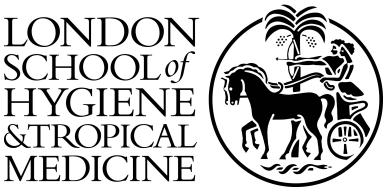


Under-five mortality estimates for sub-Saharan Africa: an inquiry into data sources and estimation methods

Hallie Carlin Eilerts

Thesis submitted in accordance with the requirements for the degree of
Doctor of Philosophy
of the
University of London

March 2023

Department of Population Health
Faculty of Epidemiology and Population Health
LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the UK Economic and Social Research Council (ESRC)

Declaration of Authorship

I, Hallie Carlin Eilerts, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

There is a lack of high-quality data on levels, trends, and age patterns of under-5 mortality (U5M) in sub-Saharan Africa. Civil registration and vital statistics (CRVS) systems are incomplete or non-existent across the region, which has led to a reliance on alternative data sources and model-based estimates. Though the available information indicates that there has been substantial improvement to child health over the past decades, the region remains confronted with the highest rates of U5M in the world. Accurate measurement will be essential to tracking and accelerating gains in U5M, and providing the evidence to guide effective health policy making.

This thesis examined bias and uncertainty in U5M estimates for sub-Saharan Africa from different data sources. Of primary interest were data collected on pregnancies, pregnancy outcomes, and early mortality in Health and Demographic Surveillance Systems (HDSS). This work was conducted through four studies which form the main chapters of this thesis. In the first three studies, HDSS monitoring of pregnancies and newborns was evaluated through cross-site and external comparisons with survey and clinical data. The analyses were performed across a number of populations in unique epidemiological settings and took into account site-specific characteristics and data collection protocols. In the final study, the direction of inquiry was reversed, and the prospective data of HDSS were used to shed light on bias in U5M estimates from retrospective maternity history questionnaires in high-HIV settings.

This thesis supports existing calls to improve data collected on perinatal and U5M in sub-Saharan Africa. HDSS sites are valuable resources for this purpose, often serving as the only population-based data systems in countries lacking complete CRVS. Collectively, the findings contribute to improving understanding of U5M in sub-Saharan Africa and our ability to accurately measure it in the context of population-based surveillance systems.

Acknowledgements

I would first like to thank my supervisors Georges Reniers, Jeffrey W. Eaton, and Julio Romero Prieto. I have learned so much from each one, and am immeasurably grateful for their dedication to my research and professional development. They have been the greatest advocates for my work, opened doors to valuable experiences and opportunities, and worked to ensure that this experience was as enriching as possible. This has truly been the most wonderful gift.

I would like to extend my deep gratitude to collaborators at the Africa Health Research Institute. I spent a brief research stay at the Somkhele campus during the early stages of my doctoral program. I was warmly welcomed and remember this experience very fondly. I appreciate all those who provided me with support and guidance during this time and after, in ongoing collaborations. In particular, thanks to Kobus and Carina Herbst, Dickman Gareta, Kathy Baisley, Philippa Matthews, Manisha Yapa, and Nomsa. One day, Carina and Philippa directed my attention towards two brown-hooded kingfishers that were perched among hundreds of busy masked weavers on a tree outside the office window. I have been a dedicated bird-watcher ever since. I thank you, though I'm not sure if those who have to listen to me talk about birds do.

I want to thank the many kind and brilliant individuals at the Kenya Medical Research Institute who have supported this research. This includes but is not limited to Sammy Khagayi, Fredrick Odongo, Daniel Kwaro, Julie Ambia, and David Obor. After a productive week of meetings at the Siaya office in October 2019, I was very excited at the prospect of returning for a longer period to continue our work together. Sadly, this did not come to pass, and I can't help but be struck with a sense of dramatic irony when reviewing past timelines and plans for activities that were slated to begin in April 2020. I regret that I was not able to work with you all more in-person, but am extremely grateful for your virtual collaboration that was invaluable all the same.

I extend my sincere gratitude to Momodou Jasseh of the Medical Research Council Unit, The Gambia at LSHTM. His 2003 doctoral thesis on age patterns of mortality within childhood in sub-Saharan Africa served as an important guide to me throughout this process. I am thankful for his support of this research and it was an honour to collaborate with him.

This doctoral research was nested within a National Institute of Child Health and Human Development grant on "Age patterns of under-five mortality in human populations" (NICHD R01HD090082). Through this project, I encountered many inspiring individuals who provided valuable counsel on my work and were extremely encouraging to me as an early career researcher. My sincere thanks to them all, and especially Michel Guillot and Andrea Verhulst.

I am grateful to the UK Economic and Social Research Council for funding my doctoral research. This grant included a year of training at the European Doctoral

School of Demography, which was instrumental to my development as a researcher. I thank the professors, mentors, and friends who made that experience so fun and enriching. I am also deeply appreciative of the funding extensions that I received from UK Research and Innovation and LSHTM due to disruptions caused by the COVID-19 pandemic.

I am thankful to all the staff members at LSHTM who have made it such a wonderful place to work. I would especially like to thank Jenny Fleming and Lauren Dalton for their kindness and ever-present support in all matters organizational and administrative.

Thank you to my mentors, colleagues, officemates, and friends (which are by no means mutually exclusive categories) at LSHTM. Over the years, I have been so grateful for your support, advice, commiseration, shared laughs, and companionship. You made each day more enjoyable. To Amy, Prima, and Komal, thank you for providing London-based housing (at various times) and friendship (all the time). To Elvis Dzelamonyuy Chem, I will forever remember you and be inspired by your intelligence, passion, and kindness.

Finally, thank you to my family. There are not words to sufficiently convey my love and gratitude for you all, or at least, I would never be satisfied with my attempts at finding them. Thank you to my parents for their infinite love and support, and to whom I owe everything. Thank you to Richard for being my love and solid ground. You encouraged me to take this on and have been there every step of the way.

Contents

List of Figures

List of Tables

List of Abbreviations

AHRI	Africa Health Research Institute
AIDS	Acquired Immune Deficiency Syndrome
ALPHA	Analysing Longitudinal Population-based HIV/AIDS data on Africa
ANC	Antenatal Care
ART	Antiretroviral Treatment
CDC	Centres for Disease Control
CHAMPS	Child Health and Mortality Prevention Surveillance
CHESS	Comprehensive Health and Epidemiological Surveillance Systems
CI	Confidence Interval
CPS	Contraceptive Prevalence Surveys
CRVS	Civil Registration and Vital Statistics
DHS	Demographic and Health Survey
EDD	Estimated Delivery Date
ENAP	Early Newborn Action Plan
ESRC	Economic and Social Research Council
FBH	Full Birth History
FPH	Full Pregnancy History
HDSS	Health and Demographic Surveillance System
HIC	High-Income Country
HIV	Human Immunodeficiency Virus
HMD	Human Mortality Database
HMIS	Health Management Information Systems
HR	Hazard Ratio
ICD	International Classification of Disease
INDEPTH	International Network for the Demographic Evaluation of Populations and their Health
IPW	Inverse Probability Weight
IQR	Interquartile Range
KEMRI	Kenya Medical Research Institute
LLMICs	Low- and Lower-Middle-Income Countries
LMICs	Low- and Middle-Income Countries
LMP	Last Menstrual Period
LSHTM	London School of Hygiene & Tropical Medicine

MDG	Millennium Development Goal
MICS	Multiple Indicator Cluster Survey
MLT	Model Life Table
MRC	Medical Research Council
NCD	Non-Communicable Disease
NIHCD	National Institute of Child Health and Human Development
NIR	No-Information Rate
NMR	Neonatal Mortality Rate
OLS	Ordinary Least Squares
PDA	Personal Digital Assistant
PIRL	Point-of-contact Interactive Record Linkage
PMR	Perinatal Mortality Rate
PMTCT	Prevention of Mother-to-Child Transmission
PNMR	Postneonatal Mortality Rate
SBH	Summary Birth History
SBR	Stillbirth Rate
SDG	Sustainable Development Goal
TIPS	Time Prior to the Survey
U5M	Under-5 Mortality
UN	United Nations
UN IGME	United Nations Inter-agency Group for Child Mortality Estimation
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
WFS	World Fertility Surveys

Chapter 1

Introduction

Under-5 mortality (U5M) is a core indicator of population health that is sensitive to the quality of health systems, socio-economic conditions, and the environment. In the last few decades, the reduction of U5M has become a central aim of the international community. There has been substantial progress in reducing U5M globally, though it has been uneven. Strong regional disparities persist, and children face vastly divergent prospects of living a healthy life depending on their place of birth (UN IGME, 2021). Sub-Saharan Africa is the global region with the highest rates of U5M. The current rates of reduction in many African countries are also insufficient to meet the child health targets of the Sustainable Development Goals (SDGs) (UN IGME, 2021). Progress must be accelerated to avoid unconscionable loss of life and improve health equity.

Accurate measurement of levels and trends of U5M is essential to bringing out its reduction. Detailed information on the distribution of deaths within the first five years of life is particularly important, given the close connection between age and cause of death. Such information enables the identification of particularly vulnerable populations and age groups, and provides important insight on the epidemiological environment. In this way, it serves as necessary evidence for formulating targeted interventions. As such health programming is put into action, data remain crucial for evaluating its impact and tracking progress. While data alone do not improve health, they are fundamental to focusing resources, guiding health policymaking, and improving accountability.

High-quality data on births, deaths, and causes of death are often lacking in the regions facing the greatest health challenges. Civil registration and vital statistics (CRVS) systems are the preferred sources for this information, though are mostly found in high-income countries due to the high costs and expansive infrastructure associated with establishing and maintaining them. With CRVS systems either incomplete or non-existent in countries throughout sub-Saharan Africa, population health data must come from other sources. Data from maternity history questionnaires collected in cross-sectional surveys have been essential for estimating U5M. Longitudinal monitoring conducted by Health and Demographic Surveillance Systems (HDSS) has

also provided rich insight into U5M trends and determinants. However, these alternative data sources are also affected by quality issues which obscure the state of child health, and lead to inaccuracies in measurement of U5M. Insufficient or inaccurate mortality metrics serve as a poor road map, ultimately leading to interventions that are wrongly targeted, less effective, and suboptimal in terms of resource allocation. It is in such settings that the link between material and data poverty can become both inextricable and self-reinforcing (Byass, 2009).

This thesis investigates bias and uncertainty in estimates of perinatal and U5M from contemporary data sources in sub-Saharan Africa. It also explores how data sources may be augmented or used in complement to improve overall understanding of health, as well as data collection strategies and estimation methods that may be used to mitigate bias.

1.1 Aims

The overall aim of this research is to examine bias and uncertainty in measurement of U5M in sub-Saharan Africa.

This research will be conducted in four separate but closely related studies. The specific objectives of each study are as follows:

1. To compare estimates of neonatal, postneonatal, infant, and child mortality from HDSS with household surveys in sub-Saharan Africa and the historic record, and evaluate whether disparities between data sources can be correlated with data quality or other attributes of the data source or population.

2. To assess whether pregnancy registration improves measurement of early mortality in HDSS.

3. To utilize linked antenatal clinic data to investigate HDSS pregnancy reporting completeness and bias in the measurement of adverse pregnancy outcomes and early mortality.

4. To leverage HDSS data to shed light on bias in birth history estimates of U5M arising from missing mothers in high-HIV settings.

1.2 Thesis structure

This thesis follows a research-paper based structure, with a total of seven chapters and four research papers.

In **Chapter One** (presented here), I outline the main aims of this thesis, lay out the thesis structure going forward, and provide information on my role in the research, ethical clearance, collaborating institutions, and funding.

Chapter Two provides broad background on U5M, data sources, and measurement challenges in sub-Saharan Africa. The research papers included in this thesis also contain more detailed background that is relevant to those chapters.

Chapter Three is the first research paper which investigates the age pattern of U5M in sub-Saharan Africa from 1990-2018. This paper has been published in Demographic Research, but is included herein using standard formatting. In this work, U5M estimates for African populations were calculated using data from 30 HDSS sites in the International Network for the Demographic Evaluation of Populations and their Health (INDEPTH Network), and full birth histories (FBHs) collected in 62 Demographic and Health Surveys (DHS) and 26 Multiple Indicator Cluster Surveys (MICS). The age pattern of U5M was compared between sources and against the historic record of high-quality data. A direct comparison was also made between levels of neonatal, postneonatal, and child mortality in HDSS and DHS/MICS estimates for the encompassing sub-national regions. Linear regression was used to examine whether differences between HDSS and DHS estimates were associated with data quality markers from each source and contextual attributes of interest.

The second research paper is presented in **Chapter Four**. This work examines whether registering pregnancies and following-up on their outcomes improves U5M estimation in HDSS. The analysis was conducted using data from three HDSS sites in The Gambia, Kenya, and South Africa that differed substantially in terms of their epidemiological profile and data collection protocols. In each site, U5M was estimated for cohorts of births with and without pregnancy registrations in the HDSS. Micro-level analysis was used to investigate differences in child survival by pregnancy registration status, controlling for associations between pregnancy registration and U5M risk factors.

Chapter Five is a research paper which leverages individually-linked antenatal care (ANC) clinic data to externally validate reports of pregnancies and their outcomes in the HDSS in Siaya, Kenya. ANC data on gestational age was used to characterize the timing of pregnancy registration in the HDSS, and investigate potential misclassification between reports of miscarriage and stillbirth. Demographic surveillance data was used to investigate the status of individuals whose ANC usage was not associated with a reported pregnancy outcome in the HDSS. Individuals who were missing pregnancy outcome reports despite their continued residency in the HDSS for at least one post-delivery data collection round were identified as likely cases of adverse pregnancy outcomes, and used to explore potential bias in HDSS measurement of stillbirths and early mortality.

The final research paper is included in **Chapter Six**. In this work, demographic and HIV surveillance data from the HDSS in Somkhele, South Africa were used to

investigate bias in birth history estimates of U5M over the course of the HIV epidemic. Differentials in child survival were examined by maternal HIV and antiretroviral treatment (ART) status over time. Standard demographic methods were applied to decompose reductions in U5M into parts attributable to changes in the share of HIV-positive mothers as opposed to changes in mortality rates. In the second part of the analysis, a series of continuous cross-sectional surveys were simulated in 12 HDSS sites belonging to the INDEPTH Network. U5M was calculated for the children of mothers who were present in each site at midyear, and compared to estimates for all children to approximate bias from missing mothers. Bias in estimates of U5M was compared across sites, survey years, and windows of retrospection. Additional comparisons were made between U5M estimates calculated for children whose mothers were not present in the site at midyear due to death or out-migration.

In **Chapter Seven**, I synthesize the results from the previous four chapters and review the key findings. I then discuss the strengths and limitations of this thesis, provide recommendations for programmes, policy, and future research, and conclude with some final remarks.

1.3 Thesis contribution

This thesis adds to a rich body of work which foregrounds issues of data quality in the measurement of U5M in settings where vital registration systems are incomplete. Investigating quality issues is essential as it provides a better understanding of the strengths and weaknesses of the data, as well as the insights to be gained from it. My research has its most direct applications to the collection and usage of HDSS data. HDSS represent an extremely valuable resource, often generating the only longitudinal population and health data for many settings throughout sub-Saharan. However, cross-site heterogeneity in data collection protocols and data quality has not been sufficiently studied, which to some extent, limits the data's usage. HDSS estimates of early mortality have also been considered implausibly low. I hope to improve understanding of bias in HDSS estimates of perinatal and U5M, how it can be influenced by data collection protocols, and avenues for attenuating it through pregnancy registration and linkage with routine programme data. Overall, this work will contribute to improving understanding of U5M in sub-Saharan Africa, and our ability to accurately measure it in the context of population-based surveillance systems.

1.4 Role of the candidate

I, the candidate, designed the studies in this thesis with the guidance and academic support of my supervisors, Georges Reniers, Jeffrey W. Eaton, and Julio Romero Prieto. For the first and fourth research papers, I obtained data through online repositories for the INDEPTH Network, DHS, MICS, and Human Mortality Database. Primary data for the second, third, and fourth research papers were provided by external collaborators. I conducted all data analysis, interpreted findings, and wrote the first drafts of each manuscript, as well as the body of this thesis. My supervisors and co-authors provided revisions on the methods of analysis and formulation of each research paper, which I incorporated in the final versions.

1.5 Ethical clearance

Ethical approval for this work was obtained from the Institutional Review Board of the London School of Hygiene & Tropical Medicine (LSHTM) under protocol #17845 and #25133. Ethics approval certificates are provided in Appendix A.

1.6 Collaborating institutions

This doctoral research was nested in a National Institute of Child Health and Human Development (NIHCD) grant on "Age patterns of under-five mortality in human populations" (award number: R01HD090082) led Principal Investigator Michel Guillot at the University of Pennsylvania. I participated in three workshops related to the project, which included researchers from the collaborating institutions of LSHTM, Johns Hopkins Bloomberg School of Public Health, and the Muséum National d'Histoire Naturelle, France.

I also worked closely with institutions operating three of the HDSS sites taking part in the NIHCD grant. This included the Africa Health Research Institute (AHRI), Kenya Medical Research Institute (KEMRI), and Medical Research Council (MRC) Unit, The Gambia at LSHTM. I spent four weeks based at AHRI's Somkhele field office in KwaZulu-Natal, South Africa. I also participated in a week-long series of strategic planning sessions at the KEMRI office in Kisumu, Kenya for a study that was to conduct follow-up interviews with individuals who were missing pregnancy outcomes in the HDSS. I met with Momodou Jasseh of the MRC Unit, The Gambia at LSHTM to discuss planned research with data from Basse HDSS. Following initial in-person meetings, I maintained regular correspondence with collaborators at each institution. My primary collaborators are listed as co-authors on each study. These individuals furnished the primary datasets used in studies two through four, provided insight

on HDSS data collection protocols, advised on analytical methods, and supported the formulation of final manuscripts.

By the invitation of my supervisor, Jeffrey W. Eaton, I was hosted as a visiting researcher at Imperial College in the MRC Centre for Global Infectious Disease Analysis throughout the duration of my doctoral research. In this capacity, I took part in weekly meetings with the HIV Inference Working Group. I spent one-year as a visiting student to the University of Pennsylvania Population Studies Center, under the supervision of Michel Guillot. However, my interactions with the Population Studies Center and colleagues were primarily virtual due to the COVID-19 pandemic.

1.7 Funding

This research was funded through the Economic and Social Research Council (ESRC) Studentship at the UCL, Bloomsbury and East London Doctoral Training Partnership. I received an award for a 1+3 ESRC European Doctoral School of Demography and PhD studentship. In addition to this, I received an Advanced Quantitative Methods (AQM) Research Training Support Grant, which promotes training in and the use of AQM in the PhD research.

This doctoral research initially included a fieldwork component that was planned to take place in Siaya, Kenya over several months beginning in April 2020. As the COVID-19 outbreak progressed and was declared a global pandemic in March 2020, these plans were suspended. It was necessary to replace this element of my dissertation and the research outputs which depended on it. This significantly delayed my progress, and I applied for two short-term funding grants. I was granted funding extensions of 3.74 months from the UK Research and Innovation (UKRI) COVID-19 Extensions Scheme and two months from the LSHTM Doctoral Student COVID-19 Support Scheme.

Chapter 2

Background

In this chapter, I provide context for this thesis as a whole. I begin with a description of the mortality metrics that will be used in subsequent chapters, and then lay out broad background on under-5 mortality (U5M) and its measurement in sub-Saharan Africa. This includes a review of historical trends, data sources, and the current state of progress towards global targets for U5M reduction.

2.1 Mortality metrics

U5M refers to the probability of dying between birth and five years of age. It is a key indicator for assessing child health, as well as socio-economic development more generally. Though useful as a summary measure, it also conceals important information about the distribution of deaths under age five. Mortality risk generally declines as children age, but the pace and shape of this decline differs across populations. Identifying ages where the trajectory of mortality stagnates or reverses is essential to understanding the epidemiological environment and formulating targeted health interventions.

Conventional mortality indicators subdivide the under-5 age range into neonatal, postneonatal, and child ages. These correspond to the probability of dying within the first 28 days of life, 28 days to one year, and one to five years. Neonatal and postneonatal ages are combined for measuring the level of mortality during infancy. Mortality during the first month of life may be further stratified by early and late neonatal mortality; the probability of dying within the first seven days, and the probability of reaching exact age 28 days, conditional on surviving the first week, respectively.

Historically, more attention has been paid to measuring mortality after birth than in-utero. This has begun to change in recent years, as the burden of stillbirths has gained in visibility on the global agenda (UN IGME, 2020a; World Health Organization, 2014). Stillbirth refers to a baby born with no signs of life after a given threshold, which is typically defined by gestational age or weight. The International Classification of Disease 10th revision (ICD-10) definition of stillbirth as a fetal death occurring at or after 28 weeks of gestation has become the standard for international comparisons (World Health Organization, 2013; Blencowe et al., 2016; UN IGME, 2020a). In

the mid-20th century, difficulties differentiating between infants who were stillborn versus those dying shortly after birth led to the practice of reporting the combined indicator of perinatal mortality; including both stillbirths and deaths occurring during the first week of life (Kramer, 2002; Lawn et al., 2008; Stanton et al., 2006). However, expert opinion has shifted regarding the suitability of this practice (Kramer, 2002; Lawn et al., 2008). The principal causes of stillbirths and early neonatal deaths differ substantially, and the use of a combined indicator risks masking the relatively more severe under-reporting of stillbirths (Kramer, 2002; Stanton et al., 2006). The importance of independent tracking of stillbirths notwithstanding, growing calls for stillbirths to be included among the key indicators for maternal and child health has led to the development of other composite measures. It has been proposed that stillbirths be combined with under-5 deaths for new measures of "stillbirth-adjusted" neonatal and infant mortality (Hathi, 2021), or "total under-5 mortality" (Liu, Hill, and Oza, 2016), which more accurately reflect the state of child health.

Fetal death is defined as one occurring "prior to the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of pregnancy" (World Health Organization, 2013). The distinction between early fetal death and stillbirth (i.e. late fetal death) was originally based on a threshold of perceived viability (Blencowe, 2020). Significant medical advances in preterm care in the past decades have made survival possible at gestational ages of 22 or 23 weeks (Ancel et al., 2015; Mehler et al., 2016; Stoll et al., 2015). However, such gains have been mostly attained in highly-resourced settings, and remain largely irrelevant in countries with the highest rates of adverse pregnancy outcomes (Stanton et al., 2006). Large heterogeneity in the reporting criteria of fetal deaths complicates international comparisons of such data (Blencowe et al., 2016; Mohangoo et al., 2013). In low-income and lower-middle-income countries (LLMICs), the issue is typically less one of comparability, and more of a general lack of high-quality data on both early and late fetal deaths (Blencowe et al., 2016; Frøen et al., 2009).

2.2 Historical perspective on U5M

The earliest attempts to measure the mortality of children date back to the origin of formal demography in the mid-seventeenth century (Pozzi and Fariñas, 2016). Using information published in the London bills of mortality, John Graunt created the first rudimentary life table consisting of tallies of births and deaths attributed to different diseases (Rusnock, 2009). As this information did not initially contain information on age of death, Graunt inferred infant mortality from deaths attributable to causes that mostly affected early ages (Rusnock, 2009). Graunt hypothesized that deaths among infants and children regularly accounted for one-third of the annual total (Rusnock, 2009).

