortality estimates adjusted for HIV/AIDS in a non-stable population. The mortality estimates obtained from this method for were consistent to the IGME and ASSA estimates for 2006.

2.4.1.2 Preceding Birth Technique method

The idea of the method was formulated by Brass (1969) and the first complete method was proposed by Brass and Macrae (1984) with the aim of simplifying the procedure of calculating infant mortality using vital data of the second most recent birth from mothers who visit a health centre. The key question asked of respondents is the vital status of the child born before the most recent birth. Brass and Macrae mention that if preceding births are used, using data from women at a health centre, the children considered are of the same age range. Therefore all the births would have been equally and fully exposed to the earliest periods of high mortality because they are of similar age. This is not the case if the normal birth orders are used. The total number of deaths which occurred from these preceding births is then divided by the total number of preceding births to get the proportion of dead children. An adjustment is then made to cater for the fact that, with increasing age, death rates of children fall. This method was applied by Aguirre and Hill (1988) using maternal data from health centres.

Hill and Aguirre (1990) extended the method to include adjustment factors that correct the biases which emerge if the sampling frame used consists of only women whose last child is alive. If this adjustment is not done, this results in underestimation of mortality. Hill (2012b) cites Oliveras, Ahiadeke and Hill (2008), Bicego, Augustin, Musgrave *et al.* (1989) and Madi (2000) among others who have applied the method using data from ante-natal clinics and refugee camps in impact evaluation studies for health interventions. In these studies the general conclusion was that the method produced reasonable estimates.

However, some scholars have identified some weaknesses in the method. For instance there is inadequate coverage of all mothers whose births are exposed to risk of dying since the sampling frame consists mainly of respondents from maternity centres. In some developing countries a significant number of mothers still deliver their babies at home. In a study using DHS data from 48 countries from 2003 – 2011, Montagu, Yamey, Viscont *et al.* (2011) found that about 70 per cent of births of women categorised as poor (lowest two wealth quintiles) occurred at home. The data were extracted from countries in Sub-Saharan Africa, South and Southeast Asia, the same regions where these techniques are applied to measure mortality. In this case, the Preceding Birth Technique (PBT) method would not have included 70 per cent of the poor women in the sampling frame. The inaccuracy of the method is further compounded because there is empirical evidence that reveals a strong correlation between poverty and infant mortality.

2.4.2 The Blacker and Brass method

The Blacker-Brass method is a further extension of the preceding birth technique. While the preceding birth technique uses the second from last births, the Blacker-Brass method uses births which occurred in a period 24 months before the survey date. The formulation of the method, assumptions used, application and its limitations are discussed in this section.

2.4.2.1 Description of the Blacker-Brass method

Blacker and Brass begin by defining the proportion of births in the past 24 months dying (D) as

$$D = 1 - \frac{2L_0}{2} = 1 - \frac{1}{2} \int_0^2 l(x) dx$$

Here, ${}_2L_0$ is the proportion of survivors of unit births in each of the past two years. For this D , the authors suggest that there is a linear relationship with the probability of dying before the age of one (${}_1q_0$). This relationship is defined by the prevailing infant and child mortality pattern in a given population as measured by other methods. Blacker and Brass then use a survival function of persons (in early childhood) alive at age x (in months) based on the assumption that early mortality follows a hyperbolic function;

$$l(x) = (1 + \alpha x)^{-\beta}$$

where α and β are constants which determine the level and shape of mortality as described by Blacker and Brass. This survival function is based on the fact that in early childhood, mortality is relatively higher, and then it decreases steeply in the form of a hyperbolic function. The hyperbolic function has been used by other authors Keyfitz (1966) in describing mortality trends in early childhood ages. Blacker and Brass (2005). The survival function is transformed into a natural logarithm to become

$$\ln l(x) = -\beta \ln(1 + \alpha x)$$

Differentiating this result as described by Blacker and Brass yields;

$\mu(x) = \frac{\alpha\beta}{1 + \alpha x}$, after using the identity $\mu(x) = \frac{-d \ln(l(x))}{dx}$ where $\mu(x)$ is the force of mortality at age x .

The model was then fitted to 120 model life tables with α ranging from 0.2 to 1000 and β from 0.01 to 0.8 and a wide range of infant mortality ranging from 26 to 146 per 1000 deaths. An adjustment factor given by $q(1)/D$ was then calculated to convert the proportion dead in the past 24 months into infant mortality. Further, ratios of the death probabilities in the first two years of life, $q(1)$ and $q(2)$ were calculated. Results of this experiment showed that there was no relationship between the conversion factor and β – the mortality level parameter. However there was a relationship with α – the shape parameter. The conversion factor was plotted against the ratio $q(1)/q(2)$ and the results are shown in Figure 2.1 below. Blacker and Brass noted that the correction factors were between 1.08 and 1.1 in the central part of the range giving a mean of 1.09. The authors adopted this value as the conversion factor.

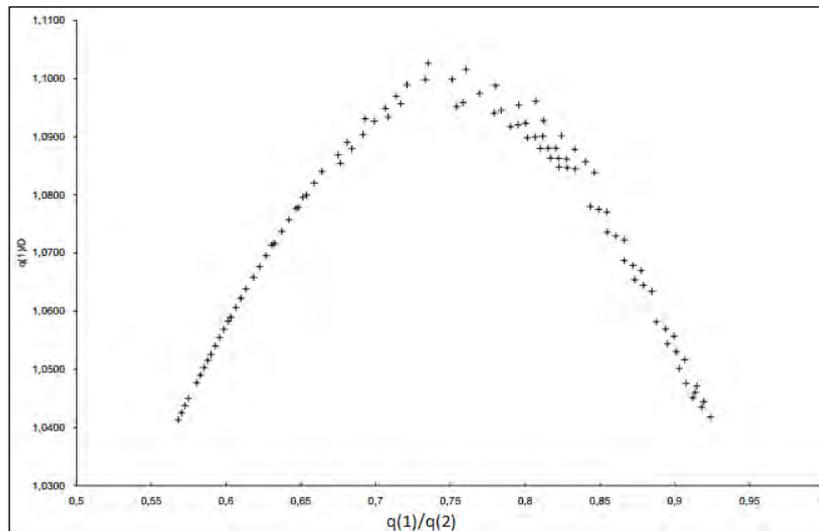


Figure 2.1 Children born in the last 24 months-estimation factors for infant mortality
Source: Blacker and Brass (2005), Figure 1

Further, the model was fitted to some African life tables and conversion factors were calculated and results applied to abridged life tables for 17 demographic surveillance sites with a very wide range of infant and child mortality. Results showed that of the 34 (17 male and 17 female) life tables, 24 had conversion factors ranging between 1.087 and 1.097. The median value calculated was 1.092, a similar value as the calculated mean from the 120 life tables described earlier with results. A similar procedure was repeated using developed countries with low mortality namely Wales, England and Japan. Slightly lower conversion factors of 1.04 for England and Wales and 1.07 for Japan were found.

These results led the authors to conclude that both the shape (α) and level (β) parameters are key determinants in the process of converting the proportion dead in the past 24 months into measures of infant mortality. They further recommend that in cases where the age pattern of a population is not known, 1.09 should be used to convert the proportion of dead children into $q(1)$. In other scenarios, for instance, where there is evidence to show that child mortality is high with respect to infant mortality or if the infant mortality is below 50 deaths per 1000 births, Blacker and Brass proposed to use a conversion factor which is close to but not less than 1.04.

In application of the method, the following assumptions were used;

- Only one birth occurs in the 24 month period used to get births and deaths (i.e. no births after the death)

- Time reference of the estimates of $q(1)$ is assumed to be the mid-point of the 24 month period
- A conversion factor of 1.09 can be used to convert proportion dead into $q(1)$.

2.4.2.2 *Application of the method*

The Blacker and Brass method has been applied to a few datasets from developing countries, those with both different HIV/AIDS prevalence and also high and low child and infant mortality. However, the method has not been used in many other studies. Blacker and Brass (2005) used the Indonesia 1992 DHS survey to test the accuracy of the estimates produced by the method.

Three provinces were selected and proportions dead in the past 24 months were calculated and multiplied by a conversion factor of 1.09 for two provinces and 1.06 for the other. Results of the mortality estimates were compared with those from the 1980 and 1990 censuses and similar results were obtained. However conclusive statements as to whether the method worked or not could not be made since only one country and three datasets were used. Blacker and Brass (2005) also mentioned that the quality of the data used was questionable due to sampling errors. An adequate and sufficient assessment of the method was therefore not possible.

The same authors made another attempt to check the consistency of the method with other estimates used the Kenya 1999 census data. However, the census data were not reliable. They reported that the data were distorted when it was captured electronically particularly affecting data on reported number of female children. Corrections were made, however, leaving a very high degree of uncertainty. Moreover, unexpected results of more female deaths than male deaths were obtained which conflicted with what was known about Kenyan mortality patterns. Despite this fact, the results produced by the method were within range of a 95 per cent confidence interval of other estimates using DHS data.

Nannan, Dorrington, Laubscher *et al.* (2012) also used the Brass and Blacker method to estimate infant mortality for South Africa in 2006. As with the Kenyan example discussed earlier, the results of the study showed higher IMR in girls compared to the boys. This result is unexpected because mortality is expected to be higher in males. The similarities in the findings from these two studies need further investigation before we reach a conclusion on whether the Blacker-Brass method underestimates mortality in males or over-estimates mortality in females or neither. Nannan,

Dorrington, Laubscher *et al.* (2012) however, concluded that produced reasonable IMRs for both sexes combined.

Hill (2012b) argues that the method assumes that in the 24-month period there is only one birth that occurs and this may not be true. Hill suggests that data from DHS shows a significant number of women with more than one birth in the 24-month period. This would then imply omission of deaths from first births of these women which occurred in that period leading to underestimation of infant mortality since early deaths increases chances of second births.

Hill's reason for this assertion is that, if a woman loses her child in early infancy, it is likely that she will have another child soon who is more likely to survive. There is not much evidence of more than one birth in a period of 24 months. One of the most relevant studies in this context was conducted by Moultrie, Sayi and Timaeus (2012) using 76 DHS datasets from 24 countries. The results of the study showed that birth intervals had lengthened. The results of the paper showed median birth intervals for all the countries way above 24 months. However, there may have been a selection bias, for example over-sampling of women who had longer birth intervals hence Hill's argument on short birth intervals remains to be investigated.

This uncertainty surrounding the accuracy of the method leads to the primary research question investigated in this dissertation whether Hill's assertion in 2012 that the method underestimates mortality is correct or not. In order to accomplish this, the need to compare the results produced by this method and other known reliable techniques used to measure infant mortality becomes paramount. One such method that has been used for quite some time is the direct method. This method will be described in detail in the following chapter.

3. DATA AND METHODS

This chapter begins with presenting the data sources used and how they were selected followed by a description of how the mortality measures were calculated. The quality of the data used to obtain the infant mortality measures is also discussed.

3.1 Data sources

The DHS is a nationally representative survey that is designed to capture health and demographic data. There are country specific and generic variables (same variables for all countries) in each dataset. The generic variables are validated and standardised so that users can apply the same analysis techniques on different datasets without changing any variables. The DHS dataset captures data on population demographics, nutrition and health indicators. In this study, DHS data on Full Birth Histories (FBHs) for women aged 15 – 49 are used. Questions are asked about children ever born to a woman in her lifetime and the exact date of births and deaths.

A total of 80 DHS datasets from sub Saharan African countries and other developing countries outside the region were chosen for use in this study. Table 3.1 below shows the selected countries and the corresponding DHS years. In selection of these datasets countries most hit by the HIV/AIDS epidemic were included. Such countries include Zimbabwe, Malawi and Kenya. The population characteristics of Southern African countries and other sub Saharan countries do not differ much hence not many countries were chosen from Southern Africa. Lower HIV prevalence African countries such as Egypt were also included so that the assessment of the infant mortality methods can be done in different HIV/AIDS contexts. Further, developing countries outside Africa such as Cambodia, the Dominican Republic, Indonesia, among others were also chosen so that the methods can be tested in diverse regional socio-economic contexts. Additionally, it was necessary to select countries with more than one DHS survey so that multiple comparisons of the estimates can be done for different time periods. This is also done so that the consistency of the estimates is seen.

Table 3.1: DHS datasets by country

<i>Country</i>	<i>DHS code</i>	<i>DHS years</i>
Bangladesh	BD	1993, 1996, 1999, 2004, 2007, 2011
Bolivia	BL	1989, 1993, 1998, 2003, 2008
Burkina Faso	BF	1992, 1998, 2003, 2010
Cambodia	KH	2000, 2005, 2010
Cameroon	CM	1998, 2004, 2011
Colombia	CO	1986, 1990, 1995, 2000, 2004, 2009
Dominican Republic	DR	1986, 1991, 1996, 1999, 2002, 2007
Egypt	EG	1988, 1992, 1995, 2000, 2005, 2008
Ghana	GH	1988, 1993, 1999, 2003, 2008
Indonesia	ID	1987, 1991, 1994, 1997, 2002, 2007
Ivory Coast	CI	1994, 1998, 2005
Kenya	KE	1988, 1993, 1998, 2003, 2008
Madagascar	MD	1992, 1997, 2008
Malawi	MW	1992, 2000, 2004, 2010
Uganda	UG	1988, 1995, 2000, 2006, 2011
Zimbabwe	ZW	1988, 1994, 1999, 2005, 2010

3.2 Methods

The Blacker and Brass and direct estimation methods were used to obtain infant mortality estimates. The procedures used to calculate infant mortality using both methods is described in detail below.

3.2.1 Blacker and Brass method

The basic information used for the Blacker and Brass method is data on the vital status of children born in a 24-month period. The vital status of these children on the survey day is captured and the month of death for those who died is also recorded. This information was extracted for each dataset described in the previous section to derive the proportion dead of those born alive in a 24-month period. The procedure can be repeated for multiple periods by artificially censoring the data to use births and deaths in 24-month periods.

It is however noted that, for countries with a high HIV/AIDS burden the bias can be significant with increasing time moving back from the survey date.

Only the children who were born and died within the 24-month period were counted as the “relevant deaths” in the method. Data were extracted from the measure DHS website using the DHS birth recode STATA file for each country and survey.

Since the dates of births and age at death from the DHS are in CMC, they are not exact. Hill (2013) describes the process of estimating exact dates of birth and age at death as follows. First, a pseudo random number was generated from dividing v002 – the household number variable – into deciles yielding values between 0 and 9. The random number was divided by 10 and 0.05 was added to the result to get values between 0.05 and 0.95 and the result was added to b3, the date of child’s birth variable to obtain the exact date of birth. To impute an exact age at death, a similar approach is used. A pseudo random number is generated using a similar procedure described above for deriving exact date of birth using the day of interview variable (v016). Hill (2013) summarises the procedure for calculating the date of death as;

For “unit” = 1 (i.e. age at death measured in days), age at death (aad) can be estimated as (“value”+ random2)/31 (for age at death in days this is not necessary, but is described for symmetry); for “unit” = 2 (i.e. age at death in months), age at death is “value” + random2; and for “unit” = 3 (age at death in years), age at death is (“value” + random2)*12.

where “value” is the observed age at death and “random2” is the pseudo-random value generated from v016.

The births and deaths which occurred in the 25th – 48th, 49th – 72nd, 73rd – 96th and finally the 97th – 120th month relative to the date of the survey were extracted to yield five, two-year periods of births and deaths which went up to the 10th year preceding the year of the survey.

To derive the mortality estimates using the ABM the proportions dead of all births in each two year period was derived by dividing the number of all children dead by those born in the same period. These proportions were each multiplied by a factor of 1.09 to obtain the ABM estimates of Blacker and Brass method. This procedure was repeated for all the DHS datasets listed in Table 3.1 and results were obtained for each country using consecutive DHS datasets and a trend of infant mortality for the past ten years before survey date was derived. The STATA ado file used to calculate the Blacker and Brass estimates is provided in Appendix A.

To obtain mortality estimates using the MRBM, the number of children dead among most recent births in each 24 month period described above was divided by the total number of most recent births in the same period. The proportions were also multiplied by 1.09 to yield the Blacker Brass estimates as proposed in the original paper.

3.2.2 Direct estimation of infant mortality

Direct methods use data from FBHs or Truncated Birth Histories (TBHs). FBHs are data obtained from women usually in the age range 15-49 which is regarded as the fertility age group for most women. FBHs include all birth records of each child that was born to a woman in her lifetime. The name of the child, the year and month of birth and month and year of death, if they died, are captured. If the child died within the first month of life, the day of death is recorded. Further, if the death occurs within the first twenty four months of life, the age at death is recorded in months and if death occurs afterwards, it is recorded in years. This aggregation is done so that a more precise age at death is recorded. If this problem is not avoided, a displacement of ages at deaths and more importantly heaping may occur resulting in distorted trends in mortality.

In cases where sex-specific measures are required, data on FBHs also contains the sex of the child. In order to make the necessary calculations, the method makes a couple of assumptions. Firstly, it is assumed that the children who died and those alive are reported with the same level of accuracy i.e. the number of children who died and not recorded is similar to the number of births which were also omitted. This assumption ensures that there is no over-representation of births or deaths in the calculation and hence the principle of correspondence is not violated when dividing the deaths by the number of births that occur in a specific period.

The second assumption relates to the periods over which the births and deaths occurred. It is presumed that the month and year of births and deaths are reported accurately. In most cases, if the biological mother is reporting these events the dates are more likely to be accurate. In addition the closer the date of the event is to the day of interview, the more accurate it becomes and as we move further back in time from interview day, these dates may become less reliable. This assumption is made because the method uses exposure to risk of dying which is obtained from the date of birth and date of death.

Another assumption is made that there is no correlation between mortality of mothers and that of children. This assumption is violated especially in populations where there is high HIV/AIDS prevalence due to vertical transmission of the virus to the child by the mother. Chances are high that if mothers die from HIV/AIDS, the children will also be dead, leading to severe under estimation of mortality. Since the prime objective of this research is based on comparing the Blacker-Brass method and the direct estimation method, correction for bias induced by HIV/AIDS was not done. Nonetheless, a more accurate comparison would be done if correction for HIV/AIDS was done first on the direct method estimates

Cohort or period measures of mortality can be calculated using the direct method. The period measures of mortality were used to obtain the infant mortality estimates for this project.

3.2.2.1 Calculating period measures of mortality

According to Hill (2013), period measures of mortality refer to deaths that occur between specific ages for a defined period of time. Period measures of infant mortality are used in this research project using FBHs from DHS datasets. The objective is to locate the exact time of births and deaths as accurately as possible. To do this there is a need to get an estimate of the date of birth and death for children who died. As described earlier, pseudo random numbers are calculated using non-correlated variables in the dataset so that the calculation is reproduced. This process has an advantage of avoiding heaping at certain ages like the 12th month for example. The process of obtaining the period measures used in this study is described in more detail below.

The process of obtaining an estimate of the date of births and deaths is similar to the procedure used in the Blacker-Brass method as described in section 3.2.1. After these dates were obtained, derivation of exposure to risk of dying then followed.

3.2.2.2 Derivation of exposure to risk of dying

Here, the upper and lower age limits of the children are defined as x_b and x_l respectively for the age range pertaining to a specified period of investigation. The time period for which the mortality measure will apply is also defined to be in range t_1 and t_2 . This age-time scenario defined is presented in Figure 3.1 below. Here, the diagonal lines represent the survivorship of children aged x_i at time t_j where i is some point between x_b and x_l and j is some time point between t_1 and t_2 .

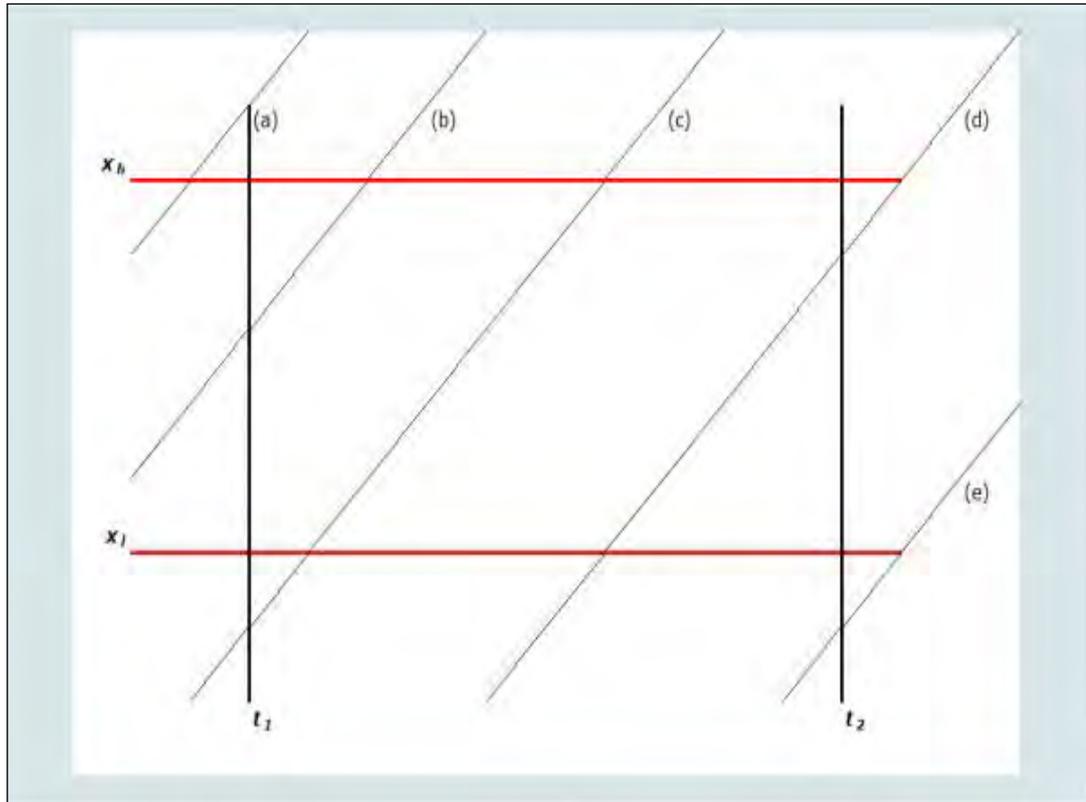


Figure 3.1: Lexis plane showing calculation of exposure to risk

Source: Hill (2013)

The next step was making sure that all births fall in the correct exposure period. Of these births, there were a set of survivors and those who died. Five possible scenarios are presented on the Lexis diagram ranging from a – e as shown in Figure 3.1 above and the description of the rules governing this allocation for the children who dies and those who survived, according to Hill (2013), is presented in Table 3.2 below.

Table 3.2: Algorithm for determining exposure to risk

Scenario	Description	Defining rule(s)	Exposure for survivors in the period of investigation	Exposure for decedents (where death occurs in the period of investigation)
(a)	Aged older than x_h at t_l	$x_{t_l} > x_h$	0	0
(b)	Aged between x_l and x_h at t_l . Attains x_h in the period of investigation	$x_l < x_{t_l} < x_h$ $x_{t_l} + (t_2 - t_1) > x_h$	$x_h - x_{t_l}$	$x_d - x_{t_l}$
(c)	Attains x_l and x_h in the period of investigation	$x_l > x_{t_l}$ $x_{t_l} + (t_2 - t_1) > x_h$	$x_h - x_l$	$x_d - x_l$
(d)	Attains x_l in the period of investigation but period ends before attainment of x_h	$x_l > x_{t_l}$ $x_l < x_{t_l} + (t_2 - t_1) < x_h$	$x_{t_l} + (t_2 - t_1) - x_l$	$x_d - x_l$
(e)	Does not attain x_l in the period of investigation	$x_{t_l} + (t_2 - t_1) < x_l$	0	0

Source: Hill (2013)

Here x_d is the age at death for deaths which occur between t_2 and t_1 .

The following step involved weighting the deaths and number of children exposed to risk of dying and cumulating them. The sample weight variable used was v005.

This variable was divided by 1,000 000 to avoid decimal weights as described by Hill (2013).

Hill (2013) defined the age specific mortality rates $M(x, j)$ obtained by summing all the observation in the sample with n observations with child i , age x and year j as,

$$M(x, j) = \frac{\sum_{i=1}^n D(i, x, j).wgt(i)}{\sum_{i=1}^n E(i, x, j).wgt(i)}$$

Here, $D(i, x, j)$ represent the death of child i , aged x in year j , $E(i, x, j)$ signifies time of exposure of child i , aged x in year j and $wgt(i)$ is the weight of child i . These age specific mortality rates apply to one-month periods of exposure Hill (2013).

The probabilities of dying were then obtained using the age specific mortality rates. To do this calculation it was assumed that the deaths are uniformly distributed across each month. Since the age ranges are quite narrow, the assumption is quite plausible. The probability that a child aged x months in year j dies before reaching age $x+1$ was then calculated using the formula;

$$q(x, j) = \frac{\frac{M(x, j)}{12}}{1 + \frac{M(x, j)}{24}}, \text{ and a product sum of the } 1 - q(x, j) \text{ terms was used to get infant}$$

mortality rate using ${}_1q_0^j = 1 - \prod_{x=0}^{11} (1 - q(x, j))$. The procedure was repeated for the 1st-2nd - , 3rd - 4th, 5th - 6th, 7th -8th and 9th - 10th year before the survey year giving five time-specific infant mortality estimates (${}_1q_0$) from each dataset corresponding to the period described earlier.

The 95 per cent confidence intervals (CI) for the infant mortality proportions calculated using the direct estimates were also calculated as follows

$$CI = \pm 1.96 \sqrt{\frac{p(1-p)}{n}} \text{ where } p \text{ is the infant mortality rate and } n \text{ the sample size. The}$$

calculated confidence intervals are not strictly accurate because they do not allow for the sample design of the DHS hence they are too broad.

The STATA program used to calculate direct estimates of infant mortality and the confidence intervals as explained above was developed by Ken Hill and presented in Appendix B.

3.3 Data quality

Measure DHS has a standard and meticulous data assessment protocol that is applied to all datasets before releasing them. Two sources of errors in calculation of infant and child mortality using DHS data have been identified by Sullivan, Bicego and Rutstein (1990). These are sampling and non-sampling errors. The DHS data processors do not focus much on correcting errors associated with sampling errors.

In most cases these are inevitable and cannot be readily addressed at a data processing or data cleaning stage according to Rutstein and Rojas (2006). Non-sampling errors can, however, be corrected. The errors associated with infant and child mortality are discussed at length in the following section and these include incomplete and unrepresented data, i.e. missing data, misreported dates of birth and death and under reporting of deceased children. These will be investigated in the datasets selected for this project.

3.3.1 Missing data

The effect of the extent of missing data in this research is secondary. Hence data quality in this regard is not done. However it can be noted that there are three sources of missing data in DHS datasets. First, eligible respondents may not be included in the initial sampling frame, in this case, 15-49 year old women due to migration hence there is total exclusion of the household in the survey.

The second source of missing data in DHS data comes from women whose households are included in the survey but were not available for interviews.

Other missing data come from errors of omission by interviewers on specific questions required to calculate infant mortality rates or misreporting date of birth or vital status of child if the mother does not know the whereabouts of their child. Missing such information means a complete discarding of that child's record when calculation infant mortality especially on vital status data of the child. Similarly if the month of death or birth of the child is missing, both the Blacker-Brass method and direction estimation method fails to work completely as exposure to risk are calculated from these dates.

Measure DHS have a mandate to provide quality data that is available in a usable format for analysis. There is a standard protocol that is applied on all datasets before they are made available for users that ensures that this policy is adhered to. Efforts have been put in place in the DHS data collection protocol to reduce the amount of missing data through improving training techniques for interviewers on probing skills especially on acquiring the correct date for births and deaths to avoid heaping. The level of missing data in the datasets considered in this research project is therefore assumed insignificant in distorting the measure of infant mortality calculated.

3.3.2 Number of births reported in 24 month period

One of the key issues raised by Hill (2012b) in criticising the Blacker-Brass method was that, since the method assumes just one birth in the 24 month period there might be bias when estimating infant mortality using census data.

This is because census questions capture vital events of recent births only yet in the 24-month period there could be more than one death.

The author argues that the 24 month period is long enough for women especially in developing countries to have more than one birth. Therefore earlier deaths are not reported leading to an under estimation of infant mortality.

This assumption is required when multiplying the proportion dead to convert it into a mortality rate as mentioned in Chapter 2. Hill argues that this assumption is not accurate because for women with more than one birth in the interval, it is likely that if the first baby dies, the birth intervals that follow are shorter compared to a situation where the first child survives. If the first child dies, and a second birth occurs within that 24-month period, then the death of the first child is not reported. This scenario results in under-estimation of mortality since the death of the first child is not captured.

To measure the extent of this bias in the datasets used in this study, the proportion dead among all births which occurred in each of the five 24-month periods described in 3.2.1 above were compared with proportion dead among last births only in the same 24-month periods for all surveys compared by country. Relative errors of the difference between these proportions were calculated and results are presented in Table 3.3 below.

There is evidence to show that proportion dead among last births only is significantly lower than proportion dead in all births in the 24 month periods in all countries. The highest differences were in the Bangladesh followed by Egypt datasets with relative errors of 48.4 per cent and 47.8 per cent respectively. The Kenya and Zimbabwean datasets had the least underestimation of proportion dead in the combined datasets with relative errors of 27.1 per cent and 27.3 per cent respectively. Nonetheless it is observed that using last births only (as done with census datasets) severely underestimates mortality rates if only the most recent birth in the 24 month period in considered.

Table 3.3: Comparison of proportion dead among all births and last births only in the five-24 month period by country

<i>Country</i>	<i>All births in 24 month period</i>			<i>Last births only in 24 month period</i>			
	Total births	Deaths of all births	Proportion dead	Last births	Deaths of last births	Proportion dead	Relative Error (%)
Bangladesh	86179	5976	0.07	46747	1673	0.04	48.4
Burkina Faso	72983	6073	0.08	29692	1402	0.05	43.3
Bolivia	68576	4456	0.06	32836	1222	0.04	42.7
Ivory Coast	24282	2063	0.08	10486	621	0.06	30.3
Cameroon	44589	2957	0.07	19038	852	0.04	32.5
Colombia	97428	2126	0.02	59142	848	0.01	34.3
Dominican Republic	72837	2654	0.04	37337	965	0.03	29.1
Egypt	129482	6832	0.05	60689	1671	0.03	47.8
Ghana	34283	2228	0.06	15466	658	0.04	34.5
Indonesia	192422	9748	0.05	107097	3200	0.03	41.0
Kenya	69359	3951	0.06	29551	1227	0.04	27.1
Cambodia	52928	3977	0.08	26127	1186	0.05	39.6
Madagascar	45923	2985	0.07	19281	750	0.04	40.2
Malawi	88091	7176	0.08	37135	1972	0.05	34.8
Uganda	65416	4849	0.07	24542	1165	0.05	36.0
Zimbabwe	40985	1968	0.05	21095	736	0.03	27.3

3.3.3 Misreporting dates of vital events

Vital events from DHS data frequently suffer from misreporting errors, especially for recent births. Schoumaker (2012) conducted a study on omissions of recent births using 52 datasets from Sub-Saharan Africa and only one did not indicate omissions of recent births. Forty datasets had statistically significant omissions. The same author also reveals an increase in omission of recent births between 1985 and 2005 for datasets in Sub-Saharan Africa. Mwale (2005) explains the implication of omission and misreporting of vital events mentioning that misreporting of birth dates affect mortality analysis because it results in transferring times of death for the children whose birth dates have been misreported. In addition, the same author also mentions that misreported ages at death distort the age pattern of mortality. The most common example cited by various authors emanates from rounded ages at death. This results in recording deaths in late infancy as if they occurred at exactly 12 months. Since in this study births and deaths which occurred in a 24-month period are considered, the problem would be if heaping occurred on the 24th month.

One of the main reasons for misreporting vital event dates is recall bias by respondents especially if the vital event occurred a long time ago. The problem of forgetting dates is more likely to occur with deaths than births. It is thus important to investigate the distribution of reported ages at interview date and also age at death.

Another source of misreporting of births comes from interviewers. The DHS questionnaire is designed to ask additional questions for children born within five years before the survey. These questions are asked for births that would have occurred from the 1st of January of the 5th year to the survey date. Births that occur in that year are reported to have occurred in the previous year by interviewers according to Hill (2013). They do this in a bid to avoid asking child anthropometry questions. As a result, there is a significant shortage of births in the year following the cut off period and a surplus in the year before the cut off. This concept is known as birth transference. These errors are not investigated, since the primary objective of this dissertation is to compare how accurate the Blacker-Brass estimates are compared to the direct estimates. Further since 24-month periods are used to calculate infant mortality, this cut-off is within a 2 year period so has no effect.

3.3.4 Sex ratios

Sex ratios at birth and death were also investigated to check for implausible patterns. Sex ratios at birth usually range from 100 – 106 in the absence of intervention programs according to Hill (2012b). Sex ratios outside this range are regarded as erroneous. Numbers of children alive at the start of each of the 5 intervals described above were extracted and their sex ratios at birth and at death determined. Results are presented in Figure 3.2 and Figure 3.3 below and the dashed lines represent the range for plausible sex ratios.