Techniques for measuring mortality were developed further by physicians, arithmeticians, and government officials in the eighteenth century, often incorporating information from parish records, local censuses, and hospital records (Pozzi and Fariñas, 2016). Such sources from England and France indicated high levels mortality among children, with close to one in four dying during infancy (Rusnock, 2009). As statistical techniques for mortality calculation were becoming more formalized in the nineteenth century, the concept of an overall rate of infant mortality came into focus (Armstrong, 1986; Pozzi and Fariñas, 2016). Changes in this rate over time were regularly assessed during the latter half of century (Armstrong, 1986).

The first five years of life have been a time of high mortality risk for the majority of recorded human history (Pozzi and Fariñas, 2016). Reductions in the severity and frequency of mortality crises were first observed in western Europe in the mid-eighteenth century (Omran, 1971). These improvements were initially modest, though they gradually transformed into sustained declines (Omran, 1971). In the nineteenth century, mortality reduction accelerated, with improvements to the health of children accounting for substantial gains in life expectancy across Europe, North America, and Australia (Defo, 2014; Pozzi and Fariñas, 2016). This decline in mortality that was followed by a fall in fertility rates was termed the "demographic transition" (Notestein, 1945). The theory of an "epidemiological" or "health" transition was used to described the mortality decline in terms of the fall in the relative importance of infectious and acute causes of death at early ages, and the increasing salience of non-communicable diseases (NCDs) and deaths at older ages (Omran, 1971).

Historical study of trends of U5M in sub-Saharan Africa has been limited by the scarcity of data. Some of the earliest information on births and deaths comes from parish records kept by African missionaries in the late nineteenth and early twentieth centuries (Notkola, 2017; Walters, 2016). Censuses conducted during the early 1900s and historical and administrative records from European colonies constitute other sources of data (Curto, 1994). However, these colonial-era sources were embedded in hegemonic power structures and selective in terms of their coverage and constituencies (Lockwood, 1990; Walters, 2021). Drawing demographic insights from such data is very difficult; an exercise that must be undertaken cautiously and requires significant engagement with the broader historical context (Walters, 2021). The volume of data grew in the second half of twentieth century, with an increase in the number of surveys, censuses, and studies focused on demographic surveillance (Cleland, 1996).

Model life tables (MLTs) were developed to assist with mortality estimation in populations where data was lacking. In 1955, the United Nations (UN) Department of Social Affairs proposed a set of model life tables for inferring age schedules of mortality using the level of infant mortality as an entry parameter (United Nations, 1955). Perhaps the most prominent set of model age patterns was developed by Coale and

Demeny in 1966 (Coale and Demeny, 1966). These MLTs were indexed by life expectancy at age 10, and included four distinct regional families of age patterns. In 1982, the UN MLTs were revised to include five families of regional patterns (United Nations, 1982). There have been several innovations in MLTs since this time, though the Coale and Demeny and UN systems are still in common use today (Guillot et al., 2022). Given that they were primarily based on data from European populations, their applicability to sub-Saharan African contexts was and remains questionable (Guillot et al., 2012; Verhulst et al., 2021).

The relevance of the theoretical frameworks of the demographic and epidemiologic transition to developing countries, and those of sub-Saharan Africa in particular, is a subject of debate (Bongaarts and Watkins, 1996; Defo, 2014; Gaylin and Kates, 1997; Maher and Sekajugo, 2010; Reher, 2004; Teitelbaum, 1975; Thornton, 2001). A substantial decline in U5M was observed in low- and middle-income countries (LMICs) during the mid-twentieth century (Ahmad, Lopez, and Inoue, 2000). This took place at a much faster pace than during the transition which began in Europe in the late nineteenth century (Collinson et al., 2014; Cutler, Deaton, and Lleras-Muney, 2006; Defo, 2014). The decline has been primarily attributed to the widespread immunization of children and rapid introduction of antibiotics (Cutler, Deaton, and Lleras-Muney, 2006; Gwatkin, 1980). Other improvements to standards of living and environmental conditions also played a role, but one that was relatively less significant than observed in nineteenth century Europe (Preston, 1975; Gwatkin, 1980). In another departure from the European experience, the pace of the mortality decline in many LMICs began to stall earlier than expected in the 1970s (Ahmad, Lopez, and Inoue, 2000; Garenne, 2006; Gould, 1998; Gwatkin, 1980). The trend towards life expectancy convergence between low and high-income countries (HICs) also weakened (Gwatkin, 1980). Trends towards globalization and urbanization had been accompanied by negative lifestyle changes related to unhealthy diets and habits (Collinson et al., 2014; Maher and Sekajugo, 2010; Tollman et al., 2008). Many LMICs faced a growing burden of NCDs, along with the continued high prevalence of infectious diseases. The impacts on the health of children would be both direct and indirect, with NCDs often afflicting the adults responsible for their care (Tollman et al., 2008).

Slowdowns in the decline of U5M in sub-Saharan Africa were likely due to several factors that varied across countries and over time. It has been argued that the "vertically implemented" health campaigns of international organizations did not penetrate into the harder-to-reach segments of populations (Cutler, Deaton, and Lleras-Muney, 2006). "Children's diseases" (e.g. diphtheria, measles, tetanus, polio, and whooping cough) that had been largely eradicated in HICs continued to have detrimental consequences (Collinson et al., 2014; Cutler, Deaton, and Lleras-Muney, 2006). Malaria, which had been fully controlled using insecticides and environmental measures in HICs, also remained a serious threat (Cutler, Deaton, and Lleras-Muney, 2006;

Gwatkin, 1980; Snow et al., 2017). Large-scale sustained efforts at malaria elimination had been primarily conducted in countries at the northern and southern bounds of endemic transmission, and islands such as Madagascar and Mauritius (Snow et al., 2012). Prevalence had slowly declined across most of sub-Saharan Africa in the 1950s and 60s, though began to increase in the 1980s (Snow et al., 2017). The reversal was influenced by factors such as growing chloroquine resistance, climate anomalies, and the wavering commitments among governments and international organizations to controlling the epidemic (Snow et al., 2017).

Existing treatment options for other diseases were often not easily or widely accessible. Newly available antibiotic regimes for tuberculosis required sustained engagement with healthcare systems which had patchy coverage and were frequently underresourced (Cutler, Deaton, and Lleras-Muney, 2006). Causes of death, such as acute respiratory infections and diarrhoeal disease, regularly accompanied by the underlying condition of malnutrition, were increasing in relative importance among under-5s (Gwatkin, 1980). Universal delivery strategies were lacking for practical inventions such as exclusive breastfeeding and oral rehydration therapy for children with diarrhoea (Cutler, Deaton, and Lleras-Muney, 2006; Hill and Amouzou, 2006; Jones et al., 2003).

The transition period in sub-Saharan Africa was also one of increasing urbanization, which had complicated effects on child health. Rates of U5M in urban areas in sub-Saharan Africa have been consistently lower than those of rural areas (Bocquier, Madise, and Zulu, 2011; Brockerhoff, 1990; Gould, 1998; Issaka, Agho, and Renzaho, 2016; Menashe-Oren and Bocquier, 2021). Initially, this was related to the superior healthcare and public health measures (e.g. improved sanitation and water supply systems) found in cities, as well as the higher socio-economic statuses of those who lived there (Gould, 1998). This disparity in health infrastructure can be traced back to colonial past, and the prioritization of strategies to serve the health needs of European immigrants residing in urban centres (Gould, 1998; Snow et al., 2012). However, the urban advantage may have faded both with gains in rural health and deterioration of urban living conditions (Bocquier, Madise, and Zulu, 2011; Gould, 1998). While there has been much variation in urbanization processes across sub-Saharan Africa, it appears that urban growth slowed in the 1980s as economic downturns disproportionately affected urban areas (Gould, 1998; Menashe-Oren and Bocquier, 2021). Public health infrastructure and health services in many cities were strained and unable to support larger populations (Gould, 1998). Under these conditions, the urban environment entailed numerous risk factors for child health related to urban poverty, overcrowded housing conditions, and issues of sanitation; all of which also hastened the spread of infectious disease (Collinson et al., 2014).

The late 1970s and early 1980s were also a time of political upheaval. U5M temporarily increased during struggles for independence in Namibia, Madagascar, Zimbabwe; and in the civil wars that followed independence in Angola, Chad, and Mozambique (Garenne, 2006). Other periods of political or economic uncertainty led to excess U5M in Ghana, Uganda, and Zambia (Garenne, 2006). Structural adjustment loans from international financial groups also required many countries to make cutbacks on health and social spending programmes (Defo, 2014). Advances in several aspects of socio-economic development were likely necessary to consolidate and sustain previous gains in child health. This included improvements to standards of living, and the promotion of a more equitable distribution of wealth and opportunities (Gwatkin, 1980; Preston, 1980). The important role of education, particularly women's education, in the decline of U5M would also become more clear with time (Cutler, Deaton, and Lleras-Muney, 2006; United Nations, 1986; Liu, Hill, and Oza, 2016).

In the 1990s, reductions in U5M in sub-Saharan Africa continued to decelerate, and reversals were observed in some countries heavily affected by the HIV epidemic[1] (Adetunji, 2000; Ahmad, Lopez, and Inoue, 2000; Boerma, Nunn, and Whitworth, 1998). At the beginning of the twenty-first century, new global commitments would be made for the eradication of preventable child deaths.

2.3 Sources of U5M data for sub-Saharan Africa

2.3.1 Censuses

Population-level censuses in their modern sense began in the mid-twentieth century in sub-Saharan Africa (Cleland, 1996; Lockwood, 1990), and as of 1960, around 22 countries had conducted at least one census (Cleland, 1996). Today, census data is available in a majority of sub-Saharan African countries (Odimegwu, Chisumpa, and Somefun, 2018). Census data should be comprehensive, aiming to quantify the size and age distribution of the total population. This permits the estimation of age-specific mortality rates that are unaffected by sampling error. Censuses may also be used to provide a sampling frame for other surveys.

Censuses are large-scale undertakings that require substantial time and resource investment. These factors limit the frequency of implementation, as well as the complexity and length of questionnaires. Censuses have generally occurred at intervals of 10 years or more in sub-Saharan African countries (Jasseh, 2003), which does not allow for timely monitoring of levels and trends in U5M. Information on child survival often consists of summary questions posed to mothers regarding the number of children they have given birth to, and how many are still surviving (Hill, 1991). Techniques for

[1] Slowdowns or reversals in the decline of U5M due to the HIV epidemic were most evident in Botswana, Namibia, South Africa, Zimbabwe, and Zambia in southern Africa; Kenya, Uganda, and Tanzania in eastern Africa; and Cameroon and Côte d'Ivoire in western Africa (Garenne, 2006).

indirectly estimating U5M from this information are described in greater detail in the maternity history questionnaires sub-section below.

2.3.2 Civil registration and vital statistics systems

The primary functions of civil registration and vital statistics (CRVS) systems are administrative, whereby they provide individuals with documentation to establish their identity, familial relations, civic and legal status (AbouZahr et al., 2015a). CRVS are also the preferred data for generating population-level, up-to-date demographic statistics. CRVS systems record the occurrence of births and deaths on a continuous basis. Most countries also collate information on cause-of-death, though this has been shown to vary substantially in quality (Mathers et al., 2005). Data collected by CRVS systems allow for the calculation of age-, sex-, and cause-of-death-specific mortality rates.

The registration of vital statistics is incomplete in many LMICs (AbouZahr et al., 2015b). It is estimated that less than one-half of births are registered sub-Saharan Africa, and under a quarter of deaths (Masquelier, Reniers, and Pison, 2014; World Bank, 2022). Notable exceptions include Seychelles, Mauritius, and South Africa where registration of births and deaths is estimated to be 90% or higher (World Bank, 2022).

International efforts to strengthen CRVS systems have not resulted in substantial improvements (Cleland, 1996; AbouZahr et al., 2015b). The United Nations Children's Fund (UNICEF) and the UN Statistics Division have provided support of a limited nature, and the World Bank statistical capacity building activities have not been focused on the improvement of CRVS systems (AbouZahr et al., 2015b). Initiatives led by the UN Economic and Social Commission and Population Fund have established international standards for CRVS, and provided financial and technical support to several countries (AbouZahr et al., 2015b). However, these were discontinued due to concerns regarding sustainability and the lack of high-level commitment from governments (AbouZahr et al., 2015b). Setting up and maintaining CRVS systems requires substantial and sustained public investment that is often difficult to secure in LMICs, or superseded by more pressing priorities (Byass, 2009; Ye et al., 2012).

2.3.3 Health management information systems

In settings where a high proportion of births take place in health care facilities, health management information systems (HMIS) may serve as an alternative source of data on pregnancy outcomes. HMIS gather information on health status and determinants, and permit performance monitoring of healthcare systems (Mutale et al., 2013). HMIS have been established in several LMICs (Mutale et al., 2013). However, the data collected are often subject to quality and completeness issues, and adverse pregnancy outcomes are not universally recorded (UN IGME, 2020a). Additionally, up to one-third of births continue to occur outside of health facilities in sub-Saharan

Africa, meaning that data collected on pregnancy outcomes is far from comprehensive (Doctor, Nkhana-Salimu, and Abdulsalam-Anibilowo, 2018). Nevertheless, substantial investments into strengthening HMIS in sub-Saharan Africa (Warren et al., 2013; Hung et al., 2020) and an increasing trend in the proportion of facility births (Doctor, Nkhana-Salimu, and Abdulsalam-Anibilowo, 2018) mean that routine programme data and HMIS may be play an important role in improving monitoring of maternal and child health.

2.3.4 Maternity history questionnaires

Maternity history questionnaires collected in cross-sectional surveys have been heavily relied upon for monitoring reproductive and child health in sub-Saharan Africa. Beginning in the 1970s, the World Fertility Surveys (WFS) and Contraceptive Prevalence Surveys (CPS) were deployed to collect data on births, deaths, and contraceptive use (Akuze et al., 2021). Both sets of surveys were primarily funded by the United States Agency for International Development (USAID), with the latter steered by the Centers for Disease Control (CDC) (IHME, 2022). The main feature of these surveys was the collection of retrospective birth history questionnaires through interviews with women of reproductive age (Cleland and Verma, 1989). As the focus of the CPS was expanded include an array of maternal and child health issues, the surveys became known as the Maternal and Child Health/Family Planning Surveys and eventually, the Reproductive Health Surveys (IHME, 2022). The WFS evolved into the Demographic and Health Surveys (DHS) and became the chief source of maternity history data for LMICs (Akuze et al., 2021).

The DHS Program has conducted over 400 surveys in 90 countries since its inception in 1984 (DHS, 2022). In the mid-1990s, the Multiple Indicator Cluster Surveys (MICS), developed by the UNICEF, became another important source of maternity history data (Khan and Hancioglu, 2019). Close to 350 surveys have been conducted in 118 countries over the course of the program (MICS, 2022). The type of maternity history questionnaire employed by each platform has evolved over time (Akuze et al., 2021).

The DHS has primarily employed full birth history (FBH) and full pregnancy history (FPH) questionnaires (Espeut and Becker, 2015). In a FBH, women of reproductive age (15-49 years) are asked to retrospectively report on previous live births with information regarding date of birth, the survival status of children, and if applicable, age at death (Akuze et al., 2021). From this information, estimates of U5M can be calculated using the date of the death and the period of time that the child was exposed to the risk of dying. Estimates are typically calculated for no more than 15 years prior to the administration of any given survey to limit survivorship bias among respondents and recall error (Pullum and Becker, 2014). Furthermore, as questionnaires are

typically administered to women who are 15-49 years of age at the time of the survey, estimates of fertility or mortality for periods more than 15 years prior would not reflect the experience of women who were over 35 years of age at that time. This becomes increasingly consequential for estimates further back in time with the exclusion of women in peak childbearing ages.

The FPH differs from the FBH in that information is collected on all pregnancy outcomes, including pregnancy terminations, miscarriages, and stillbirths (Akuze et al., 2021). In both questionnaires, the order in which information is collected may be forward or backward; starting with the woman's first birth/pregnancy or her most recent, and continuing in the opposite direction (Espeut and Becker, 2015). The deployment of FPHs has not been widespread, though it was recently adopted as the standard maternity history module in the eighth phase of the DHS (DHS, 2019). Prior to this time, the DHS has primarily used reproductive calendars to collect information on pregnancy loss (Akuze et al., 2021).

In a summary birth history (SBH), women are asked to report on the number of children they have given birth to, the survival status of each child, and proxy information to calculate exposure to risk (usually the mother's age, duration of marriage, or time since first birth) (Silva, 2012). The proportion of children that have died are calculated for five-year reproductive age groups of the mothers (e.g., 15-19, 20-24,..., 45-49), and converted into probabilities of dying under age five using a series of adjustment factors corresponding to age patterns of fertility and mortality (Brass and Macrae, 1984). SBHs allow for the indirect estimation of mortality without reliance on exact reporting of dates of birth or ages at death. They include fewer and simpler questions, and require less interviewer training than FBHs (Mahy, 2003). The MICS survey program has employed both SBHs and FBHs.

2.3.5 Demographic surveillance

The use of demographic surveillance for health research in sub-Saharan Africa dates to the 1940s (Ngom, 2001). Activities were initially oriented towards field-testing and improving primary health care strategies, studying the epidemiology of tropical diseases and post-independence changes in socio-economic dynamics, and assessing the validity of demographic data collected in surveys and pilot CRVS systems (Clark et al., 1995; Garenne and Cantrelle, 1997; Tollman and Pick, 2002). Early health research stations conducting longitudinal surveillance on populations in South Africa and Senegal paved the way for other sites to be founded across the continent (Ngom, 2001). With time, sites gradually coalesced into networks in a systematic effort to address population health data gaps in LMICs (INDEPTH Network, 2002).

The largest group of affiliated HDSS is the International Network for the Demographic Evaluation of Populations and their Health (INDEPTH Network), which was founded in 1998, initially consisting of 17 sites from 13 countries in Africa and Asia

(INDEPTH Network, 2002; Sankoh and Byass, 2012). The INDEPTH Network has grown steadily and now includes close to 50 field sites (INDEPTH Network, 2021). The Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA Network) was established in 2005, comprised of sites located in eastern and southern Africa, many of which were also members of the INDEPTH Network (Maher et al., 2010; Reniers et al., 2016). The Child Health and Mortality Prevention Surveillance (CHAMPS) project is a relatively newer umbrella organization of seven HDSS sites which are specifically oriented towards generating an evidence base to inform policies for improving child health (Cunningham et al., 2019).

HDSS sites are geographically-defined, sub-national areas in which populations of several thousand individuals reside (Sankoh and Byass, 2012). The population is typically enumerated in a baseline census, and continuously followed-up for research purposes through regular household interviews (Sankoh and Byass, 2012). At each interview, detailed demographic information is collected on births, deaths, and in- and out-migrations that have occurred since the last data collection round. Household interview rounds are spaced at intervals ranging from every month to every year, depending on the research priorities and resource constraints of the site. Many sites utilize systems of proxy reporting where interviews are conducted with one household representative who reports on behalf of all household members (Kwon et al., 2021). This can reduce expenses, as it circumvents the need to make multiple visits to the same household to interview those who were absent at the first visit. It is also a necessary tool for collecting information on subjects who are too young to be interviewed or no longer present due to out-migration or death.

Quality issues affecting U5M estimates from HDSS and maternity history questionnaires

Data collected by HDSS and maternity history questionnaires are subject to several quality issues that affect the estimation of U5M. It is sufficient to mention some of these briefly, as they will be the subject of more detailed examination in the research papers included in this thesis.

The mostly commonly identified errors in maternity history questionnaires are attributable to the retrospective nature of reporting. Recall bias, such as when a child's date of birth or age at death are not remembered accurately or not known, can produce bias in estimates of mortality for certain age segments (Schoumaker, 2014). This often results in the rounding of reported ages or dates to preferred digits, which can distort measures of the age pattern of mortality (Hill and Choi, 2006; Pullum and Becker, 2014; Pullum and Staveteig, 2017). For example, ages at death which are rounded up and reportedly inexactly as one week, month, or year of life will be transferred between indicators for early neonatal, late neonatal, postneontal, and childhood mortality. However, the most serious issue affecting maternity history questionnaires is the

tendency for respondents to omit adverse outcomes such as stillbirth or early death (Espeut and Becker, 2015; Pullum and Becker, 2014). Misclassification of stillbirths and neonatal deaths has been identified as another source of potential error in mortality estimates from maternity histories, the magnitude of which is not well-understood (Helleringer et al., 2020; Liu et al., 2016).

Though the prospective design of HDSS overcomes some of the weaknesses of retrospective maternity histories, data collected on U5M are also subject to quality issues. The recurring data collection rounds that are so effective for monitoring established residents do not confer the same benefits to tracking new and transient individuals. Identifying recent arrivals, and maintaining a detailed record of in- and out-migrations is one of the most challenging aspects of HDSS data collection (Sankoh and Byass, 2012). Similar difficulties are encountered when collecting information on adverse pregnancy outcomes and births followed by early deaths. Events such as these, occurring between data collection rounds and lacking a pre-existing record to be followed-up, are vulnerable to under-reporting. To address this, some HDSS register pregnancies in order to prompt fieldworkers to follow-up on pregnancy outcomes in the next interview round. However, collecting information on pregnancy status is also uniquely challenging due to its sensitive nature (Kwesiga et al., 2021). The use of proxy respondents is another systematic weakness of HDSS. While women report on their own pregnancies, births, and children in reproductive history questionnaires; such information is often gathered from a household representative in HDSS.

There have now been several studies comparing U5M estimates from maternity history questionnaires with HDSS (Espeut and Becker, 2015; Deribew et al., 2016; Kadobera et al., 2017; Nareeba et al., 2021), one of which can be found in Chapter 3 of this thesis. The results of have been somewhat mixed, though it appears that U5M may be systematically underestimated by both sources—particularly at early ages in HDSS and later ones in estimates generated from maternity history questionnaires.

2.4 Global goal-setting

In the last few decades, the global reduction of U5M has become a central aim of the international community. In 2000, world leaders adopted the Millennium Development Goals (MDGs) which included the target of a two-thirds reduction in the global rate of U5M between 1990 and 2015 (MDG 4) (United Nations, 2000). This target was not met, although U5M was estimated to have fallen by 53% globally, and there were considerable reductions in regions with the highest burden (Molyneux and Molyneux, 2016). Figure 2.1 displays estimates of regional trends in U5M over the MDG era from the UN Inter-agency Group for Child Mortality Estimation (UN IGME) (You et al., 2015).

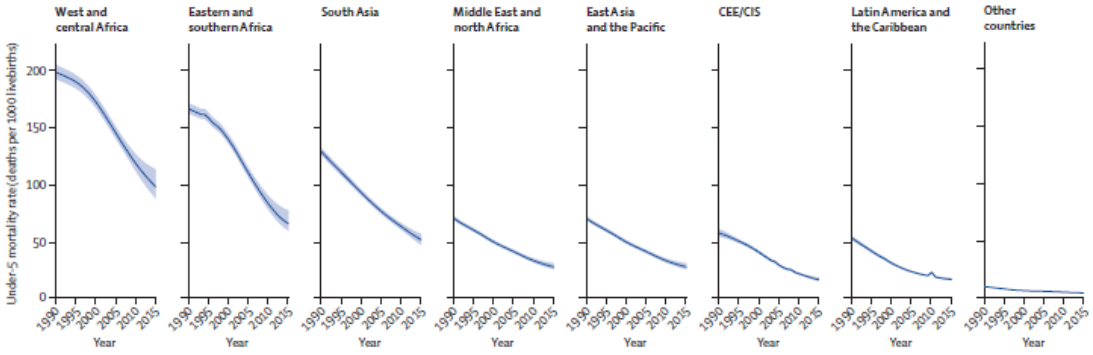


FIGURE 2.1: Trends in the under-5 mortality rate by global region from 1990-2015, with 90% uncertainty intervals. Source: You et al., 2015

The current UN IGME U5M estimation procedure involves a bias-adjusted model which was first used in the 2013 round of estimates (Alkema and New, 2014; UN IGME, 2013). In countries lacking high-quality vital statistics, the model is fit to mortality observations from nationally representative data sources. FBHs collected in DHS are the primary data underpinning the estimates for sub-Saharan African countries (Alkema and New, 2014). Each U5M observation is considered to be the true value of mortality multiplied by an error term, taking into account such factors as the standard error of the estimate, source type, and agreement with other estimates for the same period (UN IGME, 2021). For most countries in sub-Saharan Africa, a life table is used to derive estimates of infant mortality from the level of U5M. Neonatal mortality is estimated with a separate but similar model (Alexander and Alkema, 2018). The UN IGME estimates are considered the most comprehensive available, however, the insufficient quality and quantity of data inputs for sub-Saharan Africa is a barrier that is hard to overcome (Alkema and New, 2014). Large heterogeneity in the quality of FBH data in different surveys and estimation periods makes it difficult to accurately adjust for bias in U5M estimates (Alkema and New, 2014). UN IGME estimates have been key to tracking U5M globally, though it is also important to highlight the uncertainty of such estimates for settings which lack high-quality CRVS data (Alkema and New, 2014).

Though only ten countries achieved MDG 4 in sub-Saharan Africa[2], overall reductions in U5M were substantial (You et al., 2015). It has also been noted that classifying countries as either having achieved or missed MDG 4 could be misrepresentative. Measuring the decline in U5M over the period 1990-2015 was a somewhat arbitrary choice, and achieving a relative reduction of two-thirds was generally more difficult for countries starting with lower levels of U5M (Byass et al., 2015). South Africa missed the target, though started and ended the period with substantially lower levels of U5M than sub-Saharan Africa as a whole (Byass et al., 2015). It is thus in some

[2]The countries in sub-Saharan Africa which achieved MDG 4 were Ethiopia, Eritrea, Liberia, Madagascar, Malawi, Mozambique, Niger, Rwanda, Uganda, and Tanzania (You et al., 2015).

ways more informative to consider that 90% of countries in sub-Saharan Africa increased their rate of U5M reduction from the 1990s to 2000-2015, with 21 countries either tripling the rate, or reversing an increasing trend in U5M (You et al., 2015).

However, much work remained to be done. Stark regional disparities persisted, with sub-Saharan Africa continuing to have the highest rates of U5M in the world. Growth in the total number of births in the region threatened to outpace U5M reductions and lead to stagnation or increases in the total number of under-5 deaths (You et al., 2015). Additionally, declines in mortality had proceeded at a faster pace for later ages under five compared to the neonatal period, and especially the first week of life (Liu, Hill, and Oza, 2016; UN IGME, 2015; Wang et al., 2014). Though not included in the MDG 4 target, progress reducing stillbirths was the slowest of all (Liu, Hill, and Oza, 2016).

Overall declines in the level of U5M tends to coincide with the compression of risk into early ages (Lawn, Cousens, and Zupan, 2005). The intrapartum period and first few days of life are the riskiest time for human survival (Oza, Cousens, and Lawn, 2014). Causes of mortality during this period can also be harder to address compared to those afflicting later childhood ages. For instance, intrapartum complications and congenital abnormalities require high-quality obstetric and neonatal care that is not widely available in LMICs (Lawn et al., 2016; Wang et al., 2016). However, a substantial proportion of neonatal deaths in LMICs could be prevented through relatively straightforward interventions related to antenatal care, maternal health and education, and nutrition (Hug et al., 2019; Lawn, Cousens, and Zupan, 2005; Oza et al., 2015). Higher coverage of antenatal care and facility births are also strongly associated with lower rates of stillbirth (Lawn et al., 2016). Further reductions to U5M will depend on making gains at early ages, and improving access to high-quality care during pregnancy, delivery, and the postnatal period (Hug et al., 2019)

Adopted in 2015, the Sustainable Development Goals (SDGs) provided updated targets for 2030. SDG 3.2 calls for reducing neonatal and U5M to at least as low as 12 deaths per 1,000 live births and 25 per 1,000, respectively (United Nations, 2017). The Early Newborn Action Plan (ENAP), launched through a World Health Assembly resolution, set a target of 10 or less stillbirths per 1,000 total births by 2035 (World Health Organization, 2014).

As of the latest available data, approximately three-quarters of countries in sub-Saharan Africa were not on track to achieve the SDG child health targets (UN IGME, 2021). Countries in west and central Africa are the most in need of accelerating progress (You et al., 2015). The rate of reduction of stillbirths for the period 2010-2019 in sub-Saharan Africa remained among the lowest of any global region, and had not increased from the prior decade (UN IGME, 2020a).

There have been substantial improvements to newborn and child health in sub-Saharan Africa since the beginning of global monitoring, though progress must be

accelerated to save lives and reduce global inequities. Accurate measurement of levels, trends, and detailed age patterns of U5M will be essential to meeting this challenge.

Chapter 3

Age patterns of under-5 mortality in sub-Saharan Africa during 1990-2018: A comparison of estimates from demographic surveillance with full birth histories and the historic record

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1600573	**Title**	Ms
First Name(s)	Hallie		
Surname/Family Name	Eilerts		
Thesis Title	Under-five mortality estimates for sub-Saharan Africa: an inquiry into data sources and estimation methods		
Primary Supervisor	Georges Reniers		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Demographic Research		
When was the work published?	04 March 2021		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	NA		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the study in collaboration with my supervisors: Georges Reniers, Jeffrey W. Eaton, and Julio Romero Prieto. I conducted all data analysis, interpreted findings, and wrote the first draft of the manuscript. My supervisors provided revisions to the analysis and manuscript which I incorporated in the final version.

SECTION E

Student Signature	Hallie Eilerts
Date	21 March 2022

Supervisor Signature	Georges Reniers
Date	21 March 2022

3.1 Abstract

BACKGROUND
Estimates of under-5 mortality (U5M) for sub-Saharan African populations often rely heavily on full birth histories (FBHs) collected in surveys and model age patterns of mortality calibrated against vital statistics from other populations. Health and Demographic Surveillance Systems (HDSS) are alternate sources of population-based data in much of sub-Saharan Africa, which are less formally utilized in estimation.
OBJECTIVE
In this study, we compare the age pattern of U5M in different African data sources (HDSS, Demographic and Health Surveys (DHS), and Multiple Indicator Cluster Surveys (MICS)), and contrast these with the historical record as summarized in the Human Mortality Database and model age patterns.
METHODS
We examined the relative levels of neonatal, postneonatal, infant, and child mortality across data sources. We directly compared estimates for DHS and MICS subnational regions with HDSS, and used linear regression to identify data and contextual attributes that correlated with the disparity between estimates.
RESULTS
HDSS and FBH data suggests that African populations have higher levels of child mortality and lower infant mortality than the historic record. This age pattern is most explicit for Western African populations, but also characterizes data for other subregions. The comparison between HDSS and FBH data suggests that FBH estimates of child mortality are biased downward. The comparison is less conclusive for neonatal and infant mortality.
CONTRIBUTION
This study questions the practice of using model age patterns derived from largely high-income settings for inferring or correcting U5M estimates for African populations. It also highlights the considerable uncertainty around the consistency of HDSS and FBH estimates of U5M.

3.2 Introduction

Ending preventable deaths of newborns and children under five years of age by 2030 is a central aim of the international community, and codified in the United Nations' Sustainable Development Goal 3.2 (United Nations, 2017). Accurate measurement of under-5 mortality (U5M) is essential to tracking and accelerating progress towards its reduction. Civil registration and vital statistics (CRVS) systems are the preferred sources for such data. However, in sub-Saharan Africa, the region with the highest U5M in the world, CRVS are often incomplete. For sub-Saharan Africa, estimates of U5M mostly rely on retrospective reporting of child survival in full birth histories (FBHs) collected in sample surveys and model age patterns of mortality (UN IGME, 2018). However, biases in FBHs and uncertainty regarding the age pattern of U5M in sub-Saharan Africa give reason to regard such estimates with caution (Pullum and Becker, 2014; Schoumaker, 2014). The lack of reliable population-based data means there are few sources to validate these estimates against. Health and Demographic Surveillance Systems (HDSS) are designed to fill these data gaps.

HDSS collect longitudinal data through closely spaced interviews of a geographically defined population. They are valuable sources of detailed, empirical data in low-resource settings across sub-Saharan Africa. However, the recording of vital events among individuals (including newborns) who join households between interview rounds can be incomplete, and deteriorate with less frequent rounds (Sankoh and Byass, 2012).

In this contribution, we systematically compare the level and age pattern of U5M in HDSS with FBH estimates from surveys, and the historic record of high-quality CRVS data. A direct comparison between African data sources is used to investigate whether attributes of data collection, context, or quality can explain deviations from traditional age patterns of U5M. We use this approach to investigate the extent to which deviations from the historic record are driven by epidemiological factors or data quality issues. We find evidence that the age pattern of child mortality in sub-Saharan Africa deviates from the historical record assembled largely from CVRS in high-income settings. We also highlight the uncertainty in U5M estimation for sub-Saharan Africa, as well as the value to be gained from systematic comparison between HDSS and other data sources.

3.3 Methods

3.3.1 Data

Our analysis makes use of data from HDSS, FBHs in sample surveys, and high-quality CRVS.

HDSS data was obtained from the International Network for the Demographic Evaluation of Populations and their Health (INDEPTH Network) Data Repository[1]. We included data from 30 studies in 13 African countries spanning 1990 to 2017.

The largest sources of FBHs for sub-Saharan Africa are the Demographic and Health Surveys (DHS) and UNICEF Multiple Indicator Cluster Surveys (MICS). We included data from 62 DHS conducted between 1995 and 2018 in 33 countries. For MICS, we included FBHs from 26 surveys conducted between 2010 and 2017 in 20 African countries. Data was procured from the DHS data repository[2] and UNICEF MICS surveys webpage[3].

Data from countries with high-quality CRVS was obtained from the Human Mortality Database (HMD)[4]. To improve the comparison with the Sub-Saharan African data, we restricted data from the HMD to high-mortality contexts in which the level of U5M was between the 10th and 90th decile of DHS estimates. This included 99 life tables from 24 mostly European countries. Life tables which did not cover national populations were excluded (e.g. England & Wales Civilian Population, France Civilian Population, New Zealand Maori, New Zealand Non-Maori). Life tables for the total populations were excluded in favor of life tables for sub-populations with much longer time series (e.g. East Germany and West Germany were used instead of Germany, Total Population). The life table for Taiwan was excluded due to documented under-registration of infant deaths (Chen et al., 1998b).

We used a model to summarize observed regularities in the age pattern of U5M in populations with high-quality data. The model was based on a new database of CRVS data compiled from historical yearbooks and the United Nations Statistical Division, and estimated using Wilmoth's approach of flexible patterns of mortality (Wilmoth et al., 2012). It is a two-dimensional model with entry parameters for the overall level of U5M and the shape of the age pattern of mortality. More detail on the model is available elsewhere (Guillot et al., 2020).

3.3.2 Analysis

The analysis was conducted in three parts. First, we examined the age pattern of U5M in each data source. This consisted of a comparison of the relative levels of mortality for standard age breakdowns within the under-5 range. Next, we compared the absolute levels of mortality in HDSS with subnational region estimates from DHS and MICS. In the final part of the analysis, we used ordinary least squares (OLS) regression to identify fieldwork practices, contextual attributes, and indicators of data quality that correlated with the relative disparity between HDSS and DHS estimates.

[1] http://www.indepth-ishare.org
[2] https://dhsprogram.com/data/available-datasets.cfm
[3] http://mics.unicef.org/surveys
[4] https://www.mortality.org/

Age pattern of U5M

We compared the distribution of mortality under age 5 in each data source using the neonatal, postneonatal, infant, and child age groups. Neonatal mortality refers to the probability of dying in the first 28 days of life, and is denoted as $q(28d)$. Postneonatal mortality is the probability of dying before age 1, conditional on surviving the first month of life ($q(28d, 1y)$). Infant mortality is the probability of dying between birth and exact age 1 ($q(1y)$), and child mortality is the probability that a child aged 1 year will die prior to reaching age 5 ($q(1y, 5y)$).

In order to calculate mortality estimates from HDSS data, we used individual-level records with exact dates of births, deaths, migrations, and censoring to assign events and exposure time to 5-year periods extending backwards from the most recent year of available data. Within each period, events and exposure time were further disaggregated into age groups consisting of segments of 0, 1-2, 3-4, 5-11, 12-23, 24-35, 36-47, and 48-59 completed months. We performed a standard demographic calculation of observed deaths over person-years to calculate the mortality rate for each age group. Piecewise-constant rates were aggregated over age groups to produce the cumulative hazard of mortality between the ages 0 and 1 month, 1 month and 1 year, 0 and 1 year, and 1 to 5 years. We converted mortality rates into probabilities of dying $q(28d)$, $q(28d, 1y)$, $q(1y)$, and $q(1y, 5y)$.

It would be preferable to begin exposure time when births or in-migrations were first observed by the HDSS (the observation date), rather than the date of the events themselves. This is due to under-ascertainment of left-censored vital events, where the vital events of individuals that do not survive to the first HDSS round are less likely to be reported. However, dates of observation have not been universally recorded across sites and time. As a sensitivity analysis, we recalculated estimates of mortality with exposure starting at observation date when such data was available, and compared to the standard estimates.

Estimates of mortality from DHS and MICS surveys were calculated using the retrospectively reported information of FBHs. We used FBHs to assign births, deaths, and exposure time to age groups in the 0-4 and 5-9 years prior to each survey. We calculated national-level estimates of $q(28d)$, $q(28d, 1y)$, $q(1y)$, and $q(1y, 5y)$ for each 5-year period following the same methods as described for HDSS data. For MICS surveys that were not conducted at the national-level, estimates were calculated for the entire sub-population included in the survey. For the HMD, estimates of national-level $q(1y)$ and $q(1y, 5y)$ were available in 5-year period abridged life tables.

The age pattern of U5M in each source was evaluated in a descriptive manner. We limited estimates for African contexts to periods between 2000 and 2018. This was done to avoid comparing the age pattern of mortality across vastly different levels of mortality. The overall level of mortality plays an important role in determining the age pattern, though the exact nature of this relationship has been difficult to establish for

countries in sub-Saharan Africa (Blacker, Hill, and Timaeus, 1985; Guillot et al., 2012). The exclusion of data prior to 2000 was also justified by the lack of MICS FBHs prior to this time. We examined the distribution of mortality under age 5 by plotting estimates of infant mortality against child mortality. For mortality under age 1, the distribution was further stratified by a scatter plot of neonatal and postneonatal mortality. We also plotted the ratio of infant to child mortality against under-5 mortality, and the ratio of postneonatal to neonatal mortality against infant mortality.

We included a model prediction on the scatter plots displaying the distribution of mortality under 1 and 5 in each data source. The model is a log-quadratic representation of the cumulative probability of dying as a function of the level of U5M and a continuous parameter modifying the shape of the pattern of mortality (Guillot et al., 2020). We calculated the model predictions for postneonatal mortality using predefined values of neonatal mortality (that covered the observed range of DHS estimates), and five different shape parameters. With these inputs, we used to the model to solve for the level of U5M and indirectly estimate infant mortality. We then subtracted the predefined values of neonatal mortality from the model's infant mortality estimates to produce estimates of postneonatal mortality. The shape parameters were selected to produce estimates of the model's central tendency, inner, and outer bounds. The bounds represent values of postneonatal mortality corresponding to different patterns of mortality, with the outer bounds representing particularly extreme age patterns. We used the same procedure to indirectly estimate child mortality using predefined values of infant mortality.

We summarized the ratios of child to infant and postneonatal to neonatal mortality in HDSS, FBHs, and the HMD with boxplots. We further disaggregated the ratios by African subregion (Southern, Eastern, Central, Western) using the United Nations geoscheme, and compared them using median, interquartile range (IQR), and non-parametric tests of differences in median (with the DHS median as the reference). We further stratified the Western region to analyse whether the ratios differed for Coastal and Sahelian countries.

HDSS and FBH differences

We then conducted a direct comparison of HDSS and FBH mortality estimates. We recalculated mortality estimates from DHS and MICS at the subnational region level for the 0-4 and 5-9 years prior to all surveys administered between 1995 and 2018. HDSS were mapped to the DHS and MICS subnational regions using GPS coordinates for each site's approximate location, taking into account changes in survey region names and boundaries over time. We calculated HDSS mortality estimates for the same 5-year periods as the FBH estimates. We calculated the relative percentage differences between HDSS and FBH estimates by subtracting the FBH values from the HDSS, dividing by the average of the two values, and multiplying the quotient by 100. We

examined the distribution of differences between sources in boxplots.

Factors associated with HDSS-DHS differences

Finally, we used OLS regression to identify predictors of the size of the discrepancy between HDSS and DHS subnational region estimates. The relative percentage differences served as the response variable in linear regression models with covariates related to HDSS and DHS fieldwork practices, contextual attributes, and indicators of data quality. We used bivariate models to investigate associations with relative differences in neonatal, postneonatal, infant, and child mortality. We added covariates which were strongly associated with the response variables to multivariable linear regression models. The comparison between HDSS and MICS subnational region estimates was not included in this part of the analysis due to having too few overlapping observations for analysis.

As covariates related to HDSS fieldwork practices, we included the interval between HDSS rounds (in months), and whether the site conducts pregnancy reporting. Pregnancy reporting refers to a system of collecting information on pregnancy status for female residents of reproductive age during each regular data collection round, and using this information to follow-up on pregnancy outcomes in subsequent rounds. We included HDSS digit preference for day of death reporting as a data quality covariate. We measured digit preference during the 5-year periods of observation using the Index of Dissimilarity, which compares the observed and expected distributions of day of death reporting for calendar days 1-31. The expected distribution of reports was taken from the frequency of each day throughout the year (12/365 for days 1-28, 11/365 for days 29 and 30, and 7/365 for day 31). We calculated the index by summing the absolute deviation of the observed distribution of reports from the expected distribution and dividing by two. We also included the log value of the average annual births in the HDSS as a contextual attribute of interest.

For DHS data attributes, we included markers of data quality, adult mortality, and the "time prior to the survey" (TIPS). We calculated the data quality covariate by combining indicators for age heaping, date of birth displacement, and date of birth incompleteness in DHS FBHs. These indicators were calculated for each survey at the subnational region level following methods described elsewhere (Pullum and Staveteig, 2017). We calculated adult mortality as the probability of dying between ages 15 and 50 using data from DHS sibling survival modules. This covariate was used as a proxy variable for potential bias in U5M estimates from missing mothers in high-HIV settings. The TIPS covariate was a dummy variable indicating whether the mortality estimate was for the period 0-4 or 5-9 years prior to the survey. In DHS FBHs, health information is collected on children born during the period 0-4 years prior to the administration of survey, and DHS fieldworkers sometimes displace events outside of

this window in order to reduce their workload (Pullum and Becker, 2014). We used this covariate to investigate the effect of such transfers on mortality estimates.

We included malaria endemicity as a contextual attribute of interest. This was calculated as the plasmodium falciparum parasite rate in 2-10 year olds at the approximate GPS location of each HDSS using data from the Malaria Atlas Project's (MAP) global database.[5] We also included African subregion of the HDSS/DHS as a covariate.

All analyses were done in R version 3.6.1. We used the *rdhs* and *demogsurv* R packages to assist with data procurement and calculations (MRC Centre for Global Infectious Disease Analysis, 2020; Watson, FitzJohn, and Eaton, 2019).

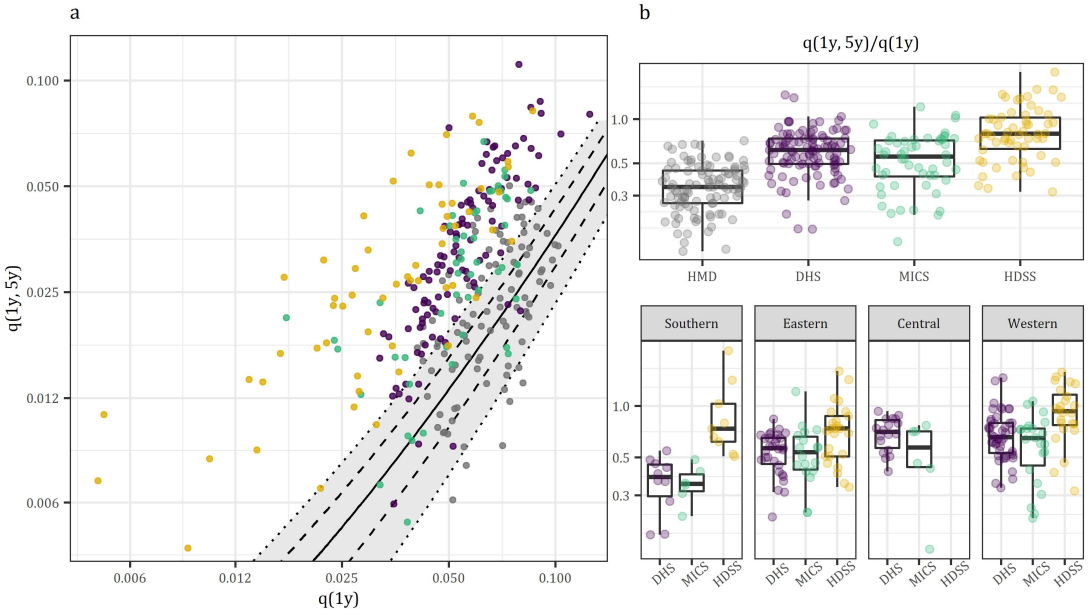
3.4 Results

The relationship between infant and child mortality in each data source is displayed in Figure 3.1. In Figure 3.1a, the relationship between infant ($q(1y)$) and child ($q(1y, 5y)$) mortality in the HMD is similar to that which is portrayed with the model. However, the pattern is much different for African data sources. Data from HDSS, DHS, and MICS have higher levels of child mortality for given levels of infant mortality than the historic record and model predictions. This was consistent across different levels of U5M (see appendix Figure B.1a).

Figure 3.1b displays the same data as panel A, but in the form of the ratio of child to infant mortality. The median ratios of child to infant mortality from African estimates were substantially higher than the HMD median of 0.34. The ratio was the highest in HDSS (0.79), followed by DHS (0.62) and MICS (0.56). When examined by African subregion, the HDSS ratio in Southern Africa was the most different from the FBH data ($p<0.01$). The highest median ratios for HDSS and MICS were in Western Africa (0.93 and 0.65, respectively). A breakdown of the western region found that this was mostly driven by HDSS in Sahelian countries rather than Coastal (see appendix Figure B.2a). The highest median ratios for DHS were in the central and western regions (0.70 and 0.66, respectively). Full results for the ratios of child to infant mortality are shown in appendix Table B.1.