The sex ratios at birth were presented by time from each of the surveys separately to get a clearer picture of the trends in over-or under-estimation of a particular sex. Results of this illustration are shown in Figure 3.2. For some countries there is evidence of increasing sex ratios back in time. Examples include Madagascar, 2007; Bolivia; 1997 and 2003 datasets; Cameroon, 1991; Colombia, 1990; Kenya, 1988 among others. In some datasets there were also trends of decreasing sex ratios moving back in time before survey date. Examples are the Cameroon, 1991 DHS (in the later periods i.e. 3rd – 5th intervals); Egypt, 1992; Indonesia, 1987; Malawi, 2004 and also Zimbabwe, 1988 dataset.

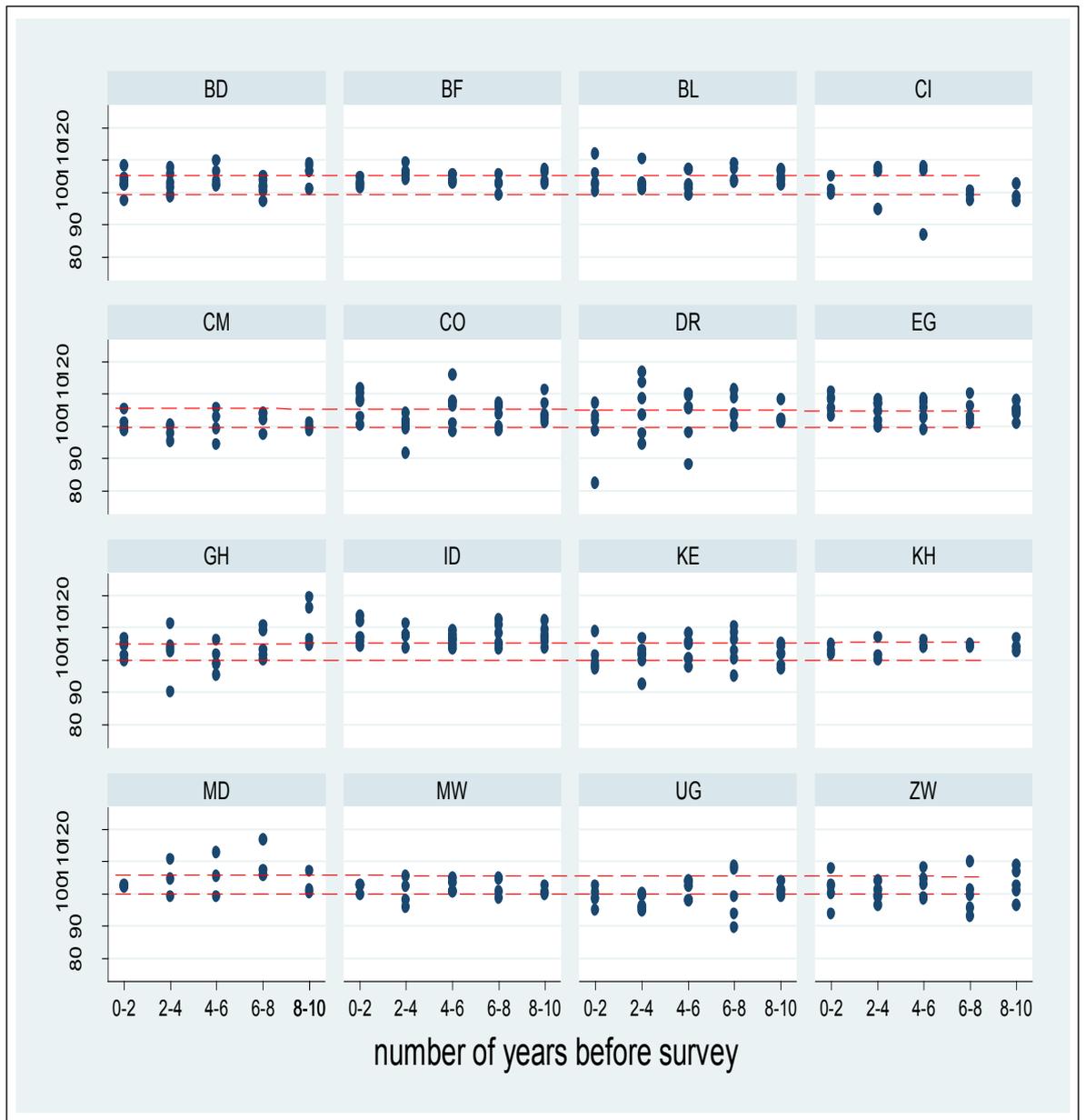


Figure 3.2: Sex ratio at birth by time before survey date

Sex ratios at death (SRAD) were calculated for children who died within a 24 month period and results presented in Figure 3.3 below. SRAD were as low as 80 and as high as 182.

In some countries all the datasets had more male deaths than females. These countries are Indonesia, Cambodia and Madagascar. However, about 18 per cent of the datasets had SRAD below 100 showing more female deaths relative to males. Datasets with the lowest SRAD included Zimbabwe's 1994 dataset with 86 for the data point in the first 24 months before the survey, Ivory Coast in the 1999 survey with 88 on the 3rd interval data point and Kenya in 1993 with 87 on the 4th interval data point. The implications of these observations are discussed in the following section.

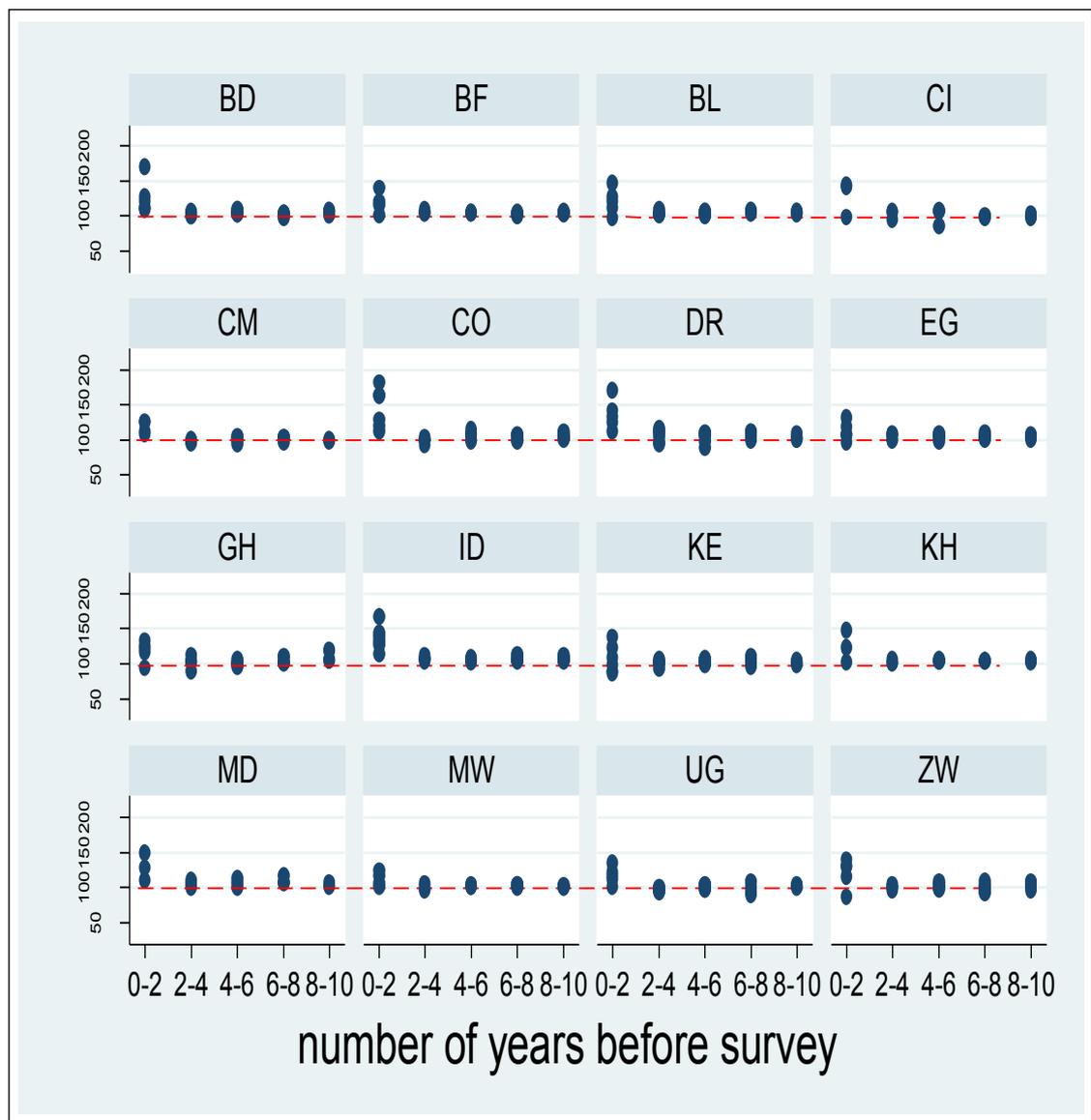


Figure 3.3: Sex ratios of dead children by time before survey date

3.4 Conclusion

This chapter introduced the DHS datasets used to calculate infant mortality rates using the Blacker-Brass and the direct methods. The procedure used to obtain the estimates was also presented and the quality of the DHS data examined. It was acknowledged that effect of HIV/AIDS on infant mortality estimates is significant especially for data pertaining to periods more than two years before the survey. However, correction for this bias was not done as the primary objective of the research is to compare direct method estimates with the ABM and MRBM estimates. Three sources of error that may affect the accuracy of the mortality estimates were identified. These were errors due to missing data, misreported dates of vital events and under reporting of deceased children. A more detailed summary of the results of the data quality assessment is presented below.

3.4.1 Sampling errors

The sampling errors associated with data in terms of the design effect of using cluster samples instead of simple random sampling were considered to be of limited effect since the main objective of the study was to compare the differences in the estimates between the two methods. Therefore the effect of sampling on the estimates cancels out as the same datasets were used in both methods.

3.4.2 Missing data

Three possible sources of missing data were identified. These were exclusion of eligible women in the initial sampling frame, missing of eligible women who were included in the sampling frame, however not available for interviewing for various reasons and finally omission of questions on vital events by interviewers. There were no further attempts to fill in missing data. As such, data were extracted from the Measure DHS website and analysed in that exact form. Since Measure DHS data are widely used and considered to be reasonably accurate, missing data were considered trivial in this project. Further, the data were used for comparative purposes; hence missing data did not have an effect.

3.4.3 Number of births in 24 month period

Results shown in Table 3.3 are evidence that using the most recent births only which occurred 24 months before the survey results in severe under estimation of mortality. Datasets from Bangladesh and Egypt had the highest underestimation of proportion of dead children relative to considering all births reported in the 24-month periods.

Overall most datasets show that using most recent births only underestimates the true proportion of children dead in the 24-month period. Violation of the assumption for converting deaths into mortality measure using the Blacker-Brass method is therefore higher in datasets from Egypt, Cambodia and Indonesia and lowest in Ghana, Ivory Coast and Uganda.

3.4.4 Misreporting of vital events

3.4.4.1 Sex ratios at birth and death

An investigation of the data quality in terms of SRAB and SRAD was executed by country for each survey.

Sex ratios at birth in the range 100 to 106 were considered plausible. Twenty-four DHS datasets had SRAB above 106 and 16 had the sex ratios below 100, an indication of errors in the reporting of deaths and births by sex.

However, the main issue considered on investigation of sex ratios is to check for a particular trend of the ratios moving further back in time from the survey year. Hill (2013) mentions that if the sex ratios increase moving back from survey date this may suggest an under-reporting of female births. If there is under-reporting of females relative to males, this could result in an over estimation of the IMR since there is more male deaths at infancy relative to females. Mortality estimates calculated using the ABM and MRBM are also affected by this misreporting. Out of the 77 datasets used in this study, 13 had increasing SRAB moving back from survey date. Hence, the IMRs in these datasets may have been slightly over –estimated. In other datasets as noted in subsection 3.3.4 the SRAB decreased with time indicating under-reporting of male births in the recent past. Seven DHS datasets had this pattern. Using the same assumption about more males deaths at infancy the IMRs obtained from these datasets may be slightly over-estimated.

Most of the SRAD were above 100 and this is expected for infant mortality rates. However, in some datasets the sex ratios were too high indicating a very high mortality level in male babies. There is a notable pattern of very high SRAD in the period 0-24 months before the survey relative to other periods further back in time. Even though the SRAB was above 100, in most datasets, the distribution pattern is not the same as the SRAD. This suggest an over estimation of male deaths for children born within the first 24 months before survey date across all countries. Therefore the IMRs in this period might have been overestimated in these datasets.

The data quality investigations presented may have an effect on the accuracy of the infant mortality measures. It is noted therefore that these estimates may be less accurate compared to a scenario where steps were taken to address the data quality issues such as correction for bias induced by HIV/AIDS and correcting the irregularities of sex ratios. Nonetheless, the primary objective of this research is based on comparing the infant mortality measures of the direct methods with ABM and MRBM estimates hence the inaccuracies may be similar across the different methods.

4. RESULTS

This chapter presents the infant mortality rates obtained from application of the ABM, MRBM and direct methods to the DHS datasets presented in Table 3.1. Relative errors of the estimates between the ABM and direct methods and also the MRBM and the direct methods are presented and statistical significance tests are done to check if the two methods significantly differ from the direct method estimates. Further in this chapter, factors associated with the relative errors between the estimates are also reviewed and tested for association in order to attribute the differences to certain characteristics in the data. The chapter concludes with adjusting the correction factors used in converting proportions dead in the past 24 months to get more accurate estimates using the direct method as the basis for both the ABM and MRBM estimates. Finally, an out-of-sample validation is conducted to test the revised method against data that did not contribute to the factors proposed under the revised method.

4.1 Direct estimation and All Births Method results

Figure 4.1 below presents the results of the two methods together with the approximate confidence intervals calculated from the direct estimation method by country and the appropriate time location as described in the previous chapter.

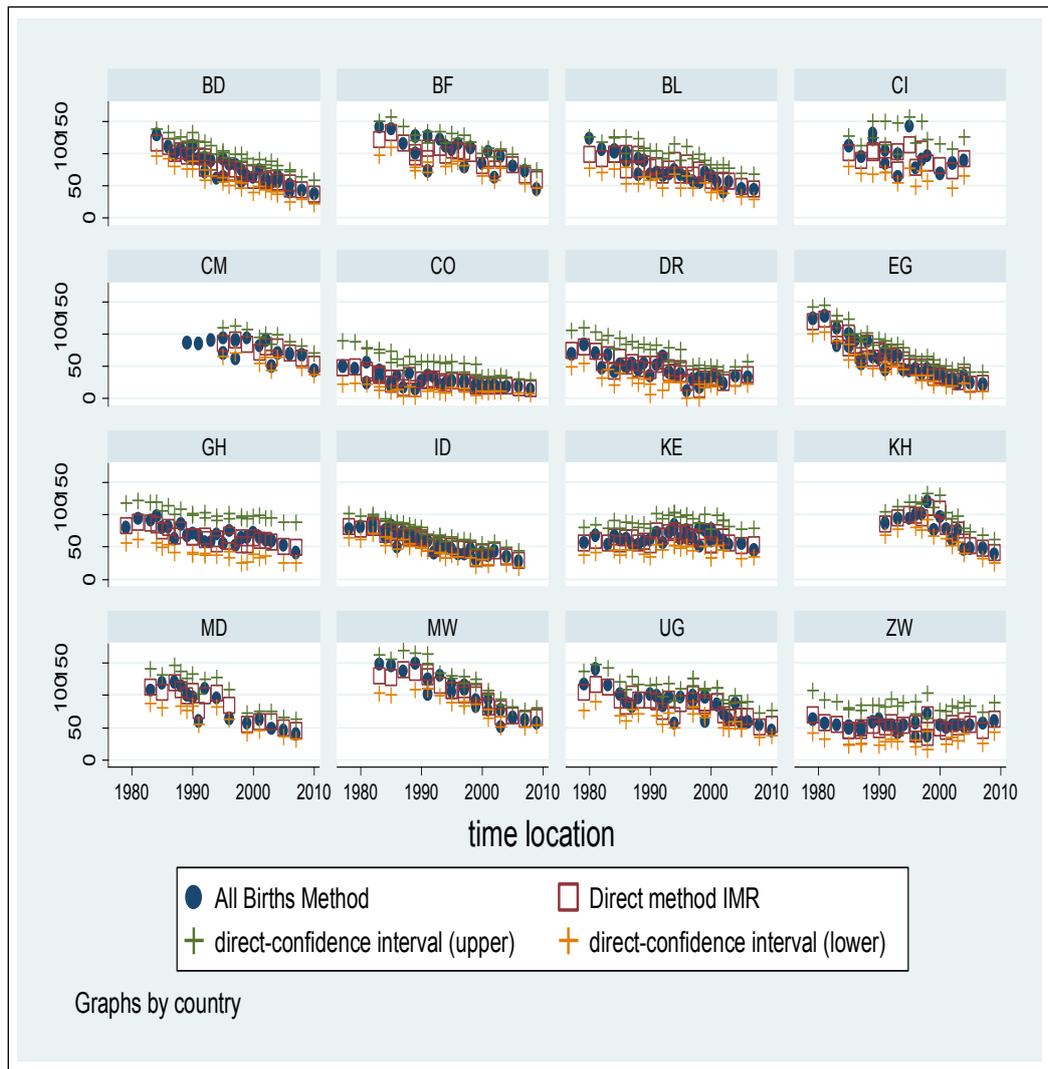


Figure 4.1: Trends of infant mortality rates using the Blacker-Brass and direct estimates by country

Across all country data sets included in the investigation, the results of the direct estimation of mortality ranged from 14 to 140 deaths per 1000 live births using datasets from 1986 to 2010. On the other hand, the ABM estimations ranged from 11 – 149 deaths per 1000 live births over the same range of time. Colombia had the lowest estimates for both methods recording about 14 deaths per 1000 live births in 2008. Malawi had the highest mortality estimates in the 1980s recording about 149 deaths per 1000 live births in 1989 from the ABM.

Both methods indicate declining child mortality over time in almost all countries. However, Ivory Coast does not show any distinct pattern as the estimates for both methods are scattered with no particular trend especially in estimates obtained by the ABM. Zimbabwe, Ghana and Kenya show an almost constant level of mortality for 3 decades for both methods. Indonesian results have the narrowest confidence intervals indicating a better level of precision in the estimates relative to other countries.

The direct method appears to show a more linear pattern in the estimates compared to the ABM. However, it is notable that all the estimates from all datasets calculated by the ABM lie within the 95 per cent confidence intervals of the direct estimation method. This indicates a reasonable level of accuracy of the method relative to the direct estimation method. However, more investigations need to be done to ascertain the differences in the two methods.

4.1.1 Differences in the Infant Mortality Estimates

To gain more insight into the difference between the estimates produced by the two methods over time, a plot of the relative errors for all the estimates obtained against the year of survey is presented in Figure 4.2 below. The relative errors were calculated using the formula;

$$relative\ error = \frac{ABM\ estimate}{Direct\ method\ IMR} - 1$$

The plots are centred around zero and are roughly horizontal showing no evidence that the ABM systematically over or underestimates infant mortality measure relative to the direct estimation method. Most of the values are in the range ± 0.2 of the relative error across all the countries indicating an error range of 20 per cent. In general, the two methods do not show a systematic difference.

In particular, the relative errors are much smaller in four countries showing that the two methods do not differ much in these countries.

In the earlier periods, the relative errors are above 0 for Bolivia, Ghana, Kenya, the Ivory Coast, Malawi and Uganda. In these countries, this shows that the ABM produced higher estimates relative to the direct estimates. In the periods closer to the survey date, the relative errors are much lower in most countries.

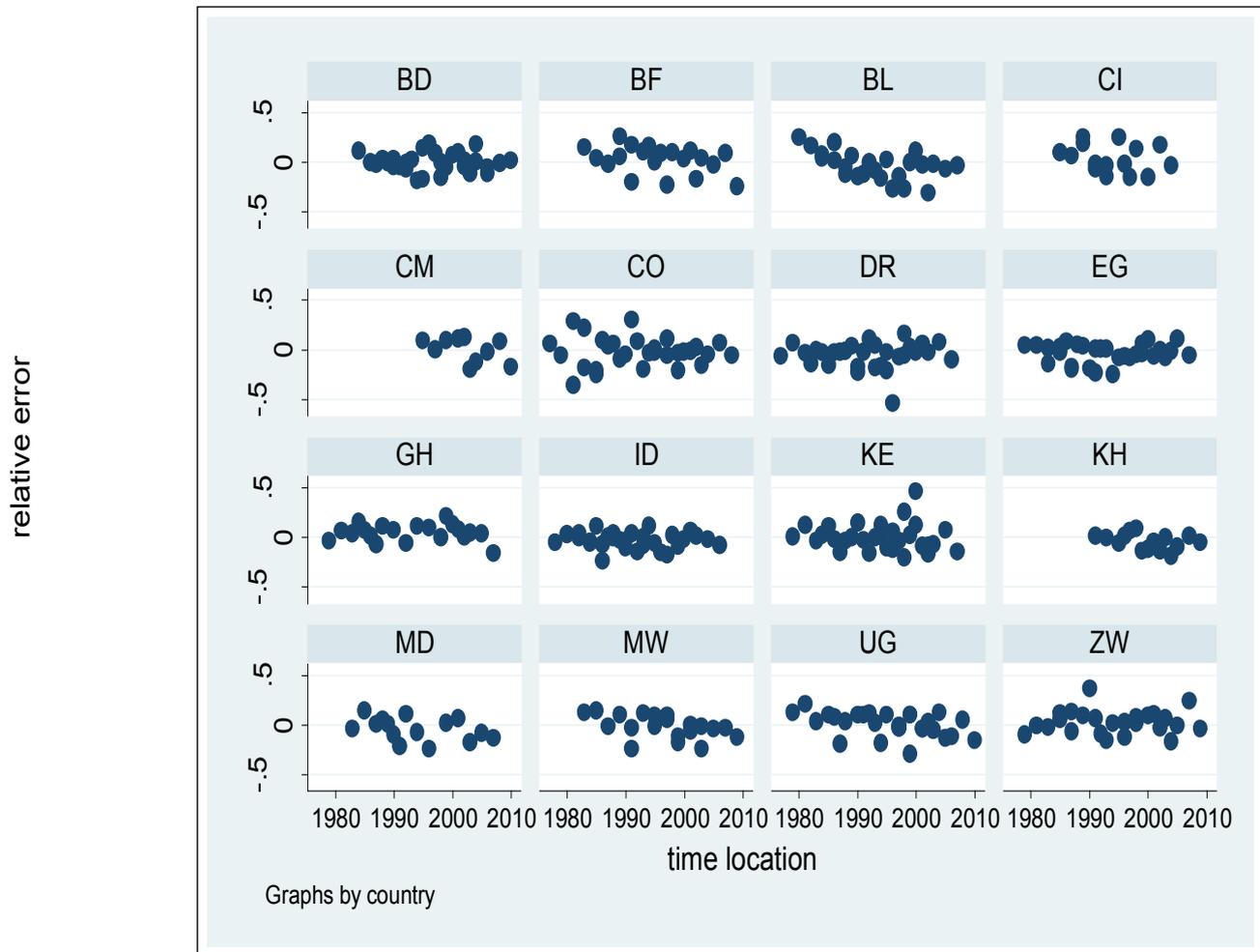


Figure 4.2 Relative errors for the Blacker-Brass method relative to the direct method by country

As noted earlier, the estimates produced by the ABM use all births which occurred within a period of 24 months before the survey or census date. An extension of this method to use births from a more distant past was proposed in section 3.2. It is therefore worth investigating whether the relative errors differ with increasing time before survey date. The time points before survey date on the x -axis refer to the midpoint of the 24-month intervals described earlier. A plot of these relative errors is presented in Figure 4.3 below. There is a similar pattern across all the countries where relative errors of the mortality estimates of the first 24-month period are all below zero. These points mean that the ABM mortality estimates for the period immediately preceding the survey are lower than the direct estimates. In some countries such as Burkina Faso, Ghana and Uganda, most of the relative errors except those in the first interval corresponding to the first 24 months before the survey date are above zero.

Most of the relative errors from Egypt, Dominican Republic, Indonesia, Colombia, Cambodia and Bolivia are below zero. This suggests that the ABM yields lower estimates in these countries relative to the direct estimation method. On the other hand, most datasets from Burkina Faso, Kenya, Ghana, Uganda, Zimbabwe, and Ivory Coast have relative errors above zero in all the periods except the first 24 months before the survey. This means that in these countries, the ABM produces higher estimates compared to the direct estimation method. The effect of time location on the relative errors needs further investigation to ascertain its effect on the infant mortality measures.

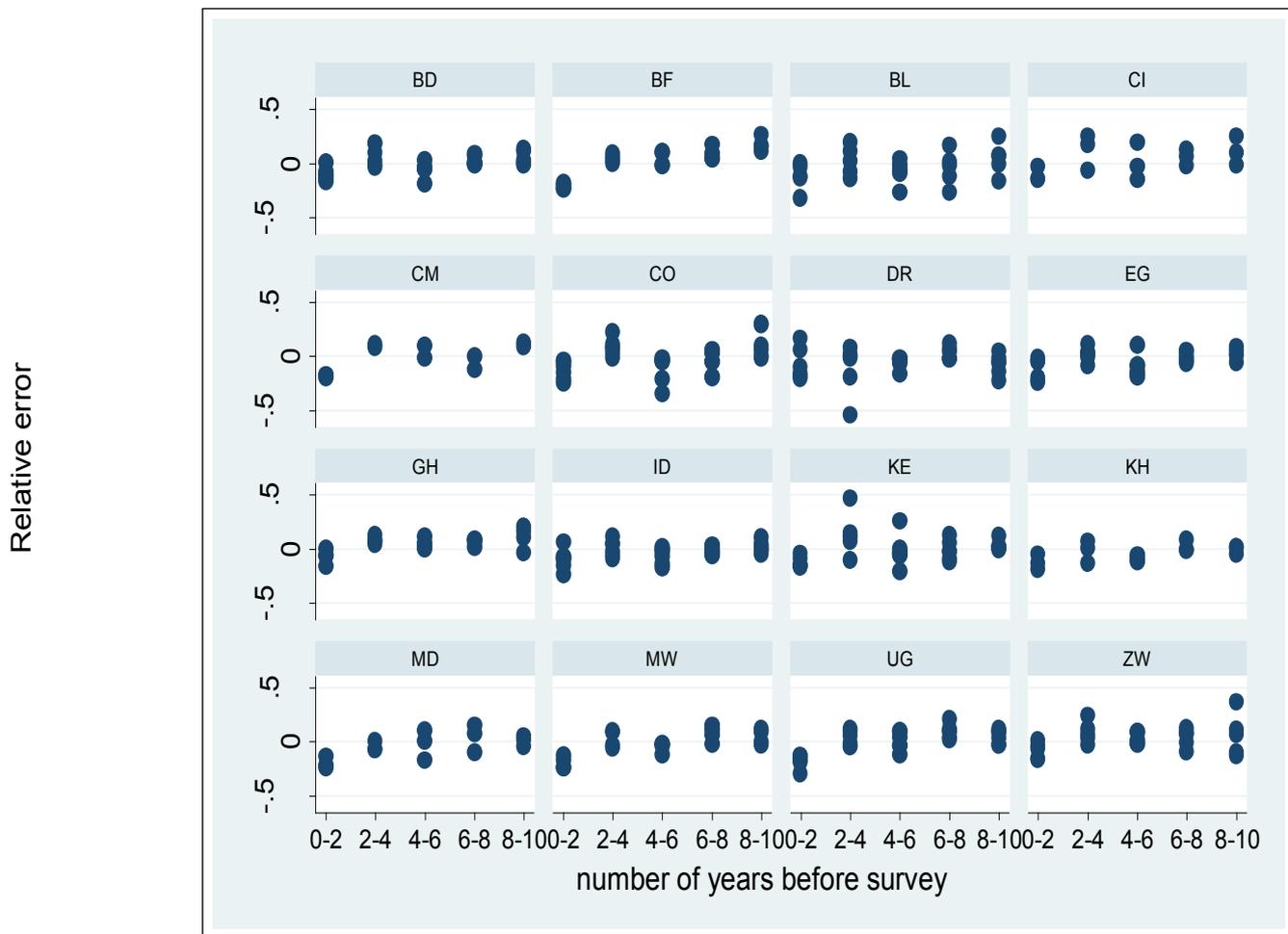


Figure 4.3: Relative errors for the Blacker-Brass method relative to the direct method by country and time before survey

Information on the difference between the two methods is not sufficient to make a conclusive statement on whether the ABM produces statistically different estimates to the direct estimation method. Figure 4.4 compares the mortality measures of both methods to shed more light on whether the ABM over-estimates or underestimates the rates. A 45° line ($y = x$) was plotted as a visual aid for this comparison. About 47 per cent (189 out of 405 data points) of the ABM estimates produced estimates that are higher than the direct method estimates for all countries combined. Most of the plots lie above the line $y = x$ when infant mortality rates are higher than 100 deaths per 1000 live births. This suggests that there could be some systematic differences in the two methods if the underlying mortality estimates are higher. However this result should be interpreted with caution because of the systematic bias introduced by HIV/AIDS moving back in time from survey date.

More than 50 per cent of the ABM estimates above the reference line are in datasets from Burkina Faso, Bangladesh, Ghana, Kenya and Zimbabwe showing that more surveys in these countries births resulted in higher estimates in the ABM.

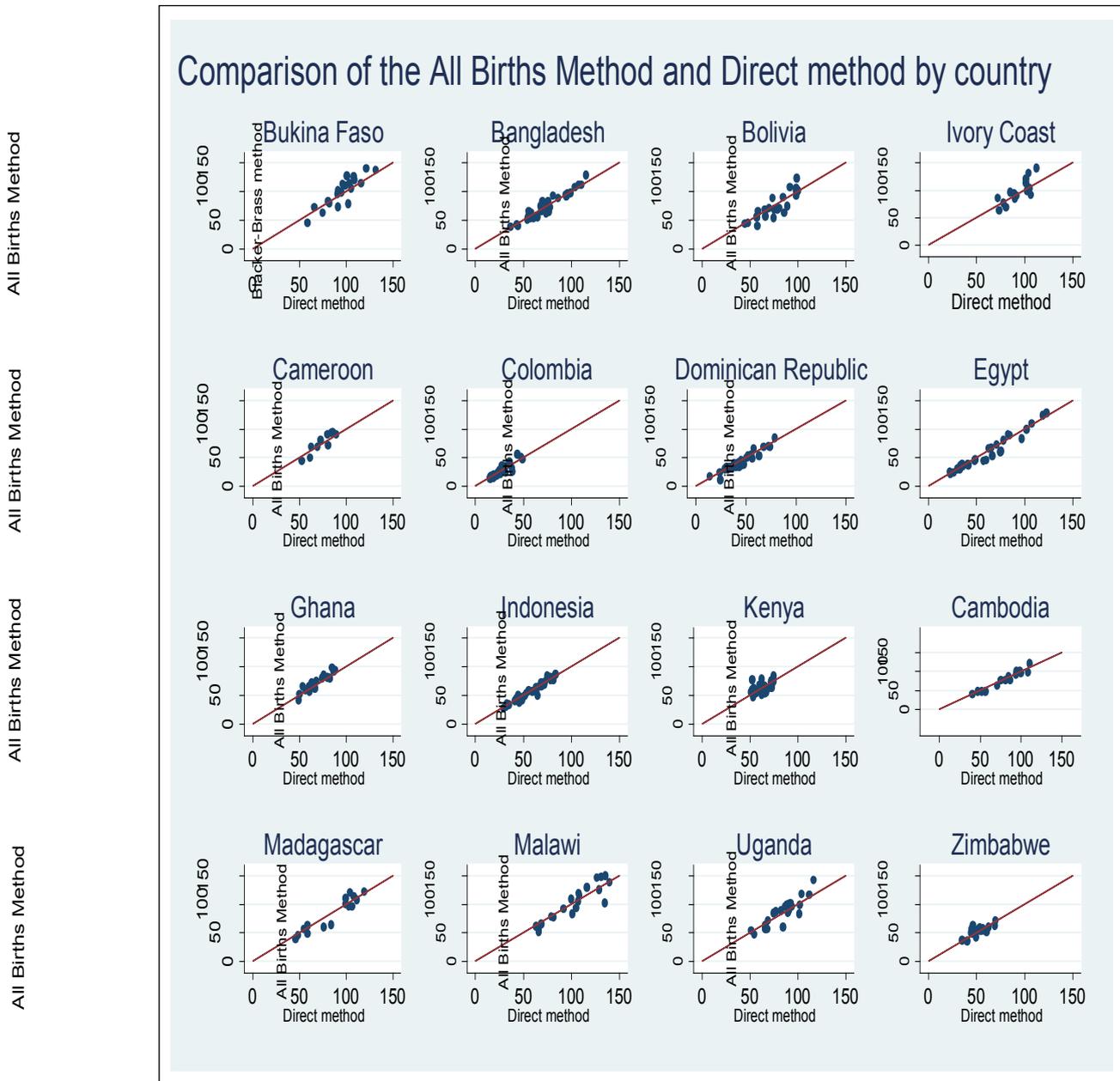


Figure 4.4: Comparison of the Blacker-Brass and direct method by country, DHS data

The results presented so far have shown some differences in the two methods across country and with time. However, these results do not ascertain whether these are statistically significant. A paired *t*-test was performed to check whether the differences between the ABM and direct estimates are statistically significant and results are shown in Appendix C.