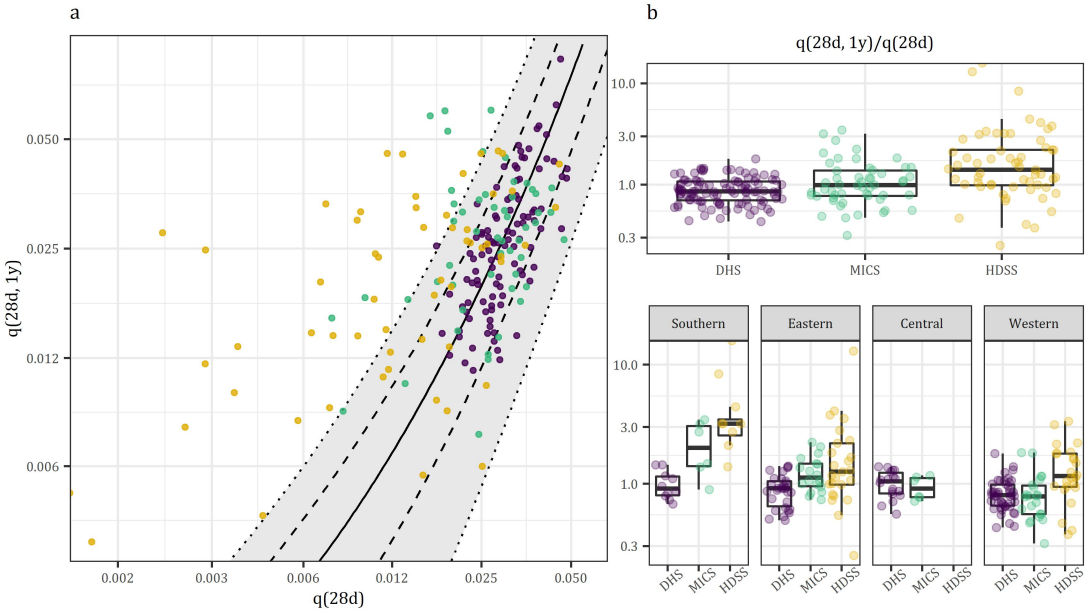
Figure 3.2 displays the relationship of neonatal ($q(28d)$) and postneonatal ($q(28d, 1y)$) mortality. In Figure 3.2a, the levels of neonatal mortality in HDSS were strikingly low compared to MICS, DHS, and modelled estimates. This resulted in HDSS having mostly lower overall mortality in the infant period than other sources, but extremely high ratios of postneonatal to neonatal mortality (see appendix Figure B.1b). The relationship between neonatal and postneonatal mortality in FBHs was consistent with modelled patterns from high quality CRVS, particularly for DHS.

[5] https://map.ox.ac.uk/



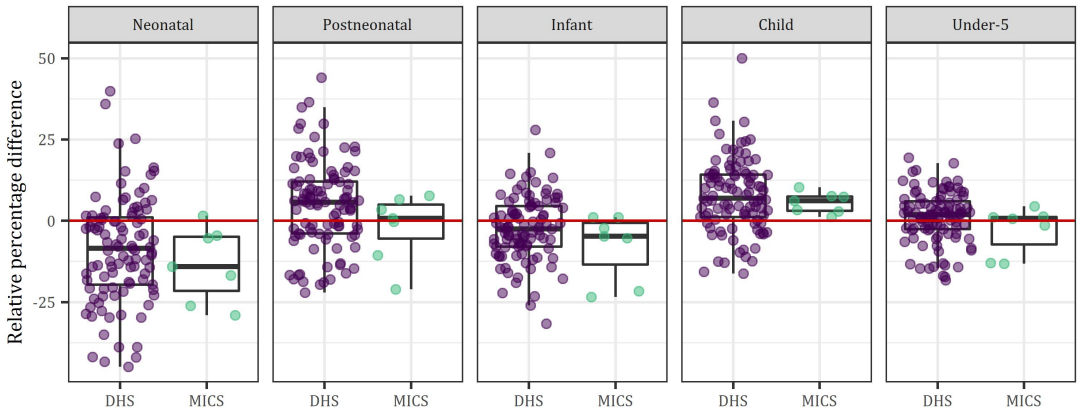
Note: a: Scatter plot showing the relationship between infant and child mortality in HDSS, DHS, and MICS from sub-Saharan Africa, and high-mortality observations from the HMD. The black solid line is the model estimate of the relationship between the two indicators, derived from high-quality CRVS data. Dashed and dotted lines indicate model predictions under different age patterns of U5M. b: Box plots showing the ratio of child to infant mortality in each source, and disaggregated by African subregion.

FIGURE 3.1: Age pattern of mortality under five years.



Note: a: Scatter plot showing the relationship between neonatal and postneonatal mortality in HDSS, DHS, and MICS from sub-Saharan Africa. The black solid line is the model estimate of the relationship between the two indicators, derived from high-quality CRVS data. Dashed and dotted lines indicate model predictions under different age patterns of U5M. b: Box plots showing the ratio of postneonatal to neonatal mortality in each data source, and disaggregated by African subregion.

FIGURE 3.2: Age pattern of mortality under one year.



Note: Relative percentage differences between HDSS and DHS/MICS subnational region mortality estimates for overlapping 5-year periods.

FIGURE 3.3: HDSS-FBH differences.

Figure 3.2b displays the ratio of postneonatal to neonatal mortality in each data source, and disaggregated by African subregion. The HDSS data had the highest median ratio of 1.32, compared to the MICS median of 0.96 and DHS of 0.87. HDSS ratios were also generally higher when compared to FBHs from the same African subregions.

In the southern region, the MICS median was double that observed in the DHS (2.11 compared to 0.92, p=0.01). HDSS in the southern region had the highest median ratio, with the level of postneonatal mortality being 2.99 times that of neonatal mortality. In the western region, more than 75% of HDSS ratios were higher than the DHS median ($p<0.01$). For HDSS and MICS, there was not a clear difference in the ratio of postneonatal to neonatal mortality for Coastal and Sahelian West African countries (see appendix Figure B.2b). For DHS, postneonatal mortality was relatively higher than neonatal mortality in Coastal countries. The full results for ratios of postneonatal to neonatal mortality are shown in appendix Table B.2.

Subnational regions from DHS and MICS were matched with HDSS located within the regions to directly compare U5M estimates. There were 101 overlapping 5-year period estimates for HDSS and DHS, and 7 for HDSS and MICS. This included data from 24 HDSS sites, 31 DHS surveys, and 3 MICS surveys. Figure 3.3 shows the distribution of relative percentage differences between HDSS and FBH estimates of neonatal, postneonatal, infant, child, and under-5 mortality. Positive values indicate higher estimates from the HDSS, while negative values indicate higher estimates from DHS or MICS.

HDSS estimates of neonatal mortality were typically lower than DHS and MICS. The median HDSS value was 7.13% lower than the corresponding DHS subnational region estimate, and 12.80% lower than the MICS estimate. The reverse was the case for postneonatal mortality: the DHS and MICS medians were 4.45% and 0.56% higher than HDSS, respectively. When neonatal and postneonatal ages were combined for

estimates of infant mortality, the HDSS median was 2.32% lower than the DHS, and 4.59% lower than MICS. For child mortality, the median HDSS estimate was 6.86% higher than the DHS and 6.14% higher than MICS. Altogether, HDSS produced U5M estimates that were roughly 1.96% higher than DHS and 0.73% higher than MICS.

We calculated the HDSS estimates using individual-level data on deaths and exposure, where individuals contributed exposure time starting from birth or in-migration to the HDSS. For a subset of studies where data was available, we also calculated estimates with exposure time beginning at the date that these events were observed by the HDSS, rather than the dates of the events themselves. Overall, estimates that used exposure time since observation date were lower for neonatal mortality, and slightly higher for postneonatal and child mortality (see appendix Figure B.3). This reinforces the above findings of HDSS-DHS differences in mortality under-5, and suggests that the gap between estimates from each source would increase if HDSS estimates used exposure time since observation date.

Table 3.1 presents the regression models for relative percentage differences between HDSS and DHS subnational region estimates. In the bivariate models for neonatal mortality, the covariates of HDSS pregnancy reporting and African region were strongly associated with the response variable. In HDSS that use information on pregnancies to follow up on pregnancy outcomes, estimates of neonatal mortality were roughly 15% higher, relative to DHS estimates. While this association disappears in the multivariable model, the effect for Southern African region becomes even stronger. Treating Eastern Africa as the reference category, estimates of neonatal mortality from HDSS in Southern Africa were approximately 36% lower than DHS. This also affects infant mortality estimation, where HDSS from Southern Africa had estimates that were approximately 15% lower.

TABLE 3.1: Linear regression on relative differences in HDSS and DHS subnational region estimates of mortality under five years.

	Dependent variable: Relative percentage difference in HDSS and DHS mortality estimates							
	Neonatal		**Postneonatal**		**Infant**		**Child**	
	Bivariate models (SE); p-value	Multivariable model (SE); p-value	Bivariate models (SE); p-value	Multivariable model (SE); p-value	Bivariate models (SE); p-value	Multivariable covariate (SE); p-value	Bivariate models (SE); p-value	Multivariable covariate (SE); p-value
HDSS avg. births	0.08 (1.51); p=0.96	2.97 (2.51); p=0.24	5.36 (1.14); p<0.01	2.50 (1.99); p=0.21	2.31 (0.89); p=0.01	1.59 (1.34); p=0.24	3.09 (1.02); p<0.01	1.60 (2.09); p=0.45
HDSS digit preference	8.70 (7.08); p=0.22	−19.58 (9.64); p=0.046	13.03 (5.82); p=0.03	26.56 (7.64); p<0.01	10.60 (4.21); p=0.01	5.67 (5.14); p=0.27	−5.50 (4.97); p=0.27	−0.70 (8.02); p=0.93
HDSS interview interval	−0.01 (0.40); p=0.99	1.71 (0.73); p=0.02	−1.06 (0.32); p<0.01	−1.13 (0.58); p=0.05	−0.39 (0.24); p=0.10	0.21 (0.39); p=0.58	−0.80 (0.27); p<0.01	−0.74 (0.61); p=0.23
HDSS preg. reporting	14.93 (4.44); p<0.01	3.78 (5.58); p=0.50	0.82 (3.92); p=0.83	0.99 (4.42); p=0.82	6.48 (2.78); p=0.02	1.49 (2.97); p=0.62	−2.51 (3.28); p=0.45	−0.21 (4.64); p=0.96
DHS adult mortality	−51.69 (36.81); p=0.17		−9.06 (25.33); p=0.72		−21.52 (20.73); p=0.30		2.89 (22.98); p=0.90	
DHS data quality	−6.50 (3.49); p=0.06	1.71 (3.93); p=0.67	−6.36 (2.90); p=0.03	−6.68 (3.11); p=0.03	−5.40 (2.09); p=0.01	−2.25 (2.10); p=0.29	−4.88 (2.44); p=0.05	−6.44 (3.27); p=0.05
DHS TIPS (ref. = 0-4 years)								
5-9 years	2.45 (3.27); p=0.45		−2.82 (2.72); p=0.30		−0.25 (1.99); p=0.90		−3.44 (2.27); p=0.13	
Malaria endemicity	7.07 (8.30); p=0.40	−0.59 (8.49); p=0.94	17.82 (6.53); p=0.01	11.78 (6.72); p=0.08	12.82 (4.33); p<0.01	7.25 (4.53); p=0.11	3.27 (6.21); p=0.60	1.56 (7.07); p=0.83
African region (ref. = Eastern)								
Southern	−26.94 (6.05); p<0.01	−35.83 (8.24); p<0.01	−11.61 (5.58); p=0.04	1.96 (6.52); p=0.77	−18.15 (3.68); p<0.01	−15.34 (4.40); p<0.01	0.91 (4.78); p=0.85	5.96 (6.86); p=0.39
Western	3.04 (3.15); p=0.34	2.40 (4.81); p=0.62	−2.09 (2.91); p=0.47	−4.71 (3.81); p=0.22	−0.24 (1.92); p=0.90	−2.77 (2.57); p=0.28	−1.27 (2.50); p=0.61	0.86 (4.01); p=0.83
Constant	−7.86 (1.61); p<0.01	−36.41 (20.85); p=0.08	4.14 (1.35); p<0.01	−13.92 (16.49); p=0.40	−2.24 (0.98); p=0.02	−16.94 (11.11); p=0.13	7.75 (1.13); p<0.01	2.64 (17.33); p=0.88
Observations		83		83		83		83
R^2		0.32		0.36		0.35		0.15
Adjusted R^2		0.25		0.29		0.28		0.06
Residual Std. Error (df = 74)		13.89		11.00		7.41		11.56
F Statistic (df = 8; 74)		4.35 (p<0.01)		5.28 (p<0.01)		5.00 (p<0.01)		1.67 (p=0.12)

Controlling for all other covariates in the models, the association of HDSS digit preference was in opposite directions for neonatal and postneonatal mortality. Increased digit preference was associated with lower levels of HDSS neonatal mortality, and higher levels of postneonatal mortality. In the bivariate models, the coefficient of HDSS interview interval was strongly associated with differences in postneonatal and child mortality. HDSS with longer interview intervals, and thus less frequent interviews, typically had lower estimates of postneonatal and child mortality relative to the DHS. The effect was weaker in multivariable models, but indicated a 1% decrease in HDSS estimates of postneonatal mortality for each one month increase in the time between interview rounds.

The coefficient for DHS data quality was negative and moderately associated with the response variable in each bivariate model, as well as the multivariable models for postneonatal and child mortality. For each one unit improvement in DHS data quality, DHS estimates increased relative to HDSS by 6-7%. Malaria endemicity was associated with higher HDSS estimates of postneonatal and infant mortality by approximately 18% and 13% respectively. Many HDSS have been set up in areas known to have high malaria transmission (Deribew et al., 2016), and therefore may have higher mortality than the surrounding DHS subnational region. However, the effect of malaria on HDSS-DHS differences was not as strong in the multivariable models. The number of average annual births taking place in HDSS was not associated with differences with DHS in any of the multivariable models. The covariates for DHS TIPS and DHS adult mortality were not included in the multivariable models, as they were not strongly associated with the response variable in any of the bivariate models.

The R^2 value for the child mortality model was lower than that of other models. This indicates that the covariates explained less of the variance observed in HDSS-DHS differences in child mortality. There were stronger associations with the covariates in the neonatal, postneonatal, and infant mortality models, which accounted for roughly one third of the variation in HDSS-DHS differences.

3.5 Discussion

Contemporary data sources for U5M in sub-Saharan are all suggestive of an age pattern that deviates considerably from the record of high-quality CRVS data from (mostly) high income countries. This atypical age pattern, characterized by higher levels of child mortality relative to infant or U5M, has previously been described for Western African populations (Abdullah et al., 2007; Billewicz and McGregor, 1981; Cantrelle, 1969; Garenne, 1981; McGregor, Billewicz, and Thomson, 1961; Pison and Langaney, 1985). Our analyses suggest that this phenomenon is not exclusive to Western Africa, but also characterizes the estimates for other African regions, albeit to a lesser extent.

Various factors could contribute to a different age pattern of mortality in African

populations. Infectious diseases like HIV, malaria, measles, and diarrhoea create an environment with relatively high levels of child mortality (Guillot et al., 2012). Mortality from these causes is typically lower in the first six months of life because children are not exposed or acquire passive immunity from their mother through breast milk. Once breastfeeding ends, however, this layer of protection wanes and mortality from these causes increases (Cantrelle and Leridon, 1971; Garenne, 1981). These factors combined with inadequate diet generate excess mortality at later ages, and thus an older age pattern of U5M (Guillot et al., 2012).

It remains, however, possible that this conclusion is compromised by data quality issues for data sources in sub-Saharan Africa (i.e. HDSS and FBHs). In the absence of well-functioning CRVS there is no gold standard measurement of U5M for many low- and middle-income populations. FBHs collected in periodic surveys have for long served as a satisfying stopgap, and they continue to be the primary data source for worldwide estimates of U5M (Alkema and New, 2014; Wang et al., 2017). FBHs are, however, subject to data quality issues that give reason to regard such estimates with caution (Dwyer-Lindgren et al., 2013; Pullum and Becker, 2014; Schoumaker, 2014).

The omission of births or deaths from FBHs is a serious issue that occurs more often when the child dies at an early age (Pullum et al., 2013). In surveys where recent births or deaths are subjected to more detailed data collection, there is evidence that data collectors displace births or deaths outside of this reporting window in order to reduce their workload (Schoumaker, 2011). There is also a tendency for deaths to be disproportionately reported as occurring at certain central ages such as 1 month and 12 months (Pullum et al., 2013; Pullum and Staveteig, 2017). If such heaping was primarily the result of upward rounding, this practice could distort FBH estimates of neonatal, postneonatal, and child mortality, and thus the age pattern of U5M.

The quality of early childhood mortality data in HDSS is also questionable. Some of the earliest research on age patterns of U5M in West Africa called attention to the issue of under-registration of early deaths in HDSS (Cantrelle, 1969; McGregor et al., 1979). The severity of this underestimation was thought to be reduced by the follow-up of pregnancies (Cantrelle and Leridon, 1971), however the imperfect and incomplete nature of this information meant that early deaths were still likely undercounted (Garenne, 1981). In our multivariable regression analysis, conducting pregnancy reporting was not strongly associated with higher estimates of neonatal mortality (relative to DHS). This is perhaps due to the heterogeneity in pregnancy reporting completeness across HDSS (Waiswa et al., 2019). There are also procedural inconsistencies regarding the use of proxy respondents and male interviewers, which have been shown to negatively impact pregnancy reporting completeness (Kadobera et al., 2017). Collecting information on pregnancies and reliably detecting neonatal deaths remain some of the most challenging issues in HDSS (Sankoh and Byass, 2012).

It is thus not surprising that HDSS estimates of neonatal mortality were approximately 7% and 13% lower than subnational estimates from DHS and MICS, respectively. The DHS and MICS early mortality estimates appear more realistic, with ratios of postneonatal to neonatal mortality that were consistent with the model based on high-quality CRVS data. However, there is also evidence of misclassification of stillbirths and neonatal deaths in FBHs that could lead to errors in mortality estimates. One study in Malawi found that 20% of 365 neonatal deaths in FBHs were classified as stillbirths by verbal autopsy questionnaires (Liu et al., 2016). This tendency to misclassify stillbirths as early neonatal deaths is corroborated by another study from Guinea-Bissau (Helleringer et al., 2020).

It is reasonable to conclude that estimation of early child mortality poses problems for both FBHs and HDSS, and further validation is needed to better understand bias in each source. In general, overestimation in DHS and underestimation in HDSS could contribute to some of the large differences in estimates of neonatal mortality, though it also important to note that there is considerable heterogeneity in the level of recorded neonatal mortality across HDSS, and it remains difficult to adjudicate whether this is due to the epidemiological context, data collection practices, or, a combination of both.

Our comparison of HDSS and subnational DHS and MICS estimates of child mortality was more conclusive, and suggests that FBH estimates are biased downward. On average, HDSS estimates of child mortality were approximately 7% higher than DHS subnational estimates and 6% higher than MICS. Accounting for the left censoring of HDSS data increased the magnitude of this difference even further. In theory, HDSS should be well-placed to prospectively track child mortality. The longer a child survives, the more likely they are to be added to the HDSS household roster, which in turn prompts fieldworkers to inquire about their vital and residency status at each follow-up interview. Thus a prospective surveillance system is well-suited to accurately measure late child mortality. In contrast, FBHs are vulnerable to retrospective reporting errors. These include the omission of past births or deaths, and date misreporting errors when the child's date of birth or age at death are not accurately reported (Pullum and Becker, 2014). This proposition is supported by our finding that DHS markers of low data quality were correlated with a larger disparity with the postneonatal and child mortality estimates from HDSS.

We identified other factors correlated with the disparity between HDSS and DHS estimates of U5M in the regression analysis. HDSS digit preference in day of death reporting was associated with lower neonatal and higher postneonatal mortality. This suggests that HDSS underestimation of neonatal mortality is not solely due to poor registration of newborns, but also rounding errors in the reported ages at death. Estimates of neonatal mortality for HDSS located in Southern Africa were also much lower than DHS estimates. It could be that HDSS estimates for this region are affected by downward bias that was not well-captured by the covariates in the model. A more

detailed examination of the fieldwork procedures and data quality of HDSS in this region is recommended.

This study has several important limitations. First, HDSS produce mortality estimates for localized areas that are generally much smaller than DHS or MICS subnational regions. HDSS data is not designed to be representative of the broader region, and some differences in HDSS and FBH mortality estimates could be attributable to geographic variations. For example, higher values of HDSS child mortality could be due to HDSS being located in areas that have higher mortality relative to the general population (Deribew et al., 2016). Conversely, it has also been suggested that HDSS residents may benefit from living in a testing ground of public health interventions (Sankoh and Byass, 2012).

There is also the issue of small sample sizes when stratifying DHS and MICS estimates by subnational region. The DHS typically reports mortality estimates for subpopulations in 10-year periods due to concerns regarding sample size and displacements of events out of the most recent 5-year period (Pullum and Becker, 2014). We used 5-year periods in our analysis, but found that the main conclusions regarding HDSS-FBH differences remained unchanged with the use of a 10-year window (see appendix Figure B.4). We also compared HDSS mortality estimates with national and residence (urban/rural) level FBH estimates, which have larger sample sizes (see appendix Figure B.5). In this investigation, the magnitude and direction of differences between HDSS and DHS estimates were similar to the subnational region comparison. The results changed slightly for MICS, where the median difference for national and residence level estimates of child mortality with HDSS was close to zero. It is important to interpret individual differences between HDSS and FBHs cautiously due to the variation in sample sizes and relative rarity of mortality as an event. Nevertheless, the aggregate differences and regression analysis provide useful insight into discrepancies between DHS and HDSS estimates, and the contextual and data attributes that may explain them.

3.6 Conclusion

Estimates of U5M for sub-Saharan Africa heavily rely on a combination of FBH data from surveys and model age patterns (UN IGME, 2018; Wang et al., 2017). This practice is problematic for two reasons. First, the comparison of African HDSS and FBH data shows that FBH estimates of U5M may be affected by downward bias. Second, our analyses suggest that the age pattern of U5M is atypical and not well-represented by existing model life tables. More specifically, our findings indicate that sub-Saharan African populations are characterized by relatively high levels of child mortality for a given level of infant mortality, and this could invalidate the use of model age patterns derived from both contemporary and historical data of populations with high-quality

CRVS. This conclusion comes with the caveat that neither of our data sources (FBH or HDSS) represent a gold standard measurement of U5M. Ultimately, there is a need for accurate primary data on U5M, and such data will be essential to reaching targets for its reduction.

3.7 Acknowledgments

We thank Bruno Masquelier, Michel Guillot, and Andrea Verhulst for providing valuable comments on early versions of the manuscript.