The data points with calculated infant mortality estimates decreased from 405 to 375. This was because 30 data points could not be paired since either one of the methods could not produce an estimate for these data points due to missing data and failure of other variables to meet specific criteria required in the method.

The p -value of 0.5036 and confidence interval which includes a zero implies that there is no statistical difference in the estimates produced by the two methods. Therefore there is no evidence to reject the null hypothesis that the two methods produce similar estimates.

4.2 Factors associated with the differences in the infant mortality measures

To further investigate the difference between the two methods, it was hypothesised that the relative errors may be associated with the country where the estimates are obtained, the level of mortality and the time before the date of the survey. Results of these investigations are presented below.

4.2.1 Significant differences in the relative errors by country

A linear regression model of the country variable against the relative error variable was done taking Bangladesh (arbitrarily) as the base category. Here the null hypothesis is that there is no significant difference in the relative errors across different countries and the alternative hypothesis is that the relative errors differ significantly by country. Table 4.1 presents results of the test. Out of 375 observations a very weak correlation coefficient of 2 per cent was obtained for the test. A p -value of 0.093 led to the conclusion that there is no evidence to suggest a relationship between the country variable and the relative errors. All the p -values and confidence intervals across all the countries have results pointing to insignificance of the test. As such, there is little evidence to show that the relative errors differ from the base country – Bangladesh.

Table 4.1: Significance difference Test in relative errors by country

<i>Country</i>	<i>Coefficient</i>	<i>P – value</i>
Bangladesh (Base)		
Bukina Faso	0.030	0.389
Bolivia	-0.033	0.308
Ivory Coast	0.038	0.31
Cameroon	0.004	0.927
Colombia	-0.020	0.515
Dominican Republic	-0.055	0.077
Egypt	-0.035	0.263
Ghana	0.050	0.149
Indonesia	-0.028	0.359
Kenya	0.007	0.831
Cambodia	-0.038	0.315
Madagascar	-0.039	0.308
Malawi	-0.016	0.652
Uganda	0.004	0.905
Zimbabwe	0.031	0.343
Constant	-0.003	0.887

4.2.2 Significance difference test in relative errors by mortality level

Mortality estimates are usually presented at country level. However, different countries may have the same level of mortality. It is therefore hypothesised that the relative errors between the two methods may be associated with the levels of infant mortality. There were 375 observations for this test and an R-squared value of about 16 per cent. A test of statistically significant difference was done to check if the relative errors are associated with the levels of mortality. Five categories of mortality levels were defined. The base category used had deaths between 0 – 24 per 1000 live births. Table 4.2 presents results of the statistical tests performed. At 95 per cent confidence level, there is a significant difference as shown by the *p*-values which are less than 0.05 in the relative errors of the base category with all the other categories except the 25-49 intervals. This shows that with increasing level of mortality, the relative errors of the estimates in the two methods changes significantly as also suggested by the increasing coefficients. However, this result could also be affected by the data quality and the bias present due to HIV/AIDS.

Table 4.2: Significance difference test in relative errors by level of mortality

<i>Mortality level (deaths per 1000 live births)</i>	<i>Coefficient</i>	<i>Standard Error</i>	<i>T-test</i>	<i>P - value</i>	<i>Confidence Interval (95%)</i>	
0-24 (base)						
25 – 49	0.028	0.024	1.18	0.237	-0.019	0.075
50 – 74	0.062	0.023	2.64	0.009	0.016	0.108
75 – 99	0.106	0.025	4.29	<0.001	0.058	0.155
/>100	0.184	0.028	6.57	<0.001	0.129	0.239
Constant	-0.081	0.021	-3.79	<0.001	-0.123	-0.039

4.2.3 Significant differences in the relative errors by time before survey date

A further statistical test to check whether the relative errors differ significantly by time before the survey year was performed. Five 2-year periods were defined as described in Chapter 3 and on the assumption that deaths are uniformly distributed across each two year period; a time reference for the mortality estimates was taken to be the mid-point of that two year period. Significant difference tests were performed between this time period and relative error and results are in Table 4.3. A total of 375 observations were used for this test with a correlation coefficient of about 28 per cent. The p -values are all less than 0.001 in the four categories and the confidence interval do not include zero. This gives evidence to reject the null hypothesis and conclude that there is a significant difference in the relative errors of estimates in the 0-23 month period and those for the other 2 year periods. The constant term pertains to relative errors for estimates at the time of the survey. The regression coefficient of the constant term = -0.128 indicates that the relative errors are reduced moving back further in time before the year of the survey, however, the time variable coefficients are positive for each period. The standard errors for both the time variable and the constant terms are quite low, almost close to zero for the time variable.

Table 4.3: Significant differences in the relative errors by time periods before survey date

<i>Time before survey date (months)</i>	<i>Coefficient</i>	<i>Standard Error</i>	<i>T-test</i>	<i>P-value</i>	<i>Confidence Interval (95%)</i>	
0-23 (base)						
24 – 47	0.17	0.02	10.04	0.00	0.134	0.200
48 – 71	0.09	0.02	5.54	0.00	0.060	0.125
72 – 95	0.15	0.02	8.78	0.00	0.114	0.179
96 – 120	0.18	0.02	10.66	0.00	0.145	0.210
Constant	-0.13	0.01	-10.88	0.00	-0.151	-0.105

4.3 Direct estimation and Most Recent Births Method results

One of the key objectives of this research is investigating the bias of the Blacker-Brass method using census data. In a census the date of birth of the last born child and its survival status on the census date are used to calculate infant mortality. In this case infant mortality rates are calculated using last births only in each of the five 24 month periods defined earlier. This procedure has been named the MRBM in section 1.3. Results of this investigation are presented in Figure 4.5 below.

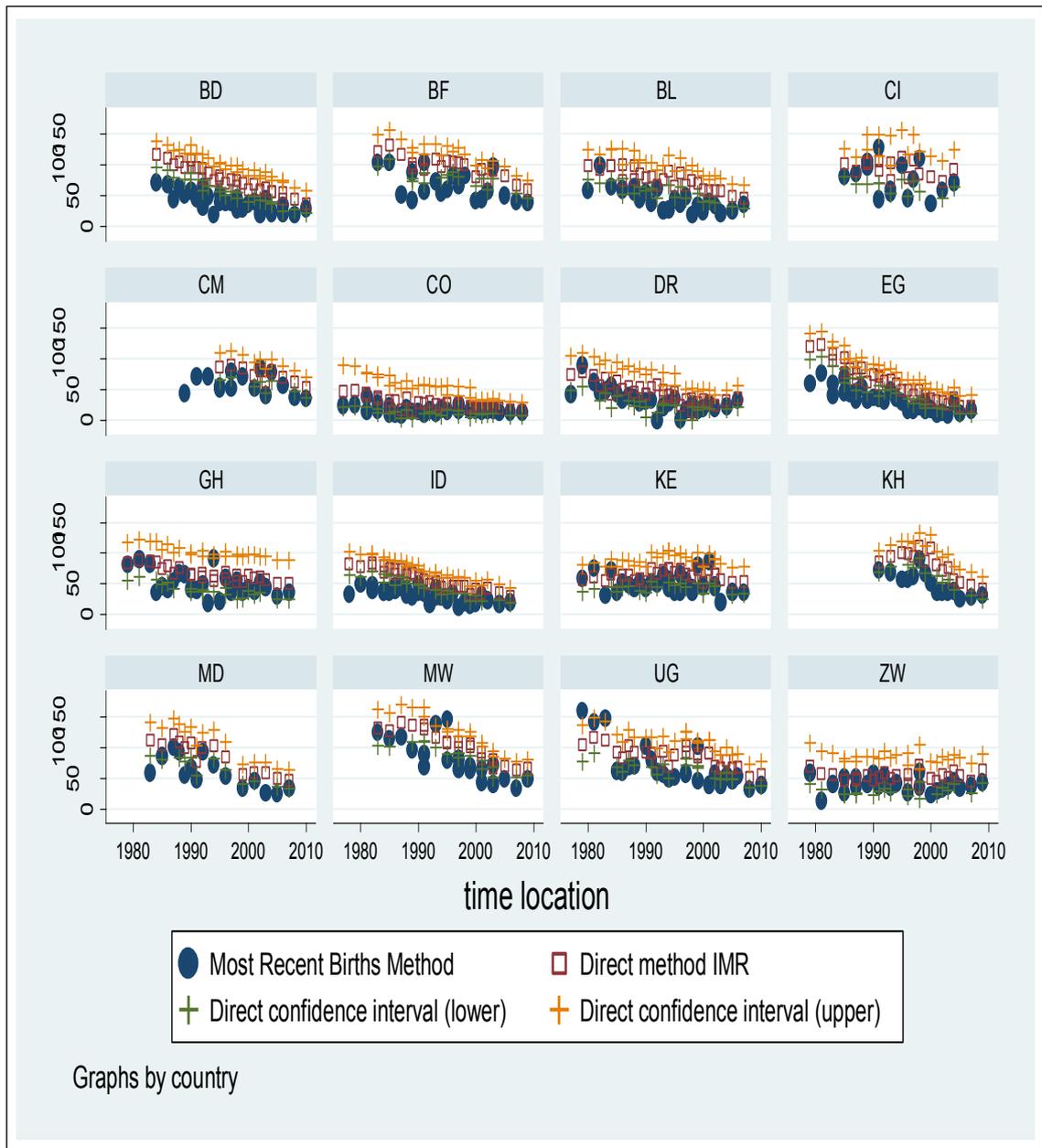


Figure 4.5: Trends of infant mortality using the direct estimates and Blacker-Brass method (using last births) by country

¹ Confidence intervals were also plotted as scatter points instead of connected lines since the data points overlap due to multiple surveys in one country.

The results from Figure 4.5 show that MRBM estimates underestimate mortality. Most of the estimates derived using the method are below the lower bound of the confidence interval of the direct estimates. In addition, no MRBM estimates are above the upper bound of the confidence interval of the direct estimates.

Comparing this result with the ABM estimates presented in section 4.1, it is observed that all estimates of the ABM were within the 95 per cent confidence interval. An average relative error was calculated for the MRBM estimates and the direct estimates in the 24-month period yielding a value of -32.3 per cent. Similarly, an average relative error was also calculated using the ABM estimates in the 24-month period and the value was only -1.2 per cent. The negative values imply that both methods underestimate infant mortality relative to the direct estimation method. However it is evident from the two values that using last births only in estimating infant mortality results in severe underestimation.

4.4 Derivation of correction factors to the All Births Method estimates

Chapter 2 noted that Blacker and Brass proposed a conversion factor of 1.09 to convert the proportion of dead children into ${}_1q_0$. This procedure was used to obtain ABM estimates that were compared with the direct method estimates. Various factors were tested for association with the relative errors of the ABM method to the direct method. The level of mortality factor showed the most significant difference in the relative errors compared to the other factors as illustrated in sub-section 4.2.2. In addition, investigations carried out in sections 4.1 and 4.3 revealed that the ABM estimates are more accurate than the MRBM estimates relative to the direct estimates.

In the current section, a linear regression model is developed using different mortality levels based on the total number of reported deaths among all births in the 24-month period against the direct method estimates. The factors used to convert reported deaths into an infant mortality measure are calculated based on the linear regression model. Adjusted infant mortality estimates are calculated by multiplying the conversion factors by the proportion of dead children reported in the 24-month periods. These estimates were compared with ABM estimates.

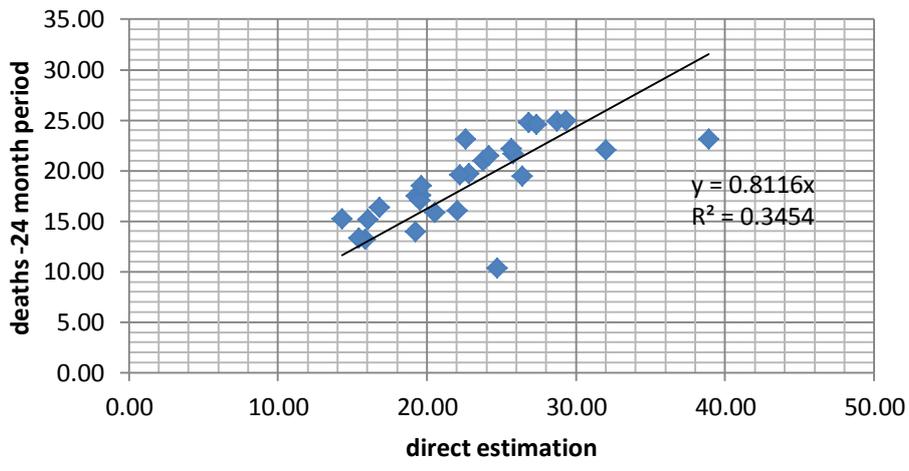
The procedure of deriving the conversion factors is described below. Reported deaths in the past 24 months were plotted against the direct method estimates and results presented in Figure 4.6 below.

The intercepts were set to zero so that the gradient only is used to define the relationship between the two variables and to ascertain that, the straight line should pass through the origin. For reported deaths between 0 – 24 per 1000 live births, the coefficient of correlation was weak at about 34.5 per cent. This weak correlation was caused mainly by one particular outlier with an IMR of 25 per 1000 live births as measured by the direct method and only 10 deaths per 1000 live births in the past 24 months for the same number of births. A unit increase in the direct method estimate was associated with about a 0.81 increase in the number of reported deaths in the 24 month period.

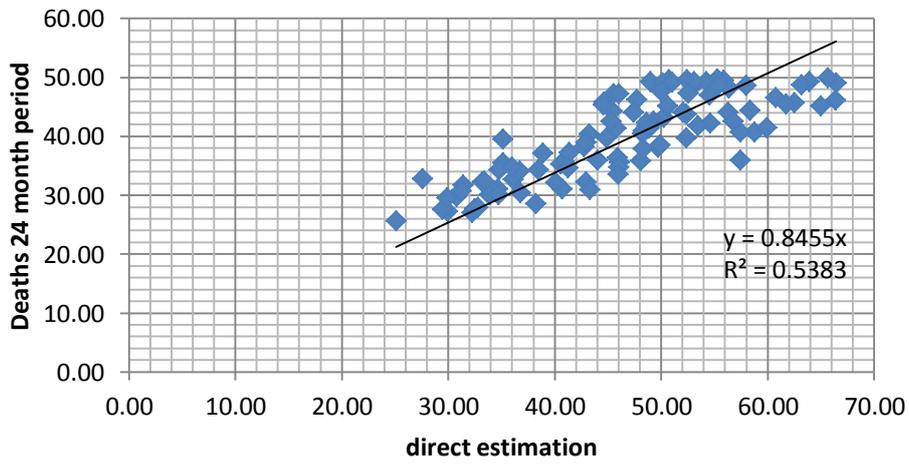
In the category with 25 – 49 reported deaths per 1000 live births, the correlation coefficient had increased to 53.8 per cent indicating a closer relationship of the reported deaths in the past 24 months and the IMR as measured by the direct method. The per-unit increase had also increased slightly to 0.85. For reported deaths in the 50 - 74 deaths per 1000 live births range, the correlation coefficient fell drastically to 10.5 per cent indicating a very weak negative relationship of reported deaths in this category with the estimates from the direct method. This weak correlation was caused by a couple of outliers as shown in the figure below. However, one unit increase in the direct estimate was associated with 0.87 decrease in the number of reported deaths in the 24-month period reported.

As the number of reported deaths increased to 75 – 99 deaths per 1000 live births the correlation coefficient also improved to 58 per cent and so did the per-unit increase of the deaths reported with increasing IMR. However there is just one outlier in this category that reduced the correlation. This data point has an infant mortality rate of about 135 deaths per 1000 live births and only about 94 reported deaths. The category with highest number of reported deaths of at least 100 deaths per 1000 live births had a weaker correlation coefficient of 43.2 per cent however the per-unit rate of increase was on average almost equal to the 75-99 category.