Chapter 4

Pregnancy reporting and biases in under-5 mortality in three HDSS in western, eastern, and southern Africa

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1600573	**Title**	Ms
First Name(s)	Hallie		
Surname/Family Name	Eilerts		
Thesis Title	Under-five mortality estimates for sub-Saharan Africa: an inquiry into data sources and estimation methods		
Primary Supervisor	Georges Reniers		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Population Studies
Please list the paper's authors in the intended authorship order:	Eilerts, Hallie; Romero Prieto, Julio; Herbst, Kobus; Gareta, Dickman; Jasseh, Momodou; Khagayi, Sammy; Obor, David; Eaton, Jeffrey, W.; Reniers, Georges
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the study in collaboration with Georges Reniers and Julio Romero Prieto. I conducted all data analysis, interpreted findings, and wrote the first draft of the manuscript. All co-authors have provided input on the analysis and comments on the manuscript.

SECTION E

Student Signature	Hallie Eilerts
Date	21 March 2022

Supervisor Signature	Georges Reniers
Date	21 March 2022

4.1 Abstract

In the absence of complete vital registration, Health and Demographic Surveillance Systems (HDSS) serve as important sources of population-based data throughout sub-Saharan Africa. However, HDSS data on the vital status of newborns is often unreliable due to omission those who were born and died between two rounds of data collection, and therefore never enumerated. This work investigates whether registering pregnancies improved under-5 mortality (U5M) estimation in three African HDSS in The Gambia, western Kenya, and KwaZulu-Natal, South Africa. Less than half of births were registered as pregnancies in each site. Births with registered pregnancies had higher mortality overall and especially at early ages. Cox proportional hazards models with inverse probability weights were used to investigate differences in the mortality of children by pregnancy registration status, controlling for potential confounding from shared risk factors for pregnancy registration and U5M. We found that pregnancy registration was not well-predicted by observable characteristics. The elevated mortality of children with registered pregnancies was not attributable to selectivity in pregnancy registration, but appeared to result from improved ascertainment of early deaths. We highlight the crucial role of pregnancy registration in estimation of mortality at early ages, and argue that improving the quality and completeness of pregnancy data is a priority for HDSS.

4.2 Introduction

An estimated 2.8 million African children under age five died from mostly preventable causes in 2019 (UN IGME, 2021). Accurate measurement of levels, trends, and age patterns of under-5 mortality (U5M) is essential to tracking and accelerating progress towards its reduction. Civil registration and vital statistics (CRVS) systems are the preferred sources for this information, generating reliable and up-to-date data on births, deaths, and causes of death (AbouZahr et al., 2015b). Unfortunately, national CRVS systems are either incomplete or non-existent in regions with the highest U5M burden, including most of sub-Saharan Africa (Mikkelsen et al., 2015). Retrospective maternity history questionnaires included in sample surveys (e.g. Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS)) often serve as the foundation for U5M estimates in such settings. Health and Demographic Surveillance Systems (HDSS) are another valuable source of population-based data which stand apart for collecting detailed epidemiological and socio-demographic information on a longitudinal basis.

HDSS conduct surveillance on geographically-defined populations of several thousand individuals (Sankoh and Byass, 2012). After an initial census, information is typically collected through regular household interviews spaced every three months to one year, depending on the study. HDSS are often considered the gold standard for detailed demographic, health, and social dynamics data in areas of sub-Saharan Africa where high-quality data is not available. However, though the prospective data collection of HDSS is highly effective at tracking the vital status of established residents, reporting on new and transient residents is not as exhaustive. This limitation applies to newborns who are born and die between two rounds of data collection. Those who survive to the first HDSS interview following their birth are more likely to be identified and included on the household roster. This facilitates the tracking of all future events such as mortality of out-migration. However, early deaths are frequently missed, either because the respondent does not report the birth and death, or as a result of insufficient probing by the interviewers. Therefore, there are persistent concerns regarding the accuracy of HDSS data collected on the mortality of children under age five, and especially within the first few weeks of life (Sankoh and Byass, 2012).

This work investigates downward bias in HDSS estimates of early mortality using pregnancy data from three sites in sub-Saharan Africa. We begin by reviewing the pregnancy data for each site, and identifying births that were registered as pregnancies. We then match mother-level information on pregnancy registration to child survival, and compare U5M estimates for cohorts of births which were registered as pregnancies to those that were not. Mortality estimates for cohorts of births with prior pregnancy registration (Cantrelle, 1969) or only counting exposure time since birth if the pregnancy was registered (Nareeba et al., 2021) have been previously found to be

higher than naïve estimates. However, what is not clear is whether this finding is due to selection bias from shared risk factors for pregnancy registration and U5M. We investigate aggregate differences in the mortality of cohorts of births with and without pregnancy registrations through micro-level analysis. Our approach utilizes Cox regression models with inverse probability weighting to standardize cohorts of births with and without registered pregnancies with respect to baseline characteristics, and estimate the marginal effect of pregnancy registration on mortality measurement. We assess whether pregnancy registration results in more accurate estimation of U5M, and conclude with a discussion of data collection priorities and protocols in HDSS.

4.2.1 Downward bias in HDSS estimates of early child mortality

Under-registration of early deaths has long been documented in surveillance-based research in sub-Saharan Africa. Longitudinal cohort studies collecting mortality and morbidity data in west Africa in the 1950s and 60s noted the tendency for deaths occurring shortly after births to be undercounted (Billewicz and McGregor, 1981; Cantrelle, 1969). Certain studies collected information on pregnancies to reduce the risk of omissions (Cantrelle, 1969; Cantrelle and Leridon, 1971; Cantrelle, 1974), though such data was incomplete, and downward bias likely persisted (Garenne, 1981; Pison and Langaney, 1985). These early surveillance studies paved the way for other population-health research stations to be founded across the continent in the 1980s and 1990s (Ngom, 2001). With time, demographic surveillance practices have become more standardized as independent research centers engaged in networks of affiliated HDSS (Utazi et al., 2018). HDSS throughout sub-Saharan Africa generate robust, prospective data on a wide array of population health indicators. Nevertheless, measurement of early mortality remains a challenge, and a regularly reported weakness in HDSS data (Alabi et al., 2014a; Assefa et al., 2016; Kahn et al., 2012; Kishamawe et al., 2015; Rossier et al., 2012).

Huge variation in HDSS estimates of neonatal mortality for given levels of U5M is indicative of data quality issues. In an analysis that utilized data from 31 HDSS sites for the period 2009-2014, four sites reported neonatal mortality rates of less than 10 deaths per 1,000 live births, while the highest estimate was 41.6 deaths per 1,000 live births (Waiswa et al., 2019). The interquartile range of such estimates was more than double that of national estimates compiled from DHS. We have previously noted the important heterogeneity of neonatal mortality in HDSS in a systematic comparison with large cross-sectional survey estimates from 1990-2018 (Eilerts et al., 2021a). The median HDSS neonatal mortality estimates were on average around 8% lower than the DHS subnational region estimate, and 14% lower than that of MICS. Given the differences in methodologies, the composition of samples, and the schedules of data collection; some divergence of HDSS and survey estimates was expected. However, both studies found that markers of good data quality such as having frequent

interview rounds and precise date reporting were linked to higher mortality HDSS estimates (Eilerts et al., 2021a; Waiswa et al., 2019).

Extremely low levels of early mortality in HDSS can also be considered implausible in light of established regularities in the age pattern of mortality. For instance, in the same analysis of data from 2009-2014 in 31 HDSS sites, the overall proportion of neonatal deaths occurring on the first day of life was also only 1.3%, compared to the expected level of around 40% (Waiswa et al., 2019). Indeed, traditional mortality models are based on the experience of mostly European countries, and there is much evidence for the existence of a unique African age pattern of U5M (Blacker, Hill, and Timaeus, 1985; Cantrelle and Leridon, 1971; Garenne, 1981; Jasseh, 2003; Pison and Langaney, 1985). However, recent research found that while African HDSS and DHS both deviate from expected patterns of late child mortality, HDSS are alone in their deviation at early ages (Verhulst et al., 2021). This disagreement between sources suggests issues of data quality, and possible under-registration of early deaths in HDSS.

4.2.2 Pregnancy registration and outcome follow-up

Demographic surveillance entails collection of data on the vital events of births, deaths, and migrations. While not typically defined as a vital event, some HDSS also record information on pregnancies. In certain sites, notifications of recent pregnancies are used to coordinate follow-up on pregnancy outcomes. The previous pregnancy status reports thus serve as prompts for fieldworkers to probe for the pregnancy outcome in subsequent data collection rounds. This can transform the process of reporting recent births from retrospective to prospective, leveraging the strengths of HDSS. However, the collection and use of pregnancy data in HDSS is varied (Kwon et al., 2021; Waiswa et al., 2019). Protocols for pregnancy surveillance are less standardized across sites than for other core components. For instance, some HDSS record information on completed pregnancies upon locating a live baby in routine data collection, or impute pregnancy notifications for data integrity purposes (Waiswa et al., 2019). In other sites, information on ongoing pregnancies is collected through pregnancy status reports, but not integrated into subsequent data collection. Such systems allow for the collection of various pregnancy-related indicators, but do not facilitate the follow-up of pregnancy outcomes.

Collecting data on pregnancies in HDSS is challenging for a variety of reasons. Most sites utilize systems of proxy reporting where one individual reports information on behalf of all household members. The use of a proxy respondent is necessary for reporting on a resident who has recently deceased or out-migrated. However, proxy reports are likely less effective for collecting information on pregnancies. Information regarding pregnancy status is sensitive, and women may conceal their pregnancy so as to avoid gossip, the shame that can accompany giving birth out of wedlock, or stigma associated with pregnancy loss (Haws et al., 2010; Kwesiga et al., 2021). In many cases,

the proxy respondent may not be aware of the pregnancy status of women in his or her own household until it is evident. Whether or not the pregnancy is reported also depends on the frequency of HDSS data collection rounds. In sites where data collection rounds only occur once or twice per year, it is a chance occurrence that the household interview will coincide with the late stages of gestation, when the pregnancy is most likely to be observed by the fieldworker or reported by or on the behalf of the subject.

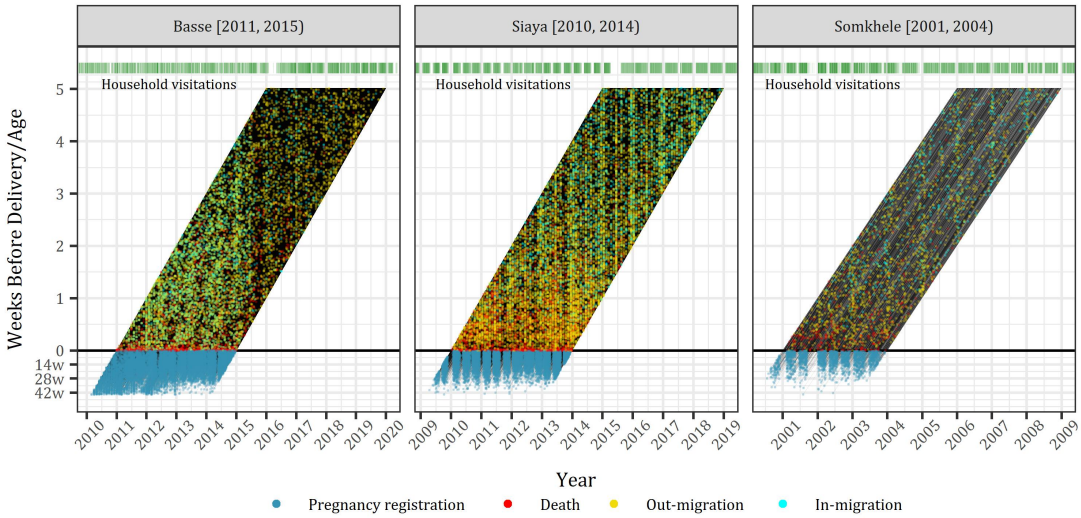
There are several other barriers to reporting pregnancies in HDSS. The use of male interviewers negatively impacts data quality and completeness (Harling et al., 2018; Johnson et al., 2009), especially for such subjects as pregnancy and childbirth (Akuze et al., 2020; Kadobera et al., 2017). Some HDSS employ female interviewers to collect data from women about pregnancy and childbirth, though recruiting and retaining staff can be a challenge (EN-INDEPTH, 2016; Alabi et al., 2014b). Other procedures such as conducting the interview in a private place, or engaging in rapport-building to make the respondent feel comfortable are essential for collecting data on sensitive events, but not always included in HDSS interviewer training (Kwesiga et al., 2021). HDSS data collection protocols have been designed to accommodate diverse research priorities under strict organizational and resource constraints. In general, they are not specifically orientated towards collecting data on pregnancies, and incomplete pregnancy registration and outcome follow-up is often the result.

Despite cross-site heterogeneity in data collection protocols and substantial barriers to reporting, pregnancy registration could be key to addressing the quality issues in HDSS estimates of U5M. We investigate these issues in a comparative analysis of pregnancy reporting and U5M in three African HDSS. This work contributes to understanding of pregnancy registration as an integral component of HDSS data collection, and one that has profound implications for the quality of early child mortality estimates.

4.3 Data

We used data from three HDSS sites located in Basse, The Gambia; Siaya, Kenya; and Somkhele, South Africa. The analytical sample from each site consisted of children born to resident mothers during certain years of interest. In all three HDSS, the nature of data collection has evolved over time with respect to the frequency of interview rounds, the size of the study area, and the availability of resources. The years of interest were thus defined to maintain consistency in these factors to the greatest extent possible. Additionally, as we were interested in estimating U5M, it was necessary that the latest available data collection round in the HDSS occurred at least five years after all births included in the sample. Lexis diagrams depicting the births included in the analysis (and associated demographic events of interest) from each site are shown in Figure 4.1. The were 29,447 births from Basse, 28,867 from Siaya, and 4,633 from

Somkhele. The years of interest in Somkhele HDSS were earlier than the other two sites. These were selected due to there being roughly equal numbers of births with and without registered pregnancies during this time, and comparatively low levels of pregnancy registration in the site in later years. More detail is given on the context and data collection protocols of each site in the following sections.



Note: The dates of data collection through regular household visitations are plotted in bars across the top of each panel, with color intensity denoting more or less data collected during the same day of fieldwork. The dates of data entry in Basse were used to infer household visitations, which were not included in the dataset.

FIGURE 4.1: Lexis diagrams depicting the cohorts included in the analysis from each HDSS site.

4.3.1 Basse HDSS, The Gambia (MRC Unit, The Gambia at LSHTM)

The Basse HDSS is run by the Medical Research Council (MRC) Unit The Gambia at the London School of Hygiene and Tropical Medicine (LSHTM). The town of Basse is on the south bank of the River Gambia, in the eastern Upper River Region of the country. The HDSS is located in the Fulladu East and Kantora Districts, in a predominantly rural setting. Demographic surveillance began in July 2007 to support ongoing studies related to pneumococcal and diarrheal diseases in infants and young children (UKRI, 2022). The prevalence of communicable diseases among the population is high. Malaria prevention measures have contributed to declining rates of infant and U5M and an increasing concentration of deaths at neonatal ages (Jasseh et al., 2015).

The HDSS collected data in interview rounds occurring three times per year from 2011 to 2020 (Rerimoi et al., 2019). Resident village reporters, recruited and trained by the HDSS, were also tasked with keeping track of demographic events on an ongoing basis. HDSS fieldworkers used this information to cross-check data collected during household interviews (Rerimoi et al., 2019). Interviews were conducted with

heads of households, or suitable representatives, who reported events on behalf of all household members. Pregnancies that were reported in household interviews were followed-up to record their outcome. In cases where the pregnancy was not registered, it was the practice of the HDSS to impute a pregnancy report for the day before birth.

4.3.2 Siaya HDSS, Kenya (KEMRI/CDC)

The HDSS in Siaya County is located to the northeast of Lake Victoria in Nyanza Province, western Kenya. It is operated jointly by the Kenya Medical Research Institute (KEMRI) and Centers for Disease Control and Prevention (CDC). In the early 1990s, the KEMRI/CDC partnership established surveillance infrastructure to collect information on malaria morbidity, mortality, and interventions as part of an insecticide-treated bed net trial in this area (Odhiambo et al., 2012). Surveillance continued after the completion of the trial, and the site was formalized as an HDSS in 2001. The HDSS initially consisted of two communities (Asembo and Gem), and was expanded to a third (Karemo) in 2007. Malaria is endemic to the area, and the prevalence of HIV and tuberculosis are among the highest in the country (Hamel et al., 2011).

Routine data collection rounds were initially conducted three times per year, and switched to twice per year in 2015. This transition did not affect the registration of pregnancies or enumeration of births in the years of interest; and if affected the reporting of mortality, it was not during the first year of life. At each household interview, fieldworkers collected information on pregnancy status and recent pregnancy outcomes from women of reproductive age (approximately 15-49 years). If a woman was not present at the time of the interview, the questions were posed to a household proxy respondent. Reports of pregnancies were used to prompt fieldworkers to follow-up on pregnancy outcomes. Prompts continued until an outcome was reported, or the woman was lost to follow-up and the event was censored. In cases where the pregnancy was not reported, births were recorded without any pregnancy registration. Local key informants have also been trained to collect data on births and deaths in a parallel continuous community reporting system (Odhiambo et al., 2012). HDSS fieldworkers referred to these records during data collection to assure data completeness.

4.3.3 Somkhele HDSS, South Africa (AHRI)

The Africa Health Research Institute (AHRI) conducts demographic surveillance activities from its Somkhele Research Campus in the uMkhanyakude district of KwaZulu-Natal province, South Africa (AHRI, 2021). In 2000, surveillance was initiated on a study area of 438km^2 with a population of approximately 85,000 individuals (Gareta

et al., 2021). In 2017, the size of the study area nearly doubled, and surveillance was expanded to a population of approximately 140,000 (Gareta et al., 2021). The area is predominantly rural, though it contains an urban township and some peri-urban settlements. HIV has severely affected the population since the start of the epidemic (Gareta et al., 2021). Prevalence of HIV was 51% for women aged 25-29 in 2003 (Tanser et al., 2008).

During the years of interest, regular data collection rounds were conducted twice annually, with the exception of 2002 which had three rounds. At each household interview, basic data were collected from the household representative who was typically the head of the household, or the next available senior adult household member. The household representative reported pregnancy statuses and recent pregnancy outcomes of female household members aged 15 years and older. Data collection forms were pre-populated with information collected in previous rounds. This would inform the interviewer if the individual in question had recently been pregnant, and prompt the interviewer to inquire about the pregnancy outcome if one had not yet been reported.

4.4 Methods

4.4.1 Descriptive summary

We identified births that were registered as pregnancies within each cohort. This was straightforward for pregnancy observations that had been successfully followed up, and were linked to pregnancy outcomes in the databases. However, it was not uncommon to observe stand-alone pregnancy registrations for which outcomes were never ascertained. When possible, we inferred the outcomes of such pregnancies from separate records of births, or records of children residing in the HDSS who belonged to the mother. Pregnancy registrations were matched to outcomes occurring in the following 44 weeks. While it is unlikely that a pregnancy would be registered 44 weeks prior to delivery, imprecision in date of birth reporting was common in each site. The 44 week window allowed for pregnancy registrations to be matched to births, even in the presence of minor date reporting errors. Pregnancy registrations from the day before or same day as a birth, or within the same data collection round, were discounted. These were considered to have been imputed, and not reflective of prospective follow-up.

Descriptive statistics were calculated for all births, stratified by the mother's pregnancy registration status. We assessed the cohort composition by mother-level characteristics such as age, education level, HIV status, marital status, and migration status. Mothers were coded as HIV-negative if they had a negative test result any time after the delivery. Alternatively, they were labelled HIV-positive if they had a positive test result prior to the delivery, or within six months postpartum. Information on mother's HIV status was not available for Basse HDSS, however this was not expected to be a

limitation given the extremely low prevalence of HIV in The Gambia (UNAIDS, 2020). Marital status was taken from the most recently recorded observation prior to the birth. If there was no marital status recorded prior to the birth, any that was recorded within the following year was used. Otherwise, the mother's marital status was considered unknown. The mother's migration status was coded as a binary indicator for whether she was resident in the HDSS for the entire year preceding the birth. Internal moves within the HDSS site were not considered migrations. For Basse, mother's ethnicity was also considered (Fula, Mandinka, Sarahule, Other).

Other characteristics of interest were related to the household of the mother. For Basse and Somkhele, information on the identity of the household head (and likely proxy respondent) was available. We assessed the age and sex of the household head at the time of the birth, and whether the mother in question was the household head. For all sites, we evaluated whether there was a sibling (i.e. child born to the same mother) under 18 months old who was residing in the household at the time of the birth in question, and whether any older sibling had died prior to the birth. Household wealth was calculated as a weighted average of variables denoting socio-economic status, with weights generated from a principal components analysis. Wealth indices were constructed separately for each site, and included such variables as access to electricity, type of toilet facility, water supply source, and ownership of assets such as a radio, television, and cooking stove. We ranked the households by a combined wealth score, and categorized them into quintiles. Chi-square tests were used to assess whether births with prior pregnancy registrations differed by measured characteristics.

4.4.2 Pregnancy and birth cohorts

We calculated mortality schedules for births that were registered as pregnancies (*pregnancy cohort*), births lacking prior registration (*birth cohort*), and all births combined (*naïve estimates*). Deaths and exposure time were enumerated for the cohorts by single days of age from birth to five years. Permanent out-migrants stopped adding exposure time after migration (right censoring) and returning-migrants did not add exposure time while they were living outside of the surveillance area (interval censoring). To transform the data from a format of cohort-period to age-cohort, the total number of deaths at zero days of life was estimated by combining the deaths from the same day of the delivery with one-half of the deaths reported for the day after (under the assumption that one-half of newborns who died the day after delivery are less than one-day old). Deaths and exposure time were aggregated over age-groups in order to calculate abridged life table rates for the following exact ages: 7, 14, 21, 28 days; 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 18, 21 months; and 2, 3, 4, 5 years. Abridged rates were used to calculate cumulative probabilities of dying under the assumption of constant

mortality rates at the age interval. Bootstrapped 95% confidence intervals were calculated from 5,000 resamples with replacement. We graphically compared the mortality of the pregnancy and birth cohorts, and naïve estimates.

4.4.3 Statistical analysis

Aggregate differences in the mortality of the pregnancy and birth cohorts were investigated using micro-level analysis. We used inverse probability (IP) weighted Cox models to estimate the effect of pregnancy registration on U5M, controlling for potential selection bias from mothers who were more likely to register pregnancies. This approach has been adapted from work by Robins et al. (2000) on marginal structural models, and subsequent applications to survival analysis (Hernán, Brumback, and Robins, 2000; Buchanan et al., 2014). The inverse probability weights (IPWs) were used to standardize the pregnancy and birth cohorts with respect to observable characteristics. Births in the pregnancy cohort were weighted by the inverse probability of pregnancy registration; while those in the birth cohort were weighted by the inverse probability of non-registration. In both cases, births that were extremely likely to have their own pregnancy registration status (whether registered or not registered) were weighted downwards, while unlikely members of either cohort were weighted upwards. The effect of the IPWs was to create a pseudo population of similar sample size in which pregnancy registration was independent from all covariates; thus approximating the random assignment of pregnancy registration (Xu et al., 2010).

The IPWs were calculated as the inverse of the covariate-conditional probability of pregnancy registration using probit regression. The outcome variable was a binary indicator for pregnancy registration. As explanatory variables, the models included covariates which were *a priori* hypothesized as potential confounders of the relationship between pregnancy registration and U5M. This consisted of the characteristics included in the descriptive statistics, in addition to the month of the birth. Covariates were removed from the models if they did not contribute significantly to model fit, as measured by AIC. When covariates in the weight model are highly predictive of the exposure, this can result in very large weights for a few individuals and a weighted estimator with high variance. To account for this, we stabilized the IPWs by multiplying by the marginal probability of pregnancy registration, estimated in a separate probit model with no covariates (Robins, Hernán, and Brumback, 2000).

We calculated pseudo-R^2 values to assess goodness of fit of the probit weighting models (McFadden, 1974). High goodness of fit (pseudo-R^2 values of 0.2-0.4) (McFadden, 1977) was considered indicative of a covariate imbalance between the pregnancy and birth cohorts, and a potential confounding effect on the relationship between pregnancy registration and U5M. We also calculated the overall accuracy of the models, which was defined as the proportion of individuals whose pregnancy registration status was correctly predicted using a probability cut-off of 50%.

Cox models were fit for each site and neonatal (birth - 28 days), postneonatal (28 days - 1 year), and child (1 year to 5 years) age groups. Survival time was measured in days, and individuals only contributed exposure time while they were resident in the HDSS. Observations were censored upon permanent out-migration or loss to follow-up. We first regressed survival status on an indicator for pregnancy registration in unweighted baseline models. We then calculated standardized hazard ratios for pregnancy registration using IP weighted models with robust standard errors. The proportional hazards assumption was verified graphically and using Schoenfeld residuals for unweighted and IP weighted samples (Schoenfeld, 1982).

All statistical analysis was performed using R version 3.6.1.

4.5 Results

4.5.1 Pregnancy and birth cohorts

Table 4.1 summarizes the characteristics of the pregnancy and birth cohorts in each HDSS site. The birth cohorts were larger than the pregnancy cohorts in all sites during the periods of analysis. The pregnancy cohort was relatively larger in Basse than the other two sites at 46% of the total sample. In Siaya and Somkhele, 38 and 41% of observations belonged to the pregnancy cohort, respectively.

In Basse HDSS, the household head was aged 40-59 for 40% of observations in the pregnancy cohort compared to 37% in the birth cohort. The vast majority of mothers lived in households with a male household head. The mother in question was the household head for 1% of observations in the birth cohort, and less than 1% in the pregnancy cohort. In Somkhele HDSS, there were not significant differences in the age or sex of household heads between pregnancy and birth cohorts, and the mother was the household head in approximately 4% of cases.

In Siaya, children in the pregnancy cohort disproportionately resided in Asembo and Gem localities, while those of the birth cohort most often lived in Karemo. In Basse HDSS, the largest share of the birth cohort was found in the top wealth quintile. In both Siaya and Somkhele, those belonging to the pregnancy cohort were more likely to be from wealthier households, while the birth cohort had a larger proportion of cases in the lower levels of wealth. Household wealth quintile was unknown in 20-25% of cases in Basse and Siaya compared to only 4-5% in Somkhele. This was due to data availability and the timing of household socio-economic questionnaires, which were administered less frequently in Basse and Siaya.

There were small but significant differences in the age and education level of pregnancy and birth cohort mothers in Basse and Siaya HDSS. In both sites, the majority of mothers in the pregnancy and birth cohorts were between 20-29 years of age. However, the birth cohorts had larger proportions of mothers who were aged 10-19 years.

TABLE 4.1: Characteristics of pregnancy and birth cohorts.

		Basse HDSS [2011, 2015) *Cohort*			**Siaya HDSS [2010, 2014)** *Cohort*			**Somkhele HDSS [2001, 2004)** *Cohort*		
		Birth n=15,998	Pregnancy n=13,449	p-value	Birth n=17,893	Pregnancy n=10,974	p-value	Birth n=2,746	Pregnancy n=1,887	p-value
Household Head Age	15-39	0.14	0.13	<0.01				0.26	0.28	0.22
	40-59	0.37	0.40					0.47	0.47	
	60+	0.43	0.42					0.26	0.24	
	Unknown	0.06	0.05					0.01	0.01	
Household Head Sex	Female	0.06	0.05	<0.01				0.29	0.30	0.73
	Male	0.87	0.90					0.70	0.69	
	Unknown	0.06	0.05					0.01	0.01	
Household Locality	Asembo				0.25	0.34	<0.01			
	Gem				0.36	0.35				
	Karemo				0.39	0.31				
Household Wealth Quintile	1	0.19	0.18	<0.01	0.14	0.15	<0.01	0.19	0.20	0.04
	2	0.17	0.17		0.15	0.16		0.18	0.21	
	3	0.11	0.12		0.15	0.15		0.19	0.19	
	4	0.17	0.19		0.16	0.17		0.18	0.18	
	5	0.14	0.15		0.15	0.12		0.21	0.18	
	Unknown	0.23	0.20		0.25	0.24		0.05	0.04	
Mother Age	10-19	0.17	0.11	<0.01	0.20	0.15	<0.01	0.24	0.25	0.71
	20-24	0.31	0.28		0.32	0.32		0.27	0.27	
	25-29	0.23	0.26		0.24	0.26		0.21	0.20	
	30-34	0.17	0.20		0.13	0.16		0.15	0.16	
	35+	0.13	0.15		0.11	0.11		0.13	0.12	
Mother Education Level	None/Primary	0.02	0.01	<0.01	0.79	0.83	<0.01	0.29	0.32	0.20
	Secondary/Religious	0.31	0.30		0.20	0.16		0.67	0.67	
	Unknown	0.67	0.68		0.01	0.00		0.03	0.02	
Mother Ethnicity	Fula	0.31	0.34	<0.01						
	Mandinka	0.20	0.23							
	Sarahule	0.46	0.41							
	Other	0.02	0.02							
Mother HIV Status	Negative				0.25	0.27	0.15	0.41	0.47	0.84
	Positive				0.03	0.04		0.01	0.01	
	Unknown				0.72	0.69		0.57	0.51	
Mother Marital Status	In union	0.38	0.52	<0.01	0.64	0.79	<0.01	0.77	0.95	<0.01
	Not in union[a]				0.25	0.15		0.16	0.05	
	Unknown	0.62	0.48		0.11	0.06		0.07	0.00	
Mother is Household Head	No	0.93	0.95	<0.01				0.96	0.96	0.83
	Yes	0.01	0.00					0.04	0.04	
	Unknown	0.06	0.05							
Mother is Recent Migrant	No	0.16	0.06	<0.01	0.28	0.14	<0.01	0.25	0.26	0.90
Older Sibling	None	0.28	0.17	<0.01	0.28	0.16	<0.01	0.41	0.44	<0.01
	Over 18m	0.67	0.81		0.66	0.78		0.54	0.53	
	Under 18m	0.05	0.03		0.07	0.05		0.05	0.03	
Older Sibling Death		0.06	0.08	<0.01	0.12	0.16	<0.01	0.02	0.02	0.74
Proportion		0.54	0.46		0.62	0.38		0.59	0.41	

Notes:
All variables presented as weighted proportions. Statistical differences tested for significance with Chi-square tests.
[a] In Basse HDSS, marital status was only recorded for women in union. There was no information to distinguish those not in union from those of unknown status.

In Siaya, 83% of mothers in the pregnancy cohort had no education or primary education, compared to 79% of mothers in the birth cohort. The ethnicity of the mother was also considered in Basse. Fula and Mandinka mothers comprised a larger share of the pregnancy cohort, while Sarahule mothers made up the largest share of the birth cohort.

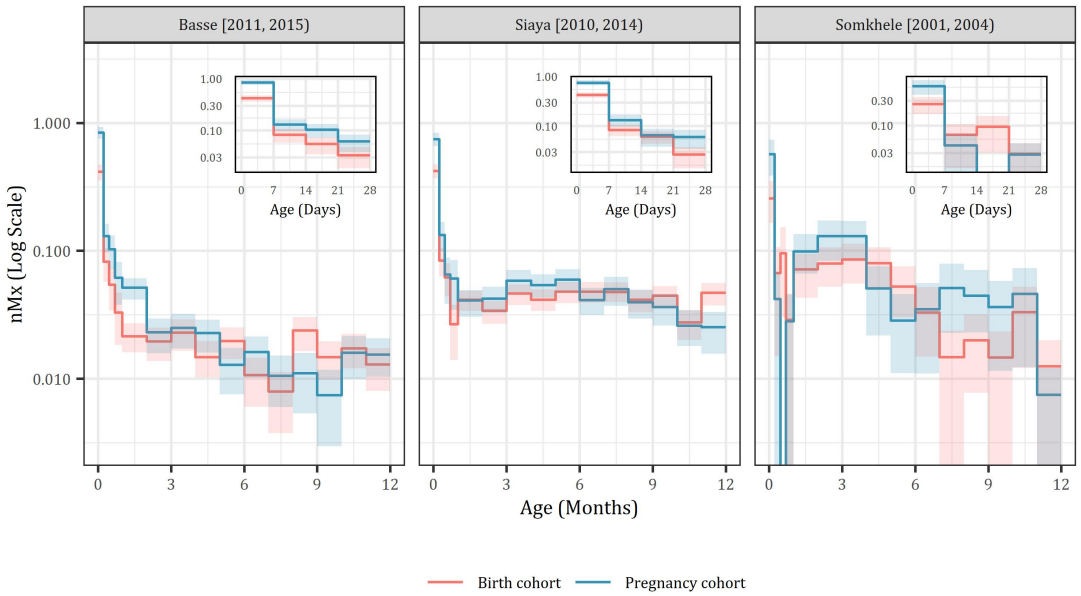
There was not a statistically significant difference in maternal HIV status between pregnancy and birth cohorts in Siaya and Somkhele. In all sites, women in union made up a larger share of the pregnancy cohort compared to the birth cohort. The difference was especially notable in Somkhele HDSS, where 95% of those in the pregnancy cohort were married, compared to only 77% of those in the birth cohort. In Basse and Siaya, the birth cohort had a higher proportion of mothers who had resided outside of the HDSS area for any part of the year preceding their delivery (i.e. recent migrants) compared to the pregnancy cohort. In all sites, observations in the pregnancy cohort were slightly less likely to have a sibling under the age of 18 months in the household at the time of their birth compared to those in the birth cohort. In Basse and Siaya, the death of an older sibling prior was more commonly observed among those in the pregnancy cohort.

Figure 4.2 displays mortality rates for the first year of life in the pregnancy and birth cohorts. In all sites, observed mortality for newborns in the pregnancy cohort was higher during the first week of life compared to the birth cohort. In Basse HDSS, mortality rates in the pregnancy cohort were significantly higher for the first two months of life, before becoming similar to those observed in the birth cohort. For Siaya HDSS, pregnancy cohort mortality rates were significantly higher in the first, second, and fourth weeks of life. There was also a slight increase in mortality risk between months three to six, which was more visible in the pregnancy cohort.

In Somkhele HDSS, the differences between the pregnancy and birth cohorts were not significant after the first week of life. There were zero deaths during the third week of life for the pregnancy cohort, and mortality rates from the second to fifth month were higher than those of the late neonatal period (i.e. from exact age seven to 28 days). This reversal is unusual, given the rapid decline in mortality that typically characterizes the first weeks and months of life. This may be partly explained by the smaller sample size of births in this site, which would make the mortality estimates subject to fluctuations. Alternatively, some of this pattern may be due to upward rounding of ages at death for deaths taking place within the first month of life.

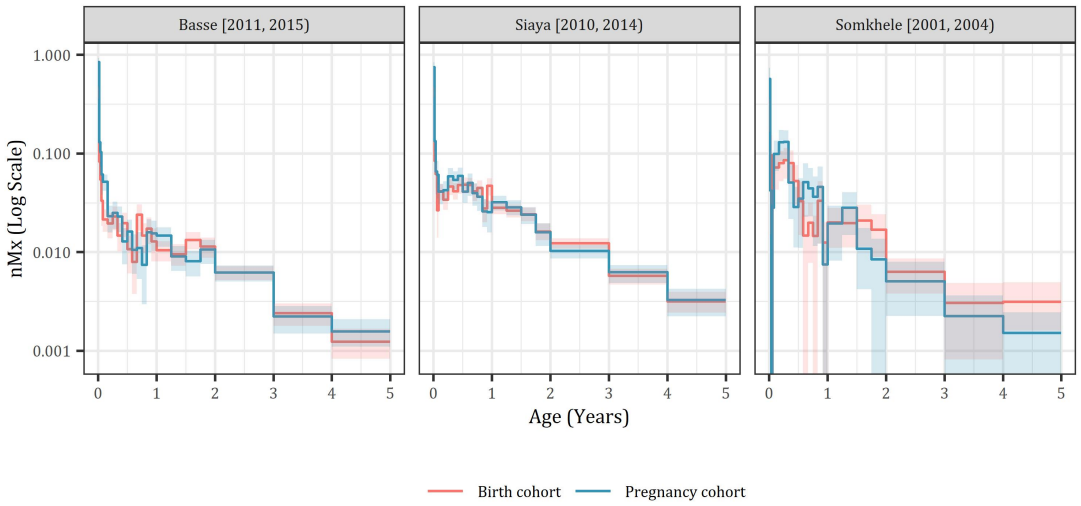
The mortality rates up to five years of age in the pregnancy and birth cohorts are shown in Figure 4.3. Across all sites, mortality rates between the cohorts became more similar with age. In general, there were few significant differences in the mortality of pregnancy and birth cohorts after the first month of life.

The cumulative probabilities of dying by age five for those in the pregnancy and birth cohorts are shown in Figure 4.4. A naïve mortality estimate is also included,



Note: Inset panels show mortality for the first 28 days of life. Bootstrap 95% confidence intervals were calculated for 5,000 resamples with replacement.

FIGURE 4.2: Pregnancy and birth cohort mortality rates up to age one year.

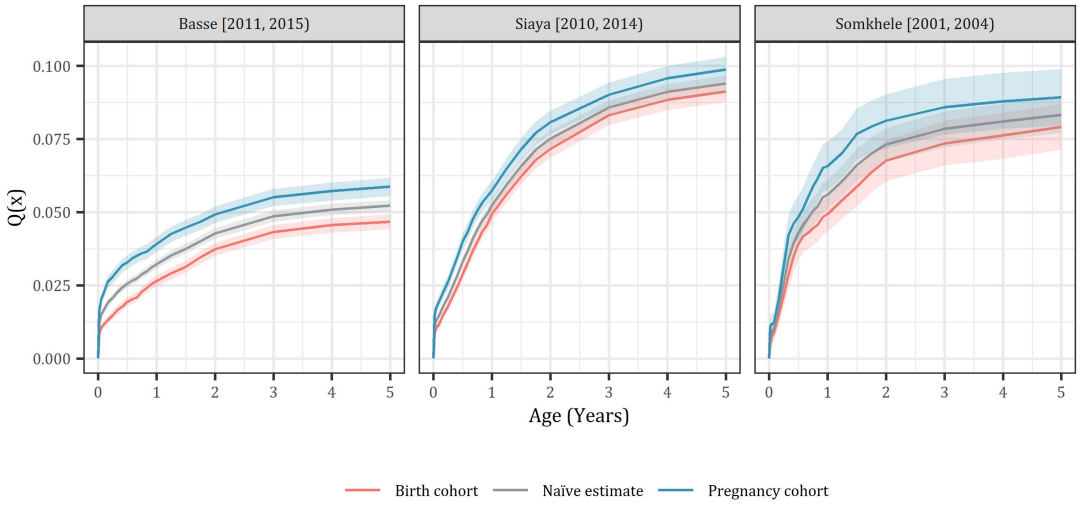


Note: Bootstrap 95% confidence intervals were calculated for 5,000 resamples with replacement.

FIGURE 4.3: Pregnancy and birth cohort mortality rates up to age five years.

which was calculated from all births combined. The pregnancy cohort yielded the highest estimate of U5M in each site. In Basse, the cumulative probability of dying by age five years in the birth cohort was 47 per 1,000 (95% confidence interval [CI] 0.044 − 0.049). In the pregnancy cohort, it was 12 deaths per 1,000 births higher at 59 per 1,000 (95% CI 0.055 − 0.062). The naïve estimate for all births was in between these

values at 52 per 1,000 (95% CI 0.050 – 0.054).



Note: Bootstrap 95% confidence intervals were calculated for 5,000 resamples with replacement.

FIGURE 4.4: Cumulative probabilities of dying by age five years in the pregnancy and birth cohorts, and both combined for the naïve estimates.

In Siaya HDSS, the pregnancy cohort had a cumulative probability of dying by age five of close to 10% (0.099, 95% CI 0.095 – 0.103). The birth cohort had lower mortality at 91 per 1,000 (95% CI 0.088 – 0.095), while the naïve estimate was 94 per 1,000 (95% CI 0.091 – 0.097). In Somkhele HDSS, mortality in the pregnancy cohort was 89 per 1,000 (95% CI 0.080 – 0.100). This was 6 deaths per 1,000 higher than the naïve estimate, and 10 deaths per 1,000 births higher than the birth cohort. The difference between the cumulative probabilities of dying by age five in the pregnancy and birth cohorts was not as statistically significant as in the other two sites, and had overlapping confidence intervals (95% CI -0.007 – 0.028).

4.5.2 Regression analysis

Table 4.2 displays the results of the probit models regressing pregnancy registration on the observable characteristics. Included in the table is the coefficient, standard error, and average marginal effect (AME) on the probability of pregnancy registration, which is expressed in percentage points out of 100. In Basse, those with household heads aged 15-39 were 2% less likely to have a registered pregnancy than those with household heads aged 40-59. The same group was approximately 4% more likely to register pregnancies in Somkhele. Female headed households were around 3% less likely to register pregnancies compared to male headed households in Basse, while this covariate was not significant in Somkhele.

The household locality was important in Siaya, where mothers residing in Asembo were roughly 15% more likely to have a registered pregnancy compared to mothers in

Gem or Karemo. The coefficients for household wealth were most significant in Siaya, where those in the lowest wealth quintile were least likely to have registered pregnancies. Month of birth was significantly associated with pregnancy registration in each site, but the AMEs were largest for Somkhele. In this site, pregnancies which came to term between March and October were significantly more likely to be registered compared to January.

Women aged 10-19 were four percentage points less likely to register pregnancies in Basse HDSS compared to those aged 20-24. In Siaya and Somkhele HDSS, mothers who were older than 25 were typically less likely to register pregnancies than those aged 20-24. Having higher or an unknown education level was associated with a higher probability of pregnancy registration when compared to those with no or primary level education in Basse. In Siaya, this relationship was reversed, while the relationship was not significant in Somkhele. Fula and Mandinka women had a 5-6% higher probability of pregnancy registration in Basse compared to Sarahule women. In Siaya and Somkhele, being of unknown HIV status significantly reduced the probability of pregnancy registration, perhaps indicating that such individuals were less likely to have complete information in the HDSS overall. Marital status had one of the strongest effects on pregnancy registration of all variables in Somkhele. Unpartnered women and women of unknown marital status were approximately 31% and 45% less likely to have registered pregnancies than women in union, respectively. The same relationship was found in the other sites, but of a much smaller magnitude. Being a recent migrant significantly decreased the probability of pregnancy registration in Basse and Siaya. In each site, pregnancies were less likely to be registered if there was a sibling in the household who was less than 18 months old at the time of birth. In Basse, pregnancy registration was more slightly likely in cases where an older sibling had died prior to the birth.

Despite the presence of several significant covariates, the goodness of fit of the probit models was somewhat low by traditional standards for binary classification models. The model for Basse HDSS had the lowest pseudo-R^2 value of 0.04, while Siaya and Somkhele HDSS were 0.06 and 0.08, respectively. The model for Basse correctly predicted pregnancy registration for 59% of cases. This was only slightly higher than the no-information rate (NIR) of 54%; which resulted from classifying all observations as belonging to the birth cohort. There was a difference of three percentage points in accuracy and the NIR for the Siaya model (accuracy = 0.65, NIR = 0.62), and four percentage points for Somkhele (accuracy = 0.63, NIR = 0.59). Thus, predicting pregnancy registration on the individual-level was difficult, and there was substantial randomness in pregnancy registration that was not well-explained by the covariates in the model.

The issue of selectivity in pregnancy registration was investigated further with Cox

TABLE 4.2: Results from probit models regressing pregnancy registration on observable characteristics of the mother and household. Probit regression coefficients, standard errors in parenthesis, and scaled average marginal effects.

		Dependent variable: Pregnancy registration								
		Basse [2011, 2015)			**Siaya [2010, 2014)**			**Somkhele [2001, 2004)**		
		Coef.	(SE)	AME	Coef.	(SE)	AME	Coef.	(SE)	AME
Household Head Age (ref.=40-59)	15-39	-0.05*	(0.02)	-2.0				0.12*	(0.05)	4.3
	60+	-0.02	(0.02)	-0.8				-0.06	(0.05)	-1.9
	Unknown	0.07	(0.42)	2.5				4.52	(36.57)	59.6
Household Head Sex (ref.=Male)	Female	-0.09**	(0.03)	-3.3				0.04	(0.04)	1.6
	Unknown	-0.09	(0.42)	-3.5				-4.48	(36.57)	-40.4
Household Locality (ref.=Karemo)	Asembo				0.40***	(0.02)	14.5			
	Gem				0.03	(0.02)	1.0			
Household Wealth Quintile (ref.=1)	2	0.01	(0.03)	0.4	-0.02	(0.03)	-0.6	0.04	(0.06)	1.4
	3	0.05	(0.03)	1.9	-0.05	(0.03)	-1.9	-0.03	(0.06)	-1.2
	4	0.06*	(0.03)	2.1	-0.02	(0.03)	-0.6	-0.05	(0.06)	-1.8
	5	0.01	(0.03)	0.3	-0.12***	(0.03)	-4.2	-0.06	(0.06)	-2.3
	Unknown	-0.07*	(0.03)	-2.5	-0.09**	(0.03)	-3.2	-0.10	(0.12)	-3.6
Month of Birth (ref. = 1)	2	-0.03	(0.04)	-1.0	-0.09*	(0.04)	-3.0	0.02	(0.10)	0.8
	3	0.06	(0.04)	2.5	0.17***	(0.04)	6.0	0.28**	(0.10)	9.6
	4	0.05	(0.04)	1.8	0.26***	(0.04)	9.2	0.40***	(0.10)	13.9
	5	0.06	(0.04)	2.3	0.11**	(0.04)	4.0	0.33***	(0.10)	11.4
	6	-0.05	(0.04)	-1.8	0.05	(0.04)	1.6	0.25**	(0.10)	8.8
	7	-0.01	(0.04)	-0.5	0.09*	(0.04)	3.3	0.32**	(0.10)	11.0
	8	-0.00	(0.04)	-0.0	0.21***	(0.04)	7.6	0.29**	(0.10)	10.1
	9	0.10**	(0.03)	3.9	0.15***	(0.04)	5.2	0.53***	(0.09)	18.9
	10	0.17***	(0.03)	6.5	0.09*	(0.04)	3.1	0.33**	(0.10)	11.5
	11	0.12***	(0.03)	4.5	0.13***	(0.04)	4.6	0.19	(0.10)	6.5
	12	0.19***	(0.04)	7.4	0.21***	(0.04)	7.5	0.14	(0.10)	4.8
Mother Age (ref.=20-24)	10-19	-0.11***	(0.03)	-4.1	0.02	(0.03)	0.8	0.01	(0.06)	0.3
	25-29	0.03	(0.02)	1.0	-0.10***	(0.02)	-3.7	-0.01	(0.06)	-0.5
	30-34	0.02	(0.02)	0.8	-0.11***	(0.03)	-3.8	0.04	(0.07)	1.6
	35+	-0.05*	(0.03)	-2.0	-0.20***	(0.03)	-6.9	-0.09	(0.07)	-3.3
Mother Education (ref.=None/Primary)	Secondary/Religious	0.18**	(0.06)	6.8	-0.09***	(0.02)	-3.2	-0.08	(0.05)	-2.8
	Unknown	0.23***	(0.06)	8.7	-0.46***	(0.10)	-15.0	-0.09	(0.17)	-3.3
Mother Ethnicity (ref.=Sarahule)	Fula	0.14***	(0.02)	5.1						
	Mandinka	0.15***	(0.02)	5.8						
	Other	0.03	(0.05)	1.0						
Mother HIV Status (ref.=Negative)	Positive				0.04	(0.05)	1.4	-0.17	(0.17)	-6.2
	Unknown				-0.09***	(0.02)	-3.2	-0.13**	(0.04)	-4.8
Mother Marital Status (ref.=In union)	Not in union				-0.43***	(0.02)	-15.1	-0.96***	(0.07)	-31.1
	Unknown	-0.20***	(0.02)	-7.6	-0.49***	(0.03)	-17.0	-2.30***	(0.28)	-45.3
Mother Recently Migrated (ref.=No)	Yes	-0.53***	(0.03)	-19.6	-0.46***	(0.02)	-16.0	0.07	(0.05)	2.4
Older Sibling (ref.=No older sibling)	Over 18m	0.20***	(0.02)	7.6	0.15***	(0.03)	5.4	-0.27***	(0.05)	-9.7
	Under 18m	-0.21***	(0.04)	-7.6	-0.08*	(0.04)	-2.8	-0.64***	(0.11)	-21.8
Older Sibling Death (ref.=No)	Yes	0.07*	(0.03)	2.8	0.03	(0.02)	1.2	0.03	(0.14)	0.9
Observations		29447			28867			4633		
No-Information Rate		0.54			0.62			0.59		
Accuracy		0.59			0.65			0.63		
Pseudo R-squared		0.04			0.06			0.08		

Notes:
*p<0.05; **p<0.01; ***p<0.001
SE - standard error, AME - average marginal effect

TABLE 4.3: Hazard ratios and 95% confidence intervals for the effect of pregnancy registration on mortality under age five.