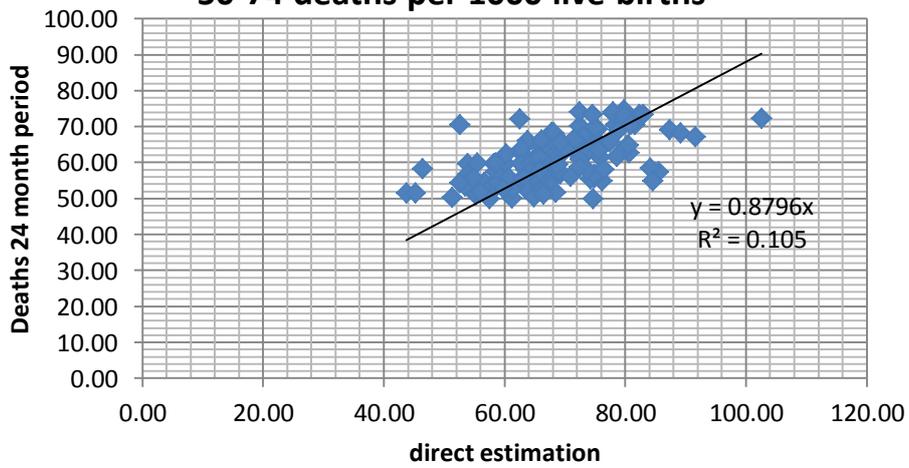
0-24 deaths per 1000 live births



25-49 deaths per 1000 live births



50-74 deaths per 1000 live births



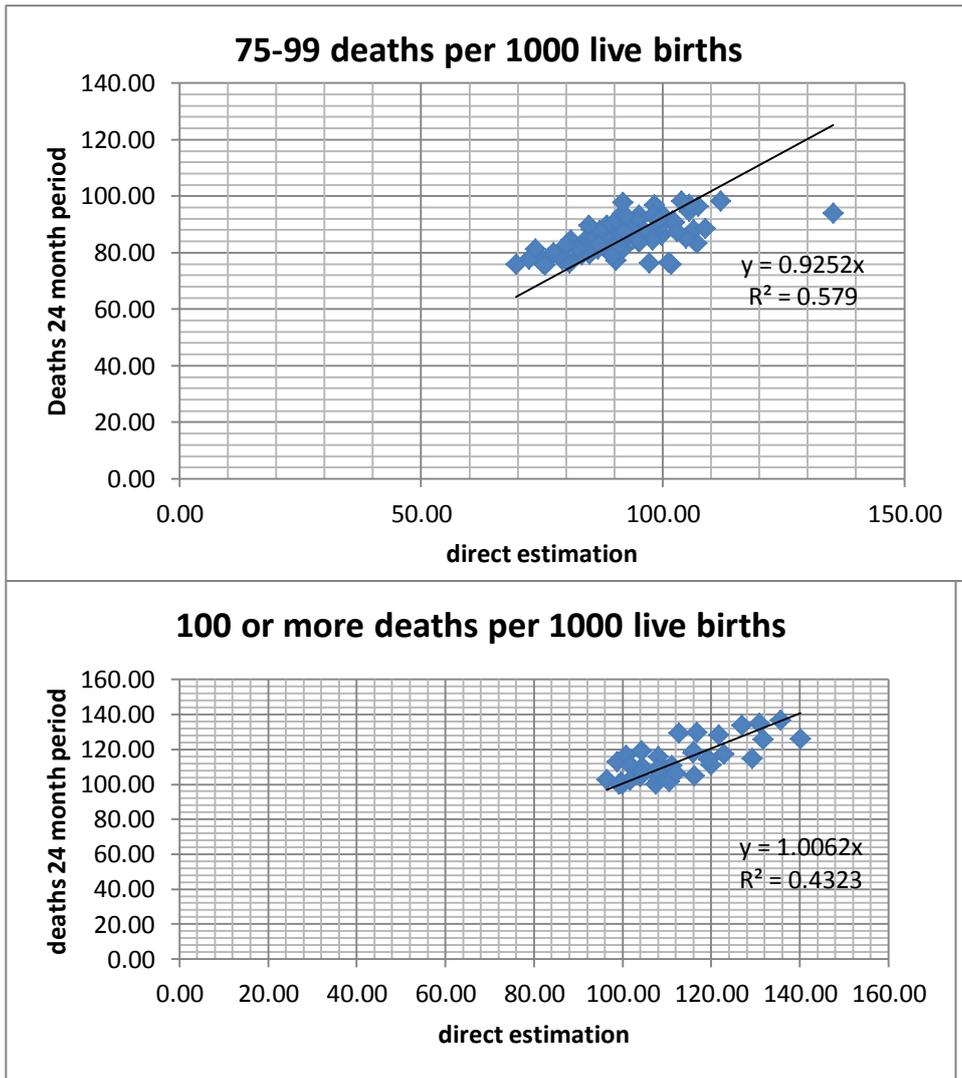


Figure 4.6: Relationship between reported deaths in the past 24 months and infant mortality from direct estimation

The following step in deriving the conversion factors involved calculation of median values of the reported deaths per 1000 live births in each of the five categories of mortality levels defined earlier. These values were plotted against the regression coefficients discussed above and presented in Figure 4.6. A linear model was fitted with $\alpha = 0.7624$ and $\beta = 0.0021$. This model was used to convert the reported deaths into a measure of mortality using the formula;

$$d^* = \frac{d}{d \cdot \beta + \alpha}, \text{ where } d^* \text{ is the adjusted deaths equivalent to the probability of}$$

dying before the age of one (${}_1q_0$), d is the number of reported deaths in the past 24 months from all births reported in the same period.

The adjusted deaths were plotted against the direct method estimates and the results are presented in Figure 4.7 below. There is a strong positive correlation of 91.6 per cent between the estimates indicating that the new mortality estimates are highly accurate.

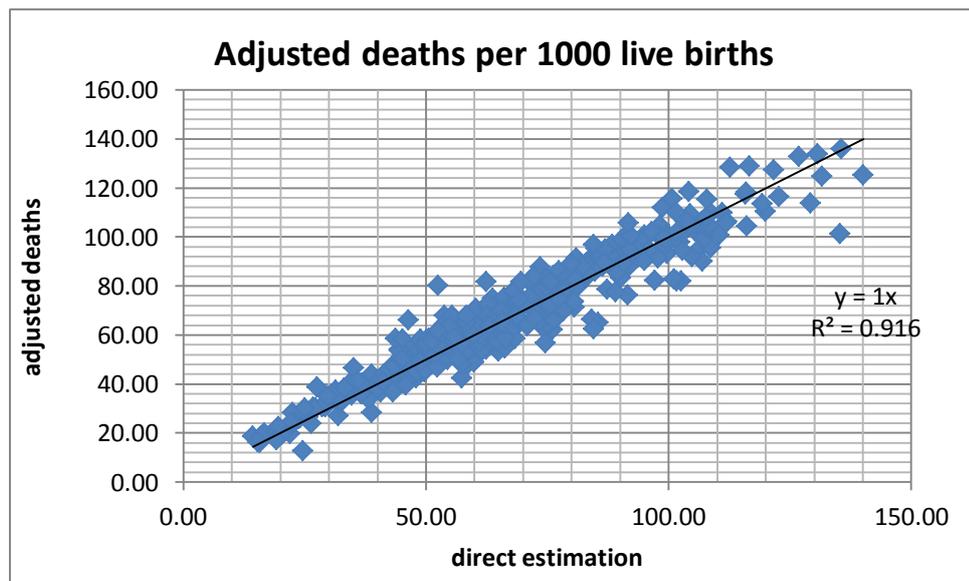


Figure 4.7: Relationship between direct method estimates and mortality rates derived from adjusted reported deaths per 1000 live births

To compare the accuracy of the new mortality estimates – adjusted deaths with the ABM estimates, an error sum of squares (SSE) was calculated between the adjusted deaths estimates and the direct method estimates across all mortality levels as follows;

$$SSE = \sum_{i=1}^{375} (d_i - d_i^*)^2 \text{ where; for each } i^{\text{th}} \text{ observation, } d_i \text{ is the direct method}$$

estimate and d_i^* the adjusted reported deaths per 1000 live births.

A SSE of 19,961 was obtained. A similar value of the SSE was calculated using the ABM estimates and an SSE of 26, 915 resulted. The ABM therefore produced a higher SSE compared to the adjusted death estimates derived from adjusting reported deaths by a relative error of 34.8 per cent derived from all mortality categories.

Results of the correction factors derived to convert the reported deaths into adjusted deaths are presented in Table 4.4. It can be noted that the correction factors are positively correlated with mortality level.

Table 4.4: Correction factors for ABM estimates for each mortality level

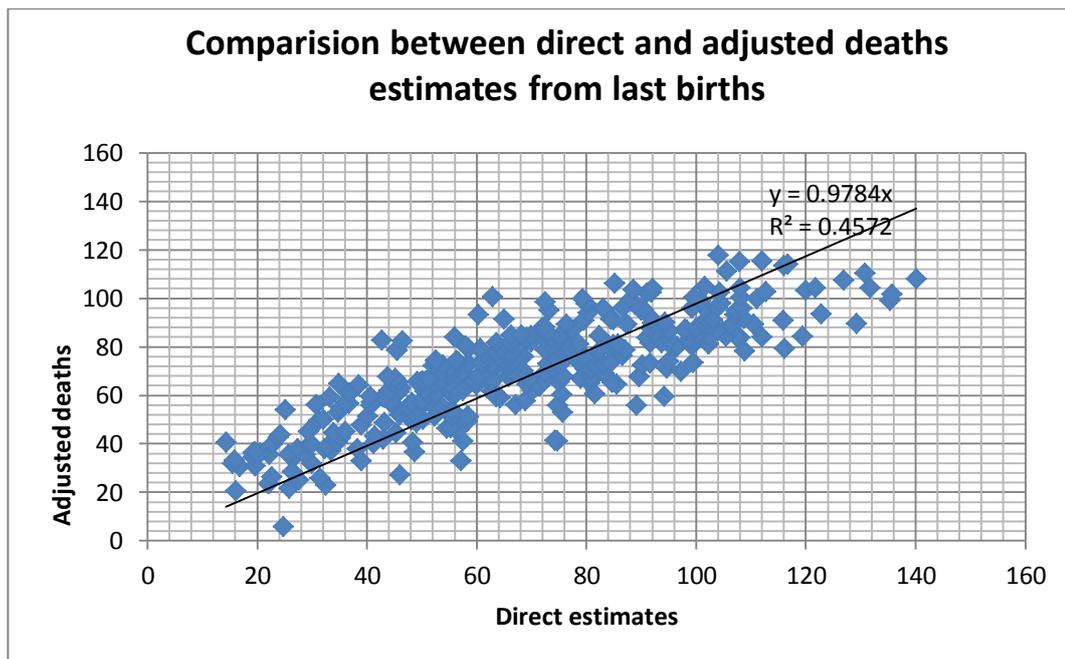
<i>Number of reported deaths in 24 month period per 1000 live births</i>	<i>Correction factor</i>
0-24	0.8116
25-49	0.8455
50-74	0.8796
75-99	0.9252
100>	1.0062

4.5 Derivation of correction factors to the Most Recent Births Method estimates

As noted in sub-section 2.4.2, the Blacker-Brass method has been applied to census data where survival status of last births only are used to derive infant mortality after multiplying the proportion dead by a correction factor of 1.09. DHS data on last births only in the 24-month period were also extracted and the correction factors were applied on these reported deaths to derive probability of dying before the age of one. Similar to the procedure described in section 4.4, a linear regression model was developed using different mortality levels based on the total number of reported deaths among last births only in the 24-month period. These estimates were plotted against the direct method estimates. The factors used to convert these reported last births into an infant mortality measure are calculated by multiplying the conversion factors by the proportion dead among the last births. This procedure is exactly similar to that described in the preceding section.

Plots of the direct method estimates and the reported deaths among the last births only in the 24-month period were plotted across the five different mortality levels as defined in the previous section. The intercepts of the linear regression models were set to zero to ascertain that the relationship between the direct method estimates and

the reported deaths is defined by the gradient only. Median values of each of the five mortality level categories were then plotted against their respective gradient coefficients and a linear model with $\alpha = 0.3175$ and $\beta = 0.0065$ was yielded. The reported deaths from these last births were adjusted using the formula described in section 4.4 to get adjusted deaths. The direct method estimates were plotted against the adjusted deaths and the results are shown in below. The R-squared value of 45.7 per cent indicates that even after correction there is still a relatively weak relation between the direct estimates and the adjusted deaths derived from deaths from the most recent births.



An SSE of 99,302 resulted from comparing the direct method and reported deaths from the most recent births and an SSE of 84,154 was obtained from the direct method and adjusted deaths. The SSE of the two sets of results were compared and the adjusted deaths estimates (method using correction factors) reduced the error by 18% compared to the MRBM using last births only. Results of the correction factors used to convert reported deaths into adjusted deaths are presented in Table 4.5 below.

Table 4.5: Correction factors for MRBM estimates for each mortality level

<i>Number of reported deaths in 24 month period per 1000 live births</i>	<i>Correction factor</i>
0-24	0.431
25-49	0.582
50-74	0.6707
75-99	0.863
100>	1.0662

4.6 Out-of-sample validation

An out-of-sample validation of the adjustment factors was performed to check for consistency of the results obtained in comparing the errors of the direct method estimates and the adjusted deaths estimates obtained from deaths in a 24 month- period. Twelve DHS datasets from four countries namely, Guatemala, Namibia, Turkey and Morocco were used. Results of comparing the ABM estimates and adjusted deaths estimates with the direct estimates are presented below in Figure 4.8 and Figure 4.9 respectively. The R-squared values of the adjusted deaths and the ABM estimates against the direct estimates are similar at about 84 per cent indicating a strong positive correlation.

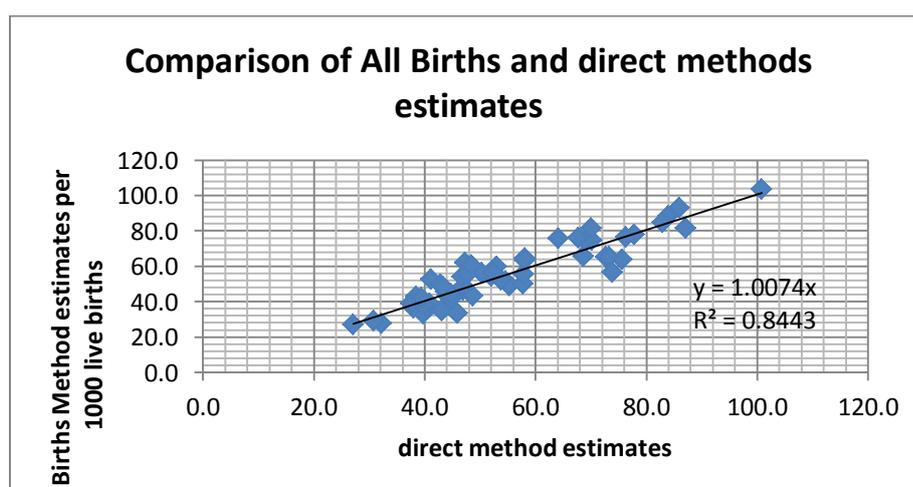


Figure 4.8: Comparison of the direct estimates and adjusted deaths in the past 24 months

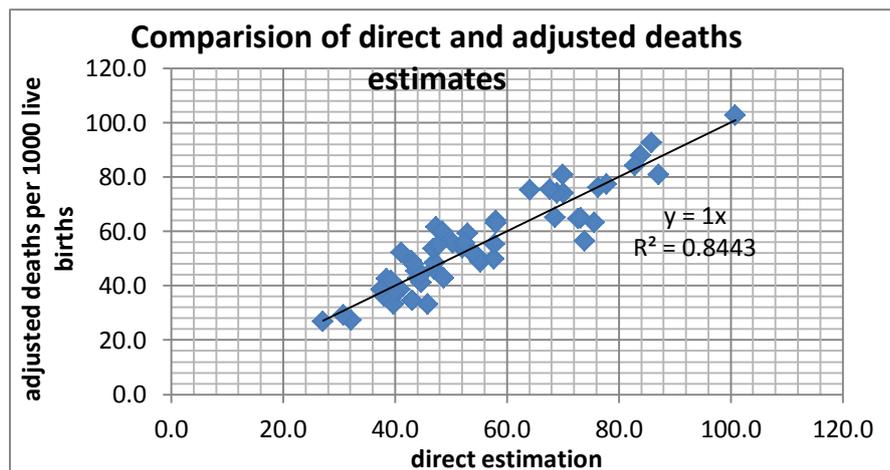


Figure 4.9: Comparison of the Blacker-Brass and direct estimates

The error sum of squares (SSE) was also calculated to check for consistency with results of the sample data. The ABM results yielded an SSE of 5,414.81 whilst the adjusted deaths estimates have an SSE of 4887.00. Hence, the relative error of the SSE between the ABM estimates to the adjusted deaths method is 10.8 per cent. This result has also showed an error reduction in the adjusted deaths results relative to the ABM results - a similar pattern with the sample results.

4.7 Discussion

In this section, the results of the mortality estimates calculated using the direct estimation method and the ABM and MRBM are discussed. Further, results of the factors associated with the relative errors between the ABM and direct method are also explained. The nature of bias involved in applying ABM and MRBM is also discussed and lastly the consistency and plausibility of the new procedure used to convert reported deaths in the past 24 months into a measure of mortality is presented.

As noted earlier, there has been a decline in infant mortality over a period of about 30 years between 1980 and 2010 in most countries. However results of the data quality assessment showed evidence of birth transferences in some datasets which might have resulted in over estimation of mortality in the more distance past. Hence mortality decline might have been slightly over-estimated in these datasets. The World Health Organisation (2013) assessed the IMR trends between 1990 and 2007 for developing countries in Africa, America and Asia using DHS datasets.

In their estimates, there was a notable decrease in mortality with time and this is consistent with the findings of this dissertation. The United Nations (2013) also provide infant mortality projections for all countries worldwide. Between 1980 and 2010 the estimates dropped from 111 to 73 deaths per 1000 live births for all countries in Africa combined. The estimates in Figure 4.1 were within the same range and even more consistent with the UN estimates at country level.

Measure DHS produces country specific reports with IMRs for children born within the past five years of the survey date. These rates are calculated using the direct estimation method. The mortality estimates obtained in this study were for children born in two year periods for the past ten years before the survey date. These estimates were compared and most of the estimates obtained by the direct estimation method were in the 95 per cent confidence intervals of the estimates published by Measure DHS. There is therefore evidence to show that the direct estimates used in the dissertation are quite plausible and generally consistent with other external sources.

The key research question for this project was to check the consistency of recent births with the direct estimation method. To answer this research question, three sets of infant mortality estimates were obtained. First, ABM estimates using all births in the 24 month period were calculated.

Second, Blacker-Brass estimates were also derived from last births only in the 24-month period (MRBM estimates) and third, direct infant mortality rates were derived using the direct estimation method. Results of the relative errors were used to characterise the differences in estimates of the two methods.

It was not clear whether the ABM over or underestimates infant mortality relative to the direct method. Further investigations of the relative errors showed that the ABM method was underestimating mortality in the first 24-month period before the survey date in most datasets. The relative errors also differed significantly across different mortality levels. There was a statistically significant positive correlation between the relative errors and the mortality level giving evidence to show that the ABM and MRBM both increase bias with increasing mortality levels. There were no significant differences in the relative errors across countries hence the accuracy of the method is not affected by application in different countries.

Results presented in section 4.3 implied that applying the Blacker-Brass method to census data which involves the use of last births and vital status of these births – MRBM results in under-estimation of IMR more severely compared to the ABM.

The MRBM produced an average relative error of -32.3 per cent for each data point across all countries and surveys. This value was much higher compared to the average relative error of just -1.2 per cent calculated from ABM estimates suggesting that using last births only (the Blacker-Brass procedure) yields more error in the estimates when compared to the direct estimates.

A simple linear regression model was formulated using all reported births in the past 24 months and the direct method estimates considering the effect of mortality level categories as described in section 4.3 to obtain different regression coefficients for each of the five mortality level categories. These were used to convert deaths reported in the 24 month period into the probability of dying before the age of one.

The new mortality rates produced a much smaller SSE by almost 35 per cent compared to the SSE of the ABM estimates when compared to the direct method estimates. Accordingly the use of all births and deaths occurring in a 24 month period appeared to be more consistent with the direct estimation method after adjusting for different mortality levels in conversion of these deaths to an infant mortality measure.

Correction factors were also made on the Blacker-Brass method which uses last births only. The estimates were improved by 18 per cent compared to the original Blacker-Brass method using last births only; however this improvement was less than that yielded from adjusting all births in the 24-month period.

5. CONCLUSION

5.1 Introduction

The main objective of this research study was to assess the accuracy of recent births in a 24-month period in measuring infant mortality. Two scenarios were considered, one which used all births in the past 24 months (ABM) and the other which used last births only – the original Blacker-Brass method (MRBM) to calculate the probability of dying before the age of one. These estimates were both compared to the direct method estimates which were used as the gold standard for infant mortality measures. The assessment was extended to births which occurred in other four 24-month periods before the survey date. As such, infant mortality rates were derived from births which occurred 10 years before the survey.

In this chapter, a conclusion on the quality of the data used to derive the mortality rates is presented in section 5.2; following that, concluding remarks on the accuracy of the Blacker-Brass estimates using last births and all births in the 24-month period and also the effect of time moving further back from survey date are presented in section 5.3. In the same section, conclusions are drawn on the effect of different mortality levels on the accuracy of the estimates. Conclusions on the effect of correction factors derived from adjusting for mortality level on improving the mortality estimates are done in section 5.4. The limitations encountered in this dissertation are presented in section 5.5 and lastly the chapter is concluded by recommendations based on the findings and on areas of possible further research in section 5.6.

5.2 Data quality

Various issues were considered in examining the quality of data used to obtain infant mortality estimates. Data quality aspects examined were on, missing data, number of births reported in the 24-month periods, and finally SRAB and SRAD. Since infant mortality was calculated using reported births and deaths by country and survey separately, for the ABM, MRBM and direct estimation, each data point used the same sample sizes hence the two estimates were comparable. On missing data, no attempts were made to fill in any data.

Since data were extracted from Measure DHS it was assumed that missing data had trivial implications in calculating the mortality measures. Evidence from examination of the sex ratios showed that infant mortality may have been slightly over estimated with increasing time before the survey date.

Seven DHS datasets had this systematic error however it was not unreasonable. There was also evidence of a slight over estimation of male deaths for children born within the first 24 months before the survey across all countries. However, no corrections were made for this error since again the error did not affect comparison of the mortality estimates in the two methods.

5.3 Estimates derived from recent births

The results showed that the ABM estimates were within the approximate 95 per cent confidence intervals of the direct estimates. However, the MRBM (Blacker-Brass) estimates derived from last births only were more biased as most of the estimates were below the approximate 95 per cent confidence intervals of the direct method estimates indicating a higher degree of mortality underestimation in comparison with the ABM estimates. Therefore the ABM estimates were considered for further investigations and some systematic differences between these estimates and the direct method estimates were observed. It was shown that the ABM slightly over-estimates infant mortality rates as we move further back from the survey date. Particularly, the first 24-month estimates significantly differed from the other four 24-month periods. Recall bias of the mothers could be the reason for this result. This might suggest that applying the ABM to data for periods greater than 24 months prior to the survey date may yield more bias. Further, the results revealed a significant difference in the relative errors by mortality level. There was a positive correlation between the mortality level and the relative errors of the ABM and direct estimation method indicating that the ABM yields more error with increasing mortality level relative to the direct estimation method. This result prompted the need to adjust for the effect of different mortality levels in converting the reported deaths into a probability of dying before the age of one.

5.4 Adjusting for mortality level – correction factors

The positive correlation between the ABM estimates relative errors with mortality level informed the next step of the dissertation of attempting to reduce the bias by using different adjustment factors for each level of mortality in converting reported births in the 24 month period into a mortality measure. Different correction factors were proposed as presented in Table 4.4 for each mortality level category. The new procedure for deriving new mortality rates was applied on reported births and deaths in the 24-month period and the SSE was reduced by 35% in comparison with the original ABM estimates.

The new method was also applied on out of sample data using 12 DHS datasets from Guatemala, Namibia, Turkey and Morocco and the SSE was also reduced by about 11% in comparison with the ABM applied to the same out of sample datasets.

In addition, the new procedure for deriving mortality rates was applied on recent births, i.e. last births only in the 24 month period (MRBM). This is a similar scenario with census data. The SSE of the new estimates and the direct method was also reduced by 18% compared to the original Blacker-Brass method using last births only with a conversion factor of 1.09 constant across all mortality levels.

5.5 Limitations of the study

The direct estimation method which was used as the basis for assessment of the methods using recent births and deaths has its shortcomings. One of the assumptions of the method is that the month and year of birth and deaths are reported accurately. This assumption may have been violated especially on month of deaths since data used were for a period up to 10 years before the survey risking recall bias on vital events. Another assumption used was that of no correlation between the mother and the child's mortality. Most of the datasets used in this dissertation are from countries with high HIV prevalence hence this assumption is violated. Even though the mortality estimates used for the investigations were validated using other external sources, an ideal examination of the robustness of recent births in measuring infant mortality could have been made possible if correction for the bias of HIV/AIDS was done to the direct method estimates. Using a more accurate infant mortality measure as the basis for deriving conversion factors from the ABM estimates would yield better results. Alternatively using IMRs derived from complete and reliable vital registration data could better approximate the bias of the Blacker-Brass method and also improve the conversion factors.

Further, DHS datasets were used to assess the Blacker-Brass method yet the original method used census datasets. Considering that census datasets use summary birth histories usually reported by a head of household during a census and DHS questions are full birth histories usually reported by women aged 15-49, the difference in the data can produce substantial bias. However, in this research DHS datasets were set up in a census format i.e. to apply the Blacker-Brass method on last births only reported in a 24-month period.

In addition, as noted in the results section, the relative errors of results pertaining to the two-year period before the survey were significantly different to the other four-two year periods pointing to the possibility of recall bias. This is also a significant limitation to the research.

The linear regression model uses a couple of assumptions that may not be an accurate representation of the relationship between the ABM and MRMB estimates and the direct method estimates. In this case the relationship between the estimates is not strictly linear and additive as the model suggests. Hence there was some loss of precision in defining the relationship. Further the errors resulting in assuming a linear relationship may also be not normally distributed.

5.6 Areas for further research

The main critique of the Blacker-Brass method, Hill's main reason for objecting the reliability of the method was that there was a significant difference in the proportion dead between all births in the 24-month period and the most recent births. The investigations done in this dissertation confirmed Hill's argument because infant mortality was severely underestimated when using last births only. The new procedure which adjusted for different mortality levels reduced the SSE in three different scenarios of, the ABM estimates i.e. all births reported in the 24-month period in the initial sample, applying the same method on out of sample data and lastly applying the Blacker-Brass method using recent births only in the 24-month period. The results concluded that the most robust approach of estimating infant mortality relative to the direct estimation method is the ABM method adjusted for different mortality levels.

It is therefore recommended that census questions should include questions on all births in the last 24 months and their survival status to reduce the bias of omitting possible births and deaths required for applying the Blacker-Brass method. The new procedure presented here can be more adequately tested on different census datasets to get more evidence on how much the correction factors improves the accuracy of the mortality estimates using recent births. In addition, since it was shown in this dissertation that, relative errors between the Blacker-Brass and direct estimates were reduced moving further back from survey date, further research can be done to adjust for this bias as was proposed for different mortality levels.

One of the major limitations of the research was also on correction for HIV/AIDS bias. Further improvement of the correction factors can be done by deriving them after correcting for the bias of HIV/AIDS in the direct method estimates. This gives a relatively more accurate infant mortality estimate derived from recent births.

REFERENCES

- Adetunji, J. 1996. "Infant mortality levels in Africa: Does method of estimation matters?", *Genus* **52**(10):89-106.
- Aguirre, A and Hill, A.G. 1988. "Estimating child mortality using the previous child technique with data from health centres and household surveys: Methodological aspects", *Notas Poblacion* **16**(46-47):9-39.
- Artzrouni, M and Zaba, B. 2003. "HIV-induced bias in the estimation of child mortality using birth history reports," Paper presented at The HIV and AIDS Child Mortality conference. London, March 2003.
- Bicego, G, Augustin, A, Musgrave, S, Allman, J and Kelly, P. 1989. "Evaluation of a simplified method for estimation of early childhood mortality in small populations", *International Journal of Epidemiology* **18**(4)
- Blacker, J and Brass, W. 2005. "The Estimation of Infant Mortality From Proportions Dying Among Births in the Last 24 Months", *Southern African Journal of Demography* **10**(1&2):25-42.
- Brass, W, Coale, A.J, Demeny, P and Heisel, D.F. 1968. "Methods of Analysis and Estimation," in *The Demography of Tropical Africa*. Princeton University Press, pp. 88-139.
- Brass, W and Macrae, S. 1984. "Childhood mortality estimated from reports on previous births given by mothers at the time of a maternity: I. Preceding-births technique", *Asian and Pacific Census Forum* **11**(2):5-8.
- Brass, W. 1969. "Disciplining demographic data," Paper presented at International Union for the Scientific Study of Population. London. 183-204.
- Chalapati, R., Bradshaw, D. and Mathers, C. . 2003. "Improving death registration and statistics in developing countries", *Southern African Journal of Demography* **9**(2):81-99.
- Coale, A.J and Trussell, J (1977) Population Bulletin of the United Nations.
- Cutler, D.M, Angus, S.D and Adriana, L.M. 2006. "The Determinants of Mortality", *Journal of Economic Perspectives* **20**(3):97-120.
- Darikwa, T.B and Dorrington, R. 2013. "The level and trends of Child mortality in South Africa",
- Ewbank, D.C. 1982. "The sources of error in Brass's method for estimating child survival: The case of Bangladesh", *Population Studies* **36**(2):459-474.
- Fernando, C. 1985. "Indirect estimation of infant mortality trends: simulation tests and the Feeney method", *Genus* **41**:65-68.
- Garenne, M and Gakusi, E (2006) In *Bulletin of the World Health Organization*, Vol. 84.
- Hill, A.G and Aguirre, A. 1990. "Childhood Mortality Estimates Using the Preceding Birth Technique: Some Applications and Extensions", *Population Studies* **44**(2):317-340.
- Hill, K. 2012a. "Effects of HIV on methods of child mortality estimation," in Moultrie, TA, RE Dorrington, AG Hill, K Hill, IM Timæus and Zaba, B (eds). *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population,
- Hill, K. 2012b. "Indirect estimation of child mortality," in Moultrie, TA, Dorrington, RE, Hill, AG , Hill, KH , Timæus, IM and Zaba, B (eds). *Tools for Demographic Estimation*. Vol. 2013. Paris: International Union for the Scientific Study of Population,
- Hill, K. 2013. "Direct estimation of child mortality from birth histories," in Moultrie, TA, RE, Dorrington, AG, Hill, K, Hill, IM, Timæus and B, Zaba (eds). *Tools for*

- demographic estimation*. Paris: International Union for the Scientific Study of Population,
- IGME. 2013. *Level & Trends in Child Mortality*. World Development Indicators: World Bank. <http://data.worldbank.org/indicator/SP.DYN.IMRT.IN>. Accessed: 25 July 2013.
- IHME. 2013. <http://www.healthmetricsandevaluation.org/gbd/visualizations/gbd-arrow-diagram>. Accessed: 10/10 2013.
- Keyfitz, N. 1966. "Finite Approximations in Demography", *Population Studies* **19**(3):281-295.
- Madi, HH. 2000. "Infant and child mortality rates among Palestinian refugee populations", *Lancet* **356**(9226):312.
- Mahy, M. 2003. "Measuring Child Mortality in AIDS Affected Countries," Paper presented at Workshop on HIV/AIDS and Adult Mortality in Developing Countries. New York, 8-13 September. Population Division.
- Mahy, M and Zaba, B. 2005. *Measuring child mortality in AIDS-affected countries*. <http://www.un.org/esa/population/publications/adultmort/Adultmortality.htm>
- Mathers, CD, Ma Fat, D, Inoue, M, Rao, C and Lopez, AD. 2005. "Counting the dead and what they die from: An assessment of the global status of cause of death data":171-177.
- Montagu, D, Yamey, G, Viscont, A, Harding, A and and Yoong, J. 2011. "Where Do Poor Women in Developing Countries Give Birth? A Multi-Country Analysis of Demographic and Health Survey Data", *Public Library of Science ONE* **2**(6)
- Moultrie, TA, Sayi, TS and Timaeus, IM. 2012. "Birth intervals, postponement, and fertility decline in Africa: a new type of transition?", *Population Studies* **66**(3):241-258.
- Mutemaringa, T. 2011. "The impact of HIV on estimates of child mortality derived using the summary birth history (CEB/CS) method." Unpublished thesis, Thesis submitted in partial fulfilment of the Degree of Master of Philosophy in the Faculty of Commerce University of Cape Town.
- Mwale, M. 2005. *Malawi Demographic and Health Survey 2004*. Calverton, Maryland: NSO and ORC Macro.: National Statistical Office.
- Nannan, N., Dorrington, R.E, Laubscher, R., Zinyakatira, N., Prinsloo, M., Darikwa, T., Matzopoulos, R. and and Bradshaw, D.. 2012. *Under-5 Mortality Statistics in South Africa: Shedding Some Light on the Trend and Causes 1997-2007*. Cape Town: South African Medical Research Council.
- Oliveras, E, Ahiadeke, C and Hill, AG. 2008. "Clinic-based surveillance of adverse pregnancy outcomes to identify induced abortions in Accra Ghana", *Studies in Family Planning* **39**(2):133-140.
- Rutstein, S.O and Rojas, G. 2006. *Guide to DHS Statistics* Macro Calverton, Maryland: http://www.ophi.org.uk/wp-content/uploads/OPHI-HDCA-SS2012-Guide_DHS_Statistics.pdf.
- Schoumaker, B. 2012. "Detecting omissions of recent births in birth histories in Sub-Saharan Africa." Unpublished thesis, Belgium: Universite catholique de Louvain.
- Sullivan, J.M. 1972. "Models for the estimation of the probability of dying between birth and exact ages of early childhood", *Population Studies* **26**(1):79-97.
- Sullivan, J.M., Bicego, G.T. and Rutstein, S.O 1990. *Assessment of the quality of data used for the direct estimation of infant and child mortality in the Demographic and Health Surveys. In: An assessment of DHS-I data quality. Demographic and Health Surveys Methodological Reports No.1*. Columbia Maryland: Institute for Resource Development / Macro systems:

- United Nations, Department of Economic and Social Affairs, Population Division. 2013. *World Population Prospects: The 2012 Revision, CD-ROM Edition*. <http://esa.un.org/wpp/Excel-Data/mortality.htm>. Accessed: 04/05 2013.
- Walker N, Hill K and F, Zhao. 2012. "Child Mortality Estimation: Methods Used to Adjust for Bias due to AIDS in Estimating Trends in Under-Five Mortality. ", *PLoS Med* **9**(8)
- Ward, P and Zaba, B. 2008. "The effect of HIV on the estimation of child mortality using the children surviving/children ever born technique", *Southern African Journal of Demography* **11**(1):11-73.
- World Health Organisation, (WHO). 2013. *Progress on infant mortality per 1000 live births*. http://www.who.int/gho/urban_health/outcomes/infant_mortality/en/. Accessed: 31/07/2013

APPENDICES

Appendix A

Code for calculation of infant mortality rates using the Blacker-Brass method

```
program IMR_BB
  version 12
  ***Calculates infant mortality rate based on children born within 24 months from survey date from those reported
  dead and alive**
  ****Syntax is <input_file_name> <initial_year> <final_year> <covar>
  args File_name Covariate
  tempvar old_file maxobs value covar new_file
  set more off
  preserve
  clear
  quietly {
  capture log close
  log using `File_name'.log, append
  log off
  quietly describe using `File_name',s
  scalar maxobs = r(N)
  local value = int((maxobs/1000)*5)
  local new_file temp
  use `File_name'
  replace v007 = v007+1900 if v007<1900 & v007>20
  replace v007 = v007+2000 if v007<20
  ** Keep only the variables needed**
  keep v005 b3 b7 v002 v008 v013 v007 b5 v000 b6
  ****generating pseudo random variables***
  egen pseudorand1=cut(v002),group(10)
  egen pseudorand2=cut(v008),group(10)
  replace pseudorand1 = pseudorand1/10+0.05
  replace pseudorand2 = pseudorand2/10+0.05
  correl pseudorand1 pseudorand2
  gen dob = b3+pseudorand1
  gen aad = b6-100*(int(b6/100)) + pseudorand2/31 if int(b6/100)==1 & (b6-100*int(b6/100))!=99
  replace aad = b6-100*(int(b6/100)) + pseudorand2 if int(b6/100)==2 & (b6-100*int(b6/100))!=99
  replace aad = 12*(b6-100*(int(b6/100)) + pseudorand2) if int(b6/100)==3 & (b6-100*int(b6/100))!=99
  gen dod = dob+aad
  format aad %5.2f
  format dod %6.2f
  **Create date of birth variable****
  gen datbrtmo = b3 + pseudorand1
  **Create age at death in months**
  gen agedthmo = b7 + pseudorand2 if b6 >= 200
  replace agedthmo = b7 + (pseudorand2*12) if b6 > 300
  replace agedthmo = cond(b6 < 200, (b6+0.5-100)/30.5, agedthmo)
  ***generating cmc between date of birth and interview date***
  gen age_cmc = v008-datbrtmo
```

```

***births within the past 10 years from survey date in 24 month intervals***
recode age_cmc (0/23.9999=1 "0-24")(24/47.9999=2 "24-48")(48/71.9999=3 "48-72")(72/93.9999=4 "72-
96")(96/120=5 "96-120"), gen(age_cmc_grp)
replace age_cmc_grp=. if age_cmc >120
***deaths within the past 10 years from survey date in 24 month intervals***
gen age_death =(v008-datbrtmo+agedthmo)
recode age_death (0/23.9999=1 "0-24")(24/47.9999=2 "24-48")(48/71.9999=3 "48-72")(72/93.9999=4 "72-
96")(96/120=5 "96-120"), gen(age_death_grp)
replace age_death=. if age_death >120
gen b5_1 =0 if age_cmc_grp==1
recode b5_1 0=1 if age_death_grp==1
gen b5_2 =0 if age_cmc_grp==2
recode b5_2 0=1 if age_death_grp==2
gen b5_3=0 if age_cmc_grp==3
recode b5_3 0=1 if age_death_grp==3
gen b5_4 =0 if age_cmc_grp==4
recode b5_4 0=1 if age_death_grp==4
gen b5_5 =0 if age_cmc_grp==5
recode b5_5 0=1 if age_death_grp==5
****generating results variables****
sum b5_1 [fw=v005]
gen imr1 = `r(sum)'/`r(N)'
gen minimr1 = (`r(sum)'/`r(N)')-(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
gen maximr1 = (`r(sum)'/`r(N)')+(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
sum b5_2 [fw=v005]
gen imr2 = `r(sum)'/`r(N)'
gen minimr2 = (`r(sum)'/`r(N)')-(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
gen maximr2 = (`r(sum)'/`r(N)')+(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
sum b5_3 [fw=v005]
gen imr3 = `r(sum)'/`r(N)'
gen minimr3 = (`r(sum)'/`r(N)')-(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
gen maximr3 = (`r(sum)'/`r(N)')+(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
sum b5_4 [fw=v005]
gen imr4 = `r(sum)'/`r(N)'
gen minimr4 = (`r(sum)'/`r(N)')-(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
gen maximr4 = (`r(sum)'/`r(N)')+(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
sum b5_5 [fw=v005]
gen imr5 = `r(sum)'/`r(N)'
gen minimr5 = (`r(sum)'/`r(N)')-(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
gen maximr5 = (`r(sum)'/`r(N)')+(1.96*sqrt(`r(sum)'/`r(N)')*(1-`r(sum)'/`r(N)')/`r(N)'))
tempfile imrs
save `imrs', replace
keep i* m* v000 v007
keep if _n==1
save `imrs', replace
use "C:\Users\Malvern\Desktop\dhs_try\Blacker_Brass_main.dta", clear
append using `imrs'
save "C:\Users\Malvern\Desktop\dhs_try\Blacker_Brass_main.dta", replace
}
end

```

Appendix B

Code for calculation of infant mortality rates using the direct estimation method

```
*version 2.0.3 14 June 2010
program IMR_1
*args File_name Start_year End_year Covariate
args File_name Covariate /*Start_year End_year*/
tempvar old_file maxobs value covar new_file
set more off
** save any data set in memory
preserve
clear
quietly {
capture log close
log using `File_name'.log, replace
log off
quietly describe using `File_name',s
scalar maxobs = r(N)
local value = int((maxobs/1000)*5)
local new_file temp
***set mem `value'm***
use `File_name'
replace v007 = v007+1900 if v007<1900
local Start_year = v007-2
local End_year = v007
de,s
***set seed for random number generator
set seed 987654321
** Keep only the variables needed
keep v005 b3 b6 b7 b5 v002 v007 v008 v013 v000
** Check to see whether data are probably DHS type
confirm numeric variable v005 b3 b6 b7 v002 v007 v008
**Create variables for U5MR_calcs
** Normalize the weight to mean 1.0
gen weight1 = v005/1000000
****generating pseudo random variables***
egen pseudorand1=cut(v002),group(10)
egen pseudorand2=cut(v008),group(10)
replace pseudorand1 = pseudorand1/10+0.05
replace pseudorand2 = pseudorand2/10+0.05
correl pseudorand1 pseudorand2
gen dob = b3+pseudorand1
gen aad = b6-100*(int(b6/100)) + pseudorand2/31 if int(b6/100)==1 & (b6-100*int(b6/100))!=99
replace aad = b6-100*(int(b6/100)) + pseudorand2 if int(b6/100)==2 & (b6-100*int(b6/100))!=99
replace aad = 12*(b6-100*(int(b6/100)) + pseudorand2) if int(b6/100)==3 & (b6-100*int(b6/100))!=99
gen dod = dob+aad
format aad %5.2f
format dod %6.2f
**Create date of birth variable****
gen datbrtmo = b3 + pseudorand1
**Create age at death in months
```

```

        gen agedthmo = b7 + pseudorand2 if b6 >= 200
        replace agedthmo = b7 + (pseudorand2*12) if b6 > 300
        replace agedthmo = cond(b6 < 200, (b6+0.5-100)/30.5, agedthmo)
        scalar startyear = `Start_year'
        scalar endyear = `End_year'
        capture gen covar = `Covariate'
** Save the data
        quietly save `new_file', replace
** Check the date ranges
if startyear < 1960 | endyear < 1960 | startyear > 2015 | endyear > 2015 {
    di in red "Beginning or ending period outside range 1960 to 2015"
    error 1
}
if startyear > endyear {
    di in red "End year must be equal to or greater than beginning year"
    error 1
}
** Start the calculations
log on
}
di in yellow _dup(63) " "
di _newline(2) "Life Table Mortality Risks in Childhood for Period " startyear " - " endyear ":"
di _column(10) "(using data from file " "File_name')"
capture confirm numeric variable covar
if _rc == 0 {
    summarize `Covariate'
    local i = _result(5)
    local j = _result(6)
    if (`j' - `i') > 10 {
        di in red as err "Values of Covariate `Covariate' exceed 10; too many values"
        error 1
    }
    forvalues k = `i'/'j' {
        noisily disp _newline(1)
        drop if covar ~= `k'
        noisily disp "Probabilities of Dying by Exact Ages of Childhood for : " "`Covariate' = " `k'
        log off
        IMR_calcs_1 `Start_year' `End_year'
        clear
        use `new_file'
    }
        clear
}
else {
    noisily disp _newline
    noisily disp "Probabilities of Dying by Exact Ages of Childhood"
    log off
    IMR_calcs_1 `Start_year' `End_year'
}
log close
end

```

```

** This subroutine does the calculations for U5MR
*! version 2.0.1 25 March 2010
program IMR_calcs_1
  version 10.0
  args Start_year End_year
  tempvar year yrpth next e_a e_d fac1 fac5
  *** Do routine called by U5MR to generate deaths, exposure, death rates
  ** and probabilities of dying in childhood in specified years using DHS data
  ** For syntax see U5MR.do
quietly {
set more off
compress
local yearmin = `Start_year' - 1900
local yearmax = `End_year' - 1900
confirm numeric variable weight1 datbrtmo agedthmo
** Create variables for deaths, exposure and weighted cases at each age
forvalues i = 0/60 {
  gen d`i' = 0
  gen exp_a`i' = 0
  gen exp_d`i' = 0
  gen n`i' = 0
}
** Calculate year of death
gen `yrpth' = int((agedthmo + datbrtmo)/12)
** Generate tally variables
gen `next' = .
gen `e_a' = .
gen `e_d' = .
** Classify and sum deaths and exposure in each year by month of age
** Starting point for century months is set to January 1970
forvalues year = `yearmin'/'yearmax' {
  capture drop adj
  gen int adj = (`year' - 70)*12
#delimit ;
forvalues i = 0/59 { ;
  replace d`i' = cond(`yrpth' == `year' & agedthmo >= `i' & agedthmo < `i'+1, d`i' + weight1, d`i');
};
** Calculate Exposure to risk in year by month of age;
forvalues i = 0/59 { ;
  replace `e_a' = weight1 if
    datbrtmo >= 841 - `i'+ adj & datbrtmo <= 852-`i'+ adj & agedthmo >= `i'+1;
  replace `e_a' = weight1*(datbrtmo - (840+ adj - `i')) if
    b3 == (840 - `i'+ adj) & agedthmo >= `i'+1;
  replace `e_a' = weight1*(853-`i'+ adj - datbrtmo) if
    b3 == (852 - `i'+ adj) & agedthmo >= `i'+1;
  replace `e_d' = weight1*(agedthmo - `i') if datbrtmo >= (841 - `i' + adj) & datbrtmo <
    (852 - `i'+ adj) & b7 == `i';
  replace `e_d' = weight1*max((datbrtmo + agedthmo - 841 - adj), 0) if b3 == (840 - `i' +adj) &
    b7 == `i';
  replace `e_d' = weight1*min((agedthmo - `i'), (853 + adj - `i' - datbrtmo)) if b3
    (852 - `i' +adj) & b7 == `i';
};
}

```

```

replace exp_a`i' = cond(`e_a' != ., exp_a`i' + `e_a', exp_a`i');
replace exp_d`i' = cond(`e_d' != ., exp_d`i' + `e_d', exp_d`i');
replace n`i' = cond(`e_a' != ., n`i' + `e_a', n`i');
replace n`i' = cond(`e_d' != ., n`i' + `e_d', n`i');
replace `e_a' = . ;
replace `e_d' = . ;
} ;
} ;

#delimit cr
forvalues i = 0/59 {
    egen event`i' = total(d`i')
    egen expa`i' = total(exp_a`i')
    egen expd`i' = total(exp_d`i')
    egen N`i' = total(n`i')
    gen expo`i' = (expa`i' + expd`i')/12
}
if _n == 1 {
forvalues i = 0/59 {
    gen nmx`i' = cond(expo`i' > 0, event`i'/expo`i', 0)
    gen a`i' = cond(`i' == 0, expd`i'/(event`i'*12), 1/24)
    gen nqx`i' = (nmx`i'/12)/(1+((1/12)-a`i')*nmx`i')
}
gen nnmr = nqx0
gen `fac1' = 1
forvalues i = 0/11 {
    replace `fac1' = `fac1'*(1-nqx`i')
}
gen imr = 1 - `fac1'
gen pnnmr = imr - nnmr
gen `fac5' = 1
forvalues i = 0/59 {
    replace `fac5' = `fac5'*(1-nqx`i')
}
gen q5 = 1 - `fac5'
gen q41 = (q5 - imr)/(1 - imr)
** Calculate "simple random sample" estimates of confidence intervals for imr and q5:
gen numimr = 0
gen denomimr = 0
gen numq5 = 0
gen denomq5 = 0
forvalues i = 0/11 {
    replace numimr = numimr + event`i'/(N`i'*(N`i'-event`i'))
    replace denomimr = denomimr + log((N`i'-event`i')/N`i')
}
gen hatsimr = sqrt(numimr/(denomimr*denomimr))
gen maximr = imr^(exp(hatsimr*1.96*-1))
gen minimr = imr^(exp(hatsimr*1.96))
forvalues i = 12/59 {
    replace numq5 = numq5 + event`i'/(N`i'*(N`i'-event`i'))
    replace denomq5 = denomq5 + log((N`i'-event`i')/N`i')
}

```

```
}
replace numq5 = numq5+numimr
replace denomq5 = denomq5 + denomimr
gen hatsq5 = sqrt(numq5/(denomq5*denomq5))
gen maxq5 = q5^(exp(hatsq5*1.96*-1))
gen minq5 = q5^(exp(hatsq5*1.96))
}
}
rename imr imr1
rename minimr minimr1
rename maximr maximr1
tempfile imrs
save `imrs', replace
keep imr1 nnmr minimr1 maximr1 v000 v007
keep if _n==1
save `imrs', replace
use "C:\Users\Malvern\Desktop\dhs_try\neonate", clear
append using `imrs'
save "C:\Users\Malvern\Desktop\dhs_try\neonate.dta", replace
```