		Unweighted		*IP weighted*	
HDSS	Age group	HR	95% CI	HR	95% CI
Basse	[0, 28d)	1.94***	(1.61, 2.33)	1.98***	(1.63, 2.41)
	[28d, 1y)	1.16	(0.97, 1.39)	1.14	(0.95, 1.38)
	[1y, 5y)	0.98	(0.83, 1.16)	0.97	(0.82, 1.16)
Siaya	[0, 28d)	1.70***	(1.40, 2.07)	1.79***	(1.46, 2.20)
	[28d, 1y)	1.03	(0.91, 1.16)	1.05	(0.92, 1.20)
	[1y, 5y)	0.99	(0.87, 1.12)	0.98	(0.86, 1.12)
Somkhele	[0, 28d)	1.46	(0.82, 2.61)	1.33	(0.72, 2.45)
	[28d, 1y)	1.32*	(1.00, 1.73)	1.50**	(1.11, 2.02)
	[1y, 5y)	0.81	(0.54, 1.21)	0.84	(0.55, 1.26)

Notes:
*p<0.05; **p<0.01; ***p<0.001
CI - confidence interval, HR - hazard ratio, IP - inverse probability
Robust 95% CIs calculated for IP weighted models.

regression. The predicted probabilities of pregnancy registration from the probit models were transformed into IPWs which had mean values of close to 1 for pregnancy and birth cohorts (see appendix Figure C.2). Results for unweighted and IP weighted Cox models regressing child survival on pregnancy registration are displayed in Table 4.3.

In the unweighted baseline model for Basse HDSS, newborns in the pregnancy cohort had almost double the risk of neonatal mortality (hazard ratio [HR] 1.94; 95% CI $1.61 - 2.33$) compared to the birth cohort. After standardizing for covariate imbalances between pregnancy and birth cohorts using the IPWs, the hazard ratio increased (HR 1.98; robust 95% CI $1.63 - 2.41$). Hence, despite the higher observed mortality of the pregnancy cohort, pregnancy registration was associated with characteristics that were protective in Basse HDSS. Standardizing the pregnancy and birth cohorts with respect to observable characteristics of the mother and household increased the hazard ratio for pregnancy registration even further.

Similar to the results for Basse, the pregnancy cohort had significantly higher neonatal mortality in Siaya HDSS. In the baseline model, newborns in the pregnancy cohort were 1.7 times more likely to die in neonatal ages (95% CI $1.40 - 2.07$) compared to those in the birth cohort. Adjustment from IPWs led to a increased hazard ratio of 1.79 (robust 95% CI $1.46 - 2.20$). In both Basse and Siaya, the increased mortality risk for the pregnancy cohort was not significant in models for postneonatal or child age groups.

Somkhele differed from both of the previous sites in having significant results for the postneonatal age group. Compared to the birth cohort, newborns in the pregnancy cohort had approximately 50% higher risk for postneonatal mortality in the IP

weighted model (HR 1.50; robust 95% CI 1.11 − 2.02). This was an increase from the baseline estimate of 1.32 (95% CI 1.00 − 1.73). For Somkhele, there was moderate evidence of violations to the proportional hazards assumption for the effect of pregnancy registration in models for the neonatal age group. In Basse and Siaya, there was stronger evidence for violations in the models for the postneonatal age group.

4.6 Discussion

Using data from HDSS in The Gambia, Kenya, and South Africa, we found that children whose births were preceded by pregnancy registrations had higher observed U5M than those lacking them. This difference in mortality was primarily attributable to neonatal and postneonatal ages. In Basse and Siaya HDSS, the pregnancy cohort had higher mortality in the first few weeks of life. For Somkhele HDSS, the mortality of the pregnancy cohort was significantly higher during the first week of life and also the postneonatal period, the latter of which may be due to reporting errors in age at death, and the transfer of deaths taking place during the first month to later ages. For all sites, the mortality of the pregnancy and birth cohorts was not statistically different between the ages of 1 and 5 years.

Births with pregnancy registrations made up a minority of the sample in each site (38-46%). Given that Basse and Siaya HDSS were conducting triannual data collection rounds compared to Somkhele's biannual, it is somewhat surprising that these sites did not have substantially higher levels of pregnancy registration. Theoretically, if births were randomly distributed throughout the year, and under the conservative assumption that pregnancy registration would only occur if a data collection round took place during the third trimester, triannual interviews would result in 75% of births having registered pregnancies. That this level was not achieved in either site indicates considerable under-reporting. Furthermore, it suggests that marginally increasing the frequency of interview rounds alone would not be sufficient to achieve exhaustive pregnancy reporting.

Across all three sites, women in union were more likely to have registered pregnancies than those outside of union. Under-reporting of pregnancies for women outside of union could be related to strong social norms against having a child outside of wedlock. Additionally, it is also the case that married women often reside in households where the household head (and likely proxy respondent) is their husband. Each site utilized proxy respondents to report on the pregnancy status of household members. Previous research conducted in a rural setting in Tanzania has shown that while women often conceal their pregnancy from the broader community for as long as possible, disclosing one's status to a sexual partner is frequently done soon after the first missed menses to secure his support in preparing for the pregnancy (Haws et al.,

2010). Thus, proxy reporting likely contributes to the under-reporting of all pregnancies, but especially those of women who are not in formal union.

Given the incompleteness of pregnancy registration in the HDSS sites, it was important to examine whether the higher mortality of pregnancy cohorts was attributable confounding from U5M risk factors. We found that the probit models for pregnancy registration had low goodness of fit. While several characteristics had statistically significant associations with pregnancy registration, they were not strongly predictive of it. These findings indicated low potential for a confounding effect between pregnancy registration and child survival. To confirm this, marginal structural Cox models were used to estimate the effect of pregnancy registration on mortality, controlling for covariate imbalances between the pregnancy and birth cohorts.

Mortality in the pregnancy cohort remained higher after standardizing the observations with respect to measured characteristics of the mother and household. These results support the proposition that mortality estimates for pregnancy cohorts are an improvement over naïve estimates; and one that is not attributable to selection bias. It is important to note that there was evidence for violations to the proportional hazard assumption in the models for certain age groups in each site. Namely, this affected the models for postneonatal mortality in Basse and Siaya, and neonatal mortality in Somkhele. As such, the hazard ratios for pregnancy registration in these ages should be interpreted cautiously. However, these violations also support the proposition that pregnancy registration only resulted in an increased hazard for early ages, and not the entirety of the first five years of life. Thus the overall higher U5M of the pregnancy cohort was not likely due to being negatively selected, but rather improved ascertainment of early deaths.

This study draws on the richness of HDSS data to investigate macro-level differences in the mortality of births with and without registered pregnancies, controlling for micro-level factors. However, the study is subject to some important weaknesses. Pregnancy registration was not randomly assigned in a controlled trial, and determining its effect on measurement of U5M was not straightforward. While the IPWs used in the Cox regression models controlled confounding by measured covariates, they did not adjust for potential bias from unmeasured variables. While there remains a possibility for residual confounding of the association between pregnancy registration and U5M by omitted or insufficiently controlled for factors, it is not likely. Clinical markers are usually a better predictor of early mortality, however, for these to explain the higher estimates of mortality in pregnancy cohorts, they would also have to be strongly associated with pregnancy registration and there is not a strong reason to presume that this would be the case.

It can be safely assumed that pregnancy registration is more likely to take place at later gestational ages, when the pregnancy is evident to the woman in question, household proxy respondent, or HDSS interviewer. As such, the timing of data collection,

as well as the identity of the respondent and interviewer, are likely key variables in predicting pregnancy registration. We had information on the identity of the household head for Basse and Somkhele HDSS, however, these sites allow for household interviews to be conducted with any suitable household representative. Information on the dates of data collection and interviewer identities was also incomplete across the sites included in the analysis during the periods of interest. Future analyses which aim to elucidate the how such variables may affect and explain pregnancy registration are recommended.

Finally, we did not investigate whether issues of misclassification of stillbirths and neonatal deaths differentially impact pregnancy cohorts. Recent research has found evidence of potential overestimation of neonatal mortality in African DHS from stillbirths that are misreported as neonatal deaths (Liu et al., 2016; Helleringer et al., 2020). This work raises important questions regarding the accuracy of data collected on early deaths in maternity history questionnaires, however there has not been much investigation of issues of misclassification in HDSS data. Nevertheless, it has been observed that stillbirths which are misclassified as neonatal deaths are most likely to be recorded as very early deaths—taking place in the first or second day of life (Helleringer et al., 2020). To account for this, we conducted a sensitivity analysis where deaths taking place on the first day of life were omitted from the pregnancy cohorts. These results showed that even in the extreme case that all deaths taking place on the first day of life were stillbirths wrongly classified as neonatal deaths, the pregnancy cohort estimates of neonatal and U5M were still higher than birth cohort estimates for all sites (see appendix Figure C.1). It is thus not likely that differential misclassification of stillbirths could fully account for the higher mortality observed in pregnancy cohorts.

4.7 Conclusion

Our research indicates that pregnancy registration improves follow-up on the vital events of newborns in HDSS. Enhancing the quality and consistency of pregnancy data can address long-standing concerns surrounding the accuracy of HDSS estimates of early mortality. It would also contribute to tracking of maternal health and pregnancy outcomes such as stillbirths, which have not been systematically measured in high-burden countries. These potential payoffs suggest that improving the quality and completeness of pregnancy data should be a top priority in HDSS.

Chapter 5

Evaluating pregnancy reporting in Siaya Health and Demographic Surveillance System through record linkage with ANC clinics

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1600573	**Title**	Ms
First Name(s)	Hallie		
Surname/Family Name	Eilerts		
Thesis Title	Under-five mortality estimates for sub-Saharan Africa: an inquiry into data sources and estimation methods		
Primary Supervisor	Georges Reniers		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	International Journal of Population Data Science
Please list the paper's authors in the intended authorship order:	Eilerts, Hallie; Romero Prieto, Julio; Ambia, Julie; Khagayi, Sammy; Kabudula, Chodziwadziwa; Eaton, Jeffrey, W.; Reniers, Georges
Stage of publication	**In press**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the study in collaboration with Georges Reniers and Julio Romero Prieto. I conducted all data analysis, interpreted findings, and wrote the first draft of the manuscript. All co-authors have provided input on the analysis and comments on the manuscript.

SECTION E

Student Signature	Hallie Eilerts
Date	18 August 2022

Supervisor Signature	Georges Reniers
Date	18 August 2022

5.1 Abstract

INTRODUCTION
Health and Demographic Surveillance Systems (HDSS) are important sources of population health data in sub-Saharan Africa, but the recording of pregnancies, pregnancy outcomes, and early mortality is often incomplete.
OBJECTIVE
This study assessed HDSS pregnancy reporting completeness and identified predictors of unreported pregnancies that likely ended in adverse outcomes.
METHODS
The analysis utilized individually-linked HDSS and antenatal care (ANC) data from Siaya, Kenya for pregnancies in 2018-2020. We cross-checked ANC records with HDSS pregnancy registrations and outcomes. Pregnancies observed in the ANC that were missing reports in the HDSS despite a data collection round following the expected delivery date were identified as likely adverse outcomes, and we investigated the characteristics of such individuals. Clinical data were used to investigate the timing of HDSS pregnancy registration relative to care seeking and gestational age, and examine misclassification of miscarriages and stillbirths.
RESULTS
From an analytical sample of 2,475 pregnancies observed in the ANC registers, 46% had pregnancy registrations in the HDSS, and 89% had retrospectively reported pregnancy outcomes. 1% of registered pregnancies were missing outcomes, compared to 10% of those lacking registration. Registered pregnancies had higher rates of stillbirth and perinatal mortality than those lacking registration. In 77% of cases, women accessed ANC prior to registering the pregnancy in the HDSS. Half of reported miscarriages were misclassified stillbirths. We identified 141 unreported pregnancies that likely ended in adverse outcomes. Such cases were more common among those who visited ANC clinics during the first trimester, made fewer overall visits, were HIV-positive, and outside of formal union.
CONCLUSIONS
Record linkage with ANC clinics revealed pregnancy underreporting in HDSS, resulting in biased measurement of perinatal mortality. Integrating records of ANC usage into routine data collection can augment HDSS pregnancy surveillance and improve monitoring of adverse pregnancy outcomes and early mortality.

5.2 Introduction

There have been substantial improvements in maternal and newborn health since the beginning of global monitoring, however stark regional disparities persist. Sub-Saharan Africa is the region with the highest burden of stillbirths and early mortality (UN IGME, 2020a; UN IGME, 2021). The United Nations Sustainable Development Goals call for reducing neonatal mortality (deaths within the first 28 days of life) to 12 deaths per 1,000 live births by 2030 (United Nations, 2017). While the reduction of stillbirths has not been a focus of international policy and investment agendas (Lawn et al., 2016), visibility has increased in recent years. The Early Newborn Action Plan, launched through a World Health Assembly resolution in 2014, established the target of reducing the stillbirth rate to 10 per 1,000 total births in all countries by 2035 (World Health Organization, 2014). Accurate measurement is seen as key to accelerating progress towards both targets. Health management information systems (HMIS) or civil registration and vital statistics (CRVS) systems are the ideal sources for this information, but they are incomplete or non-existent in most low- and lower-middle-income countries (AbouZahr et al., 2015b).

Nationally representative household surveys are the primary sources of data on pregnancy loss and early mortality in sub-Saharan Africa (Kwesiga et al., 2021). The most prominent among these, the Demographic and Health Surveys (DHS), have been instrumental to monitoring maternal and newborn health. However, the omission of stillbirths and early deaths is a serious limitation, which causes downward bias in mortality estimates (Bradley, Winfrey, and Croft, 2015; Pullum and Becker, 2014).[1] There are persistent challenges associated with collecting survey data on such topics due to their sensitive nature (Kwesiga et al., 2021). Qualitative research in rural Tanzania and Uganda found that strong social norms discouraged adverse pregnancy outcomes such as miscarriage and stillbirth from being discussed publicly (Haws et al., 2010; Kiguli et al., 2015). The use of male interviewers has been found to negatively impact data quality and completeness, especially for such topics as pregnancy and childbirth (Akuze et al., 2020; Kadobera et al., 2017). Furthermore, distinguishing between reported instances of miscarriage and stillbirth can be impacted by survey translation, and overlap in the local language terms for such events (Haws et al., 2010; Kwesiga et al., 2021). It is worth noting that these classification issues are not unique to lower-income settings. A lack of uniformity in benchmarks of gestational age, birthweight, and body length used to distinguish between stillbirth and miscarriage has complicated global monitoring efforts (Blencowe et al., 2016; Lawn et al., 2016).

Health and Demographic Surveillance Systems (HDSS) are important sources of

[1] The DHS recently updated their core reproductive history questionnaire to address data quality issues related to information collected on pregnancy loss and early mortality (DHS, 2019; Akuze et al., 2020).

longitudinal population health data throughout sub-Saharan Africa. HDSS sites routinely collect information on demographic events such as births, deaths, and migrations through recurring household interviews of contiguous populations (Sankoh and Byass, 2012). Though this prospective data collection process is highly effective at tracking the vital status of established residents, it is less reliable for collecting information on pregnancy outcomes and newborns. In the case that a pregnancy ends in a live birth, the newborn will likely be enumerated in the HDSS at the subsequent household interview round. However, pregnancies resulting in adverse outcomes such as miscarriage, stillbirth, and early deaths (before the next household enumeration) are subject to underreporting (Sankoh and Byass, 2012).

Some HDSS sites register ongoing pregnancies to facilitate follow-up on pregnancy outcomes and early mortality (Kwon et al., 2021). We previously found that mortality during the first year of life was higher for cohorts of births with registered pregnancies compared to those where the pregnancy was not observed, implying omission of early deaths from the latter group (Eilerts et al., 2021b). While pregnancy registration appears to be an important tool for improving measurement of early mortality, it is often incomplete in HDSS sites (Waiswa et al., 2019). Furthermore, little is known about the number of pregnancies that are entirely missing from HDSS records, lacking both pregnancy registration and outcome reports. Given indications that pregnancies ending in adverse outcomes are vulnerable to underreporting (Nareeba et al., 2021; Rerimoi, 2019; Waiswa et al., 2019), missing pregnancy reports are a likely source of downward bias in HDSS measurement of stillbirths and neonatal mortality. As HDSS are often deliberately set up in locations where there is limited availability of other population-based data (Sankoh and Byass, 2012), external validation of mortality estimates is difficult.

Record linkage with antenatal care (ANC) data is a promising avenue for improving HDSS pregnancy data, and can help address the data gap on pregnancy loss and early mortality in sub-Saharan Africa. The World Health Organization recommends that all pregnant women have at least eight ANC assessments (World Health Organization, 2016). While only 52% of women in sub-Saharan Africa make four or more visits, close to 90% of women seek ANC services at least once during pregnancy (Adedokun and Yaya, 2020). Integrating records of ANC usage into HDSS data could reduce frequency of surveillance rounds required to comprehensively capture pregnancies and births. ANC data also have more reliable information on the underlying conditions of mothers affecting the survival of newborns. Record linkage has been widely practised to support epidemiological research and health services evaluation in high-income settings (Ford, Roberts, and Taylor, 2006; Chen et al., 1998a; Howard, 2004; Riordan et al., 2012; Holian, Mallick, and Zaremba, 2004; Holian, 2000). It has been less common in sub-Saharan Africa, but the field is growing, with several studies demonstrating its feasibility (Joubert et al., 2014; Kabudula et al., 2014; Rentsch et al.,

2017b).

In this work, we leveraged linked HDSS and ANC data to shed light on pregnancy reporting and potential downward bias in measurement of stillbirths, perinatal, and neonatal mortality in an HDSS in western Kenya. This work provides a framework for how individually-linked HDSS and ANC records can be used to complement HDSS data and improve information collected on pregnancies and their outcomes in settings lacking adequate HMIS and CRVS.

5.3 Methods

5.3.1 Data

The study area is located in Siaya County in western Kenya. The Siaya HDSS was established in 2001 as a collaboration between the Kenya Medical Research Institute (KEMRI) and the Centers for Disease Control (Odhiambo et al., 2012). The site includes the rural communities of Karemo, Asembo, and Gem; covering a total area of approximately 700km^2 with around 224,000 residents in 2020. Data collection has been conducted through household interviews every four months up to 2015, and every six months thereafter. A household proxy respondent reports to the HDSS fieldworker on behalf of all household members, providing information on births, deaths, and migrations which have occurred since the previous data collection round. Information on pregnancy status is collected from women of reproductive age directly, though a proxy respondent may be used if the individual is not present at the time of the interview. In the case that a pregnancy is registered, the HDSS fieldworker is prompted to follow up on its outcome at subsequent data collection rounds. Local community reporters have also been trained to collect data on births and deaths in their villages. HDSS fieldworkers refer to these records during data collection to assure data completeness. More information on the data collection protocols of the HDSS is available elsewhere (Odhiambo et al., 2012).

Beginning in February 2018, the HDSS initiated Point-of-contact Interactive Record Linkage (PIRL) with 14 ANC clinics in the Gem District of Siaya County. We will briefly describe the record linkage process in Siaya, and more detail on the PIRL approach is available elsewhere (Kabudula et al., 2017a; Rentsch et al., 2017a; Rentsch et al., 2017b). Data clerks stationed in clinic waiting rooms invited pregnant women aged 13 or older who were seeking ANC to participate in the record linkage study. After obtaining written informed consent, women were enrolled in the study, and PIRL was attempted for those who self-reported residence in the HDSS. The data clerk collected information on up to three names for the individual, date of birth, location of household in the HDSS, and up to three names of another household member. This information was entered into the PIRL software to search for the individual in the HDSS database. Using the probabilistic framework developed by Fellegi and Sunter (Fellegi

and Sunter, 1969), HDSS and ANC record pairs were compared using a series of identifiers. Each comparison contributed an agreement or disagreement weight towards the total match score of the record pair, with weights calculated as a ratio of the probabilities that true and false matches agreed on the identifier (Cook, Olson, and Dean, 2001; Jaro, 1989). The match probabilities used in the PIRL software were adapted from a previous study which conducted record linkage between ANC clinics and an HDSS in rural Tanzania (Rentsch et al., 2017b). The software returned the 20 highest scoring potential matches from the HDSS ranked in descending order, and the data clerk then consulted the subject to identify the matching record.

5.3.2 Analysis

Sample of linked records

This analysis pertains to individuals who visited ANC clinics after the start of record linkage on February 7, 2018 and had an EDD prior to October 1, 2020. At the time of this analysis, the most recent HDSS data was collected between October and December 2020. The upper bound for EDD thus ensured that at least one HDSS data collection round occurred following the expected completion of all pregnancies included in the sample. For women making their first ANC visit for a given pregnancy during the first trimester, the EDD recorded in the clinic data was typically estimated as 40 weeks after their last menstrual period (LMP). For those with a later gestational age at first visit, the EDD was either estimated from the LMP or by the clinic nurse through fundal height palpation. In cases where the EDD was updated at subsequent ANC visits, we used the latest recorded EDD to evaluate the sample inclusion criteria.

Individuals who were not residing in the HDSS at the time of ANC linkage (as per HDSS records) were excluded from the analysis. As PIRL was only attempted for those who self-reported as current residents of the HDSS, it is possible that the individual incorrectly reported their residency status or that their HDSS residency record was not up-to-date. Alternatively, it could also be the case that the individual had not resided in the HDSS long enough to be considered a permanent resident. New arrivals to the HDSS are subject to a four-month preliminary registration period. Depending on the timing of the individual's in-migration relative to the HDSS data collection rounds, their status as a permanent resident may not be confirmed for several months more. HDSS pregnancy reporting could not be evaluated for such individuals.

External validation of HDSS pregnancy reporting

For linked individuals, we applied a series of deterministic rules to assess whether pregnancy registrations and outcomes in the HDSS pertained to the same pregnancy observed in the ANC records. ANC pregnancies were matched to HDSS records if (i) the registration occurred between the individual's LMP and EDD, and if there was

an outcome, it was no more than 36 weeks after the first ANC visit, (ii) the outcome was between 8 weeks before and 20 weeks after the EDD, and the total duration from pregnancy registration or first ANC visit (whichever was earliest) to outcome was no more than 36 weeks, (iii) the outcome was between the first and last ANC visit for a given pregnancy, and if there was a pregnancy registration, the total duration between pregnancy registration and the last ANC visit was no more than 36 weeks.

Some of the matches meeting the above criteria had HDSS pregnancy outcomes which predated ANC visits for a given pregnancy. This was likely due to pregnancy outcome date misreporting. Reported dates of pregnancy outcomes may be subject to recall errors and rounding, especially if the outcome occurred early on in the interval between data collection rounds, and several months passed before the next household interview. Dates of pregnancy registrations, on the other hand, are more reliable as they simply indicate whether the individual in question was pregnant on the date of the household interview. As such, the matching criteria were set to be sufficiently lenient to allow for some date reporting errors in pregnancy outcomes, but not pregnancy registrations.

We calculated the proportions of pregnancies observed in the ANC clinics that were matched to pregnancy registrations and pregnancy outcomes in the HDSS. For those matched to pregnancy outcomes, we calculated the rates of stillbirth, perinatal, and neonatal mortality by pregnancy registration status, and for all pregnancies combined. Perinatal mortality compromises stillbirths and deaths occurring within the first week of life (i.e. early neonatal period). The denominators for the rates of stillbirth and perinatal mortality were total births. For neonatal mortality, deaths and exposure time were aggregated for the first 28 days of life. The abridged life table mortality rate was converted to the probability of dying by exact age 28 days, known as $q(28d)$. In some cases, a pregnancy was reported to have resulted in a live birth, but the child was not enumerated in the HDSS. The status of such newborns was considered unknown, and they were excluded from calculations of perinatal and neonatal mortality.

For pregnancies observed in the ANC records that were matched to HDSS pregnancy registrations, we assessed the timing of registration relative to care seeking and gestational age. Among those matched to a miscarriage or stillbirth in the HDSS, we evaluated potential misclassification between the two using a threshold of 28 weeks gestation, consistent with the International Classification of Diseases 10 system (World Health Organization, 2013). HDSS pregnancy outcomes were also used to examine uncertainty in ANC estimates of gestational age. We compared the ANC clinic estimates of gestational age against that which could be inferred from the date of the pregnancy outcome, under the assumption that the pregnancy came to term at 40 weeks gestation

Characteristics associated with missing pregnancies

For linked cases that were missing pregnancy outcomes in the HDSS, we investigated whether the individual was residing in the site for the data collection round immediately following their EDD. We identified individuals that had out-migrated, died, or been lost to follow-up prior to this data collection round, and those whose households had not yet been visited for data collection. These missing outcomes were distinguished from cases where the individual continued to reside in the site, and had no pregnancy outcome despite the occurrence of a household interview following their EDD. This was considered evidence of a potential unreported adverse pregnancy outcome. We investigated the characteristics of such individuals using Chi-squared tests and logistic regression.

Covariates of interest from the HDSS included age, area of residence, duration of residency, education level, household wealth, and marital status. We generated a household wealth index from a principal component analysis of variables denoting socio-economic status including type of toilet facility, water supply source, and ownership of assets such as a radio, television, and cooking stove (Amek et al., 2015). From the clinical data, we included information from the ANC register on total number of ANC visits for the given pregnancy (taken from a pregnancy-specific visit order number), gestational age at first visit, parity, and HIV status. Variables strongly associated with having a missing pregnancy outcome in Chi-squared tests were added to multivariable logistic regression models. Variables were removed from the model if they did not contribute significantly to model fit, as measured by Akaike Information Criterion. Age, household wealth quintile, number of ANC visits, and parity were tested as both categorical and continuous variables.

5.4 Results

The ANC dataset consisted of 6,626 individuals who had visited clinics since the start of record linkage in February 2018 to December 2020 (flowchart: Figure 5.1). There were 5,794 individuals that self-reported to reside in the HDSS area. From this group, 3,173 were linked to their record in the HDSS through PIRL, resulting in an overall match rate of 54.8%. Matches were independently verified by three HDSS personnel through manual review, and no false matches were identified.

The match rate was highest (1,162/1,867; 62.2%) for individuals who were enrolled in the study between February and December 2018. The match rate was 52.9% (1,497/2,830) for those enrolled from January 2019 to March 2020. Linkage activities were interrupted in the third week of March 2020 due to the COVID-19 pandemic. Once linkage was resumed in September 2020, the match rate was 46.9% (514/1,097) for those enrolled prior to December 2020. Women who were younger, HIV-negative, in union, and had given birth to fewer children were less likely to be matched to

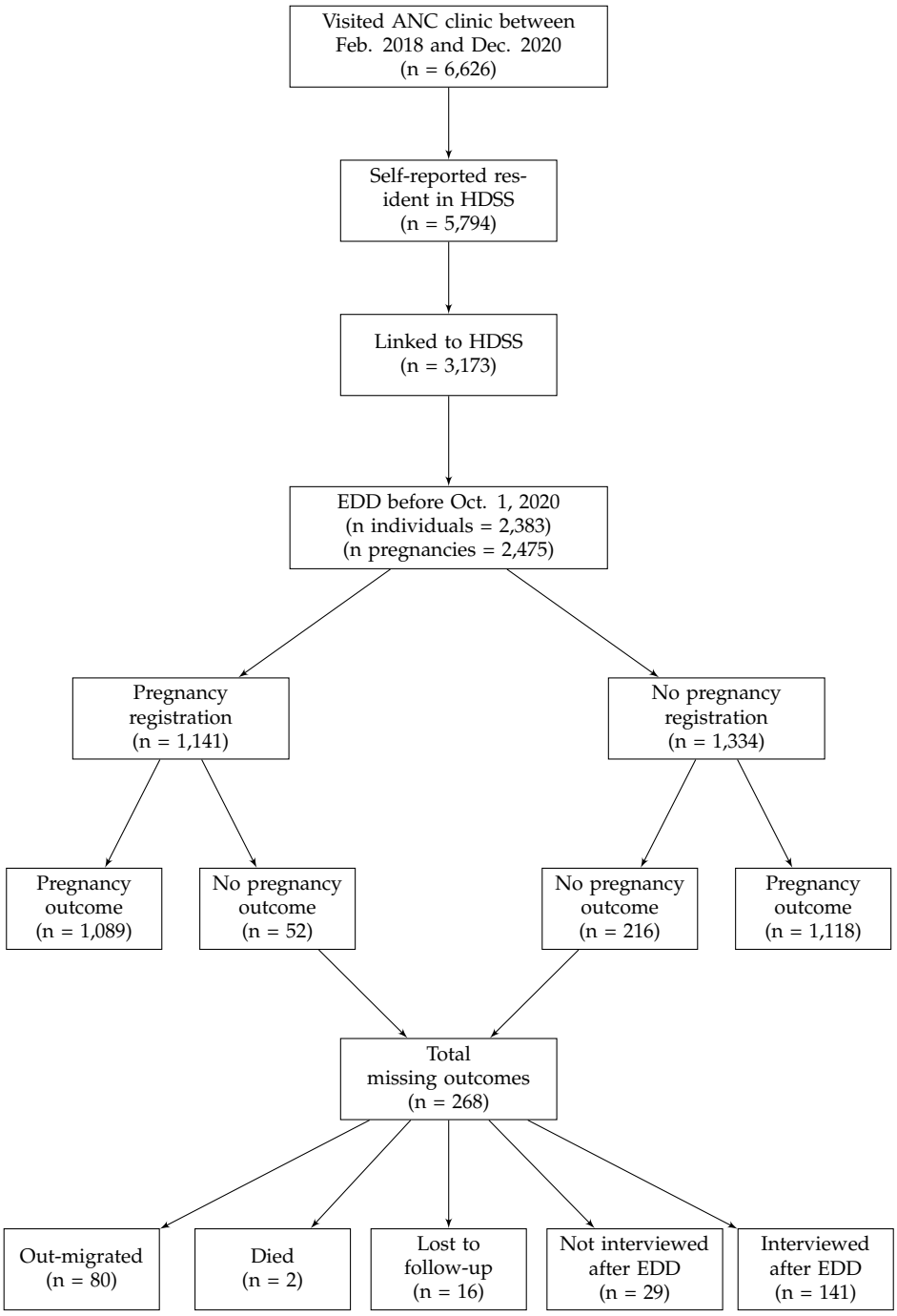


FIGURE 5.1: Flowchart of the sample included in the analysis, with information on the number of pregnancies observed in the antenatal care (ANC) clinics that had pregnancy registrations and outcome reports in Siaya Health and Demographic Surveillance System (HDSS). For pregnancies that were missing outcomes in the HDSS, information is provided on the status of the individual in question in the HDSS as of the first data collection round that occurred following their expected delivery date (EDD).

the HDSS. Those residing in Gem were more likely to be matched than residents of Asembo and Karemo. More details on the match rates stratified by socio-demographic characteristics are provided in appendix Table D.1.

5.4.1 External validation of HDSS pregnancy reporting

Among the individuals with linked HDSS-ANC records, there were 2,475 pregnancies with EDDs prior to October 1, 2020, attributed to 2,383 individuals. From this total, there were 1,141 (46.1%) pregnancies that were registered in the HDSS. Among those that were registered, 1,089 (95.4%) also had reported pregnancy outcomes. There were 1,334 (53.9%) pregnancies that were not registered in the HDSS. In this group, missing outcomes were more common (n = 216; 16.2%). In total, 268 pregnancies were missing outcome reports in the HDSS.

For pregnancies with missing outcomes, we investigated the residency status of the women in question in the HDSS as of the first household interview following their EDD. In 29.9% (n = 80) of cases, the individual had out-migrated from the HDSS area prior to the occurrence of a post-delivery data collection round. There were two individuals (0.01%) who died prior to the occurrence of a post-delivery data collection round and 16 (6.0%) who were lost to follow-up. For 29 (10.8%) cases, the individual's household had not been visited in a routine HDSS interview round following their EDD. For the largest share (n = 141; 52.6%), the individual was residing in the HDSS during the first household interview following their EDD, and no outcome was reported.

HDSS reporting for pregnancies observed in the ANC clinics is shown in greater detail in Table 5.1. After excluding cases where a woman's missing pregnancy outcome was attributable to out-migration, death, censoring, or the lack of a subsequent household interview; the remaining missing outcomes consisted of those identified as likely cases of unreported adverse pregnancy outcomes. The share of such outcomes was smaller for pregnancies that were registered in the HDSS (n = 11; 1%) compared to those that were not (n = 130; 10.4%), and made up 6% (n = 141) of the total analytical sample.

Reporting of adverse pregnancy outcomes was more common for pregnancies that had been registered in the HDSS. There were three miscarriages and 14 stillbirths among registered pregnancies. This yielded a stillbirth rate of 12.6 per 1,000 total births (95% confidence interval [CI] 6.3 − 18.9) and perinatal mortality rate of 27.1 (95% CI 18.1 − 37.1). There was one miscarriage and no reported stillbirths among pregnancies lacking registration, which had a perinatal mortality rate of 14.0 (95% CI 7.9 − 21.0). Neonatal mortality for registered pregnancies was 18.4 per 1,000 (95% CI 11.0 − 25.8), compared to 20.2 (95% CI 12.3 − 29.2) for those that were not registered. In both samples, close to three-quarters of neonatal deaths took place in the first week of life.

TABLE 5.1: Distribution of HDSS pregnancy reporting for women with linked ANC records.

	Pregnancy registration					
	Yes		*No*		*All*	
	n	(%)	n	(%)	n	(%)
Pregnancy outcome reporting						
Missing - potential adverse	11	(1.0)	130	(10.4)	141	(6.0)
Reported	1089	(99.0)	1118	(89.6)	2207	(94.0)
Total	1100	(100.0)	1248	(100.0)	2348	(100.0)
Pregnancy outcomes						
Live birth	1097[a]	(98.4)	1144[b]	(99.9)	2241	(99.2)
Miscarriage	3	(0.3)	1	(0.1)	4	(0.2)
Stillbirth	14[c]	(1.3)	0	(0.0)	14	(0.6)
Total	1114	(100.0)	1145	(100.0)	2259	(100.0)
Neonate status						
Early neonatal death	16	(1.5)	16	(1.4)	32	(1.4)
Late neonatal death	4	(0.4)	7	(0.6)	11	(0.5)
Unknown	9	(0.8)	12	(1.0)	21	(0.9)
Survived	1068	(97.4)	1109	(96.9)	2177	(97.1)
Total	1097	(100.0)	1144	(100.0)	2241	(100.0)
Mortality estimates	*Est.*	*(95% CI)*	*Est.*	*(95% CI)*	*Est.*	*(95% CI)*
Stillbirth rate	12.6	(6.3, 18.9)	0.0	(0.0, 0.0)	6.2	(3.1, 9.8)
Perinatal mortality rate	25.2	(18.2, 32.1)	11.3	(6.8, 15.8)	18.1	(13.9, 22.3)
q(28d)	18.4	(11.0, 25.8)	20.2	(12.3, 29.2)	19.3	(13.9, 25.2)

Notes:
Bootstrap 95% confidence intervals (CI) calculated from 10,000 samples with replacement. Stillbirth and perinatal mortality rates shown per 1,000 births, q(28d) shown per 1,000 live births.

[a] Includes 49 multiple births
[b] Includes 52 multiple births
[c] Includes 1 multiple birth

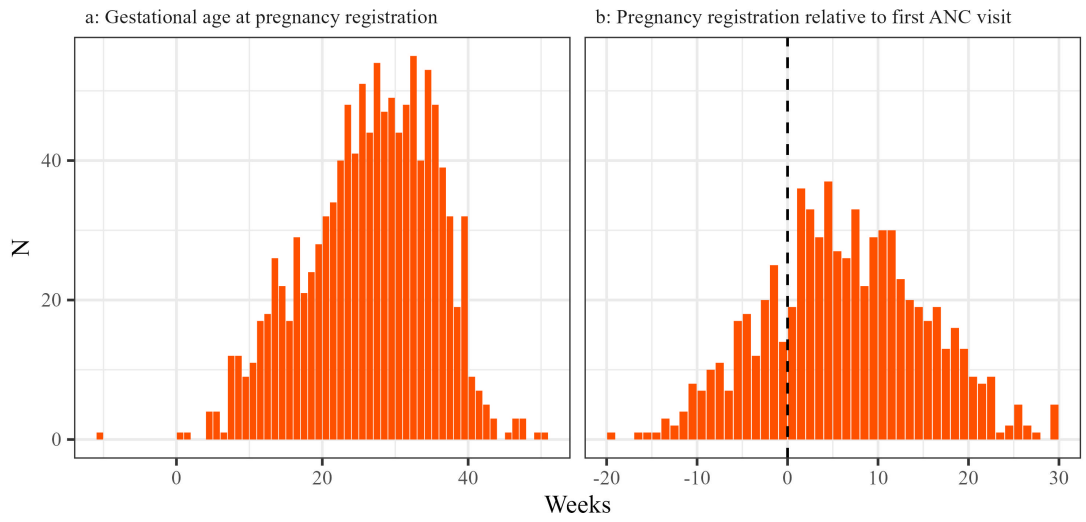


FIGURE 5.2: Timing of pregnancy registration in HDSS relative to gestational age and seeking antenatal care (ANC). (a) Histogram of gestational age at the time of HDSS pregnancy registration (n = 1,141). Gestational age was calculated from the ANC register using the value from the individual's latest recorded ANC clinic visit. (b) Histogram of number of weeks between HDSS pregnancy registration and the individual's first ANC clinic visit for the pregnancy. Pregnancies where the first ANC visit preceded the start of PIRL or those missing information for visit number were excluded (n = 446). Negative values indicate that the pregnancy was registered in the HDSS prior to the individual's first ANC clinic visit (represented by the vertical dashed line), while positive values show the reverse.

For pregnancies in the ANC records that were registered in the HDSS, Figure 5.2a displays the timing of registration relative to the individual's gestational age. The median gestational age at pregnancy registration in the HDSS was 26.4 weeks, and half of all pregnancy registrations took place between 20 and 33 weeks gestation. For 19 cases (1.7%), pregnancy registration was estimated to have taken place at implausible gestational ages, such as negative values or at more than 42 weeks gestation. These unrealistic values were indicative of error in the ANC clinic estimates of gestational age. Figure 5.2b shows the timing of pregnancy registration relative to the individual's first ANC clinic visit for the given pregnancy. There were 446 pregnancies that were excluded from this assessment due to the individual's first ANC visit preceding the start of the study or there being missing data for ANC visit number. For the remaining 695 cases, 76.7% (n = 533) received ANC prior to registering the pregnancy in the HDSS. Close to 62.2% (n = 432) of pregnancy registrations took place in the 16 weeks following the first ANC visit, and 14.5% (n = 101) were registered more than 16 weeks later.

Figure 5.3 shows the results of the investigation into misclassification of miscarriages and stillbirths in the HDSS. Among the reported stillbirths, 92.9% (n=13) had

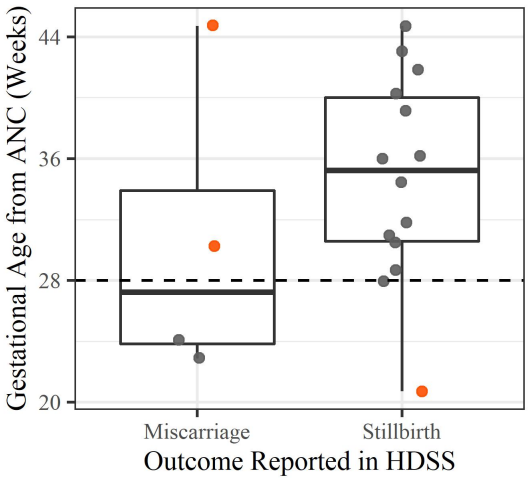


FIGURE 5.3: Gestational age for adverse pregnancy outcomes reported in the Health and Demographic Surveillance System (HDSS). Gestational age at the date of the pregnancy outcome was calculated from the latest recorded value in the ANC data. ANC-estimated gestational age has been plotted separately for pregnancies reported as miscarriages and stillbirths in the HDSS. The horizontal dashed line indicates the 28 week gestational age threshold used to distinguish between miscarriage and stillbirth.

gestational ages of >28 weeks, indicating accurate classification. There were fewer reported miscarriages (n=4), though only half occurred prior to 28 weeks gestation. The other two reported miscarriages appear to have been stillbirths that were incorrectly classified.

Gestational age in the ANC records tended to correspond fairly well with reported dates of HDSS pregnancy outcomes. In Figure 5.4a, we compared gestational ages recorded in the ANC data to that which could be inferred from pregnancy end dates in the HDSS. For the first ANC visit for a given pregnancy, the median ANC estimate for gestational age was 1.9 weeks lower than inferred gestational age (interquartile range [IQR] = 5.3). This decreased to a median difference of 1.6 (IQR = 4.1) for the fifth ANC visit, and 0.3 (IQR = 2.6) for the eighth visit (n = 25 pregnancies). In Figure 5.4b, the ANC record gestational age was plotted against the number of weeks to the HDSS pregnancy end date. A linear regression was fit to the data, showing that with each real increase of one week, the recorded gestational age in the ANC register increased by an average of approximately 0.85 weeks (95% CI $0.83 - 0.85$).

5.4.2 Characteristics associated with missing pregnancies

Table 5.2 presents descriptive statistics for women without a recorded pregnancy outcome in the HDSS, despite their continued residence during a household interview occurring after their EDD. The variables for number of ANC clinic visits, age, and

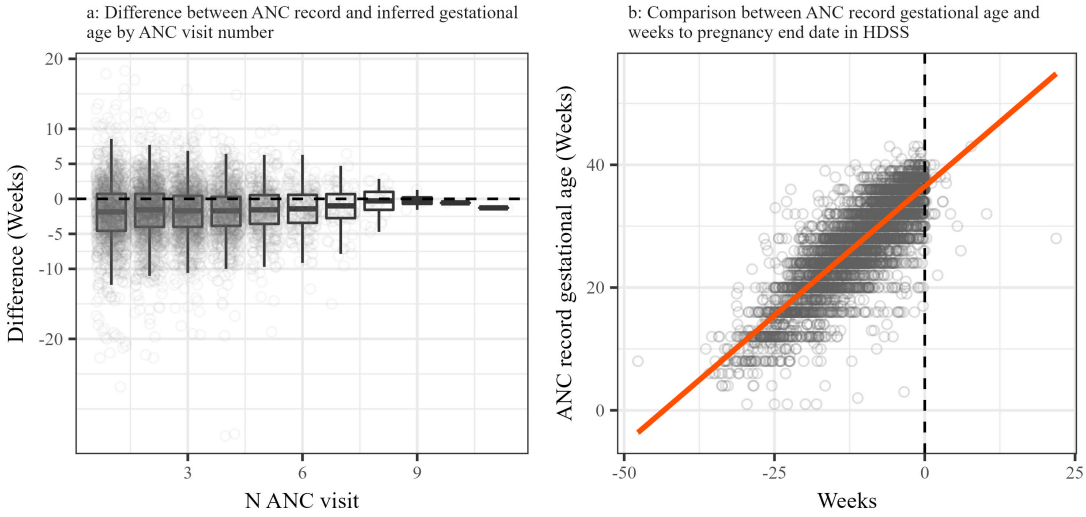


FIGURE 5.4: Comparison of gestational age from antenatal care (ANC) register and gestational age inferred from the pregnancy end date in the Health and Demographic Surveillance System (HDSS). (a) Difference between the gestational age recorded in ANC records and the gestational age inferred from the pregnancy end date (assuming a 40 week pregnancy) by ANC visit number. The horizontal dashed line denotes absolute agreement between ANC-estimated gestational age and the gestational age inferred from the HDSS pregnancy end date. Pregnancies missing outcomes or those missing visit numbers were excluded (n = 354). (b) Comparison between gestational age from ANC register and weeks to pregnancy end date in HDSS. The pregnancy end date is represented by the vertical dashed line. The orange line has been plotted from a linear regression of weeks to pregnancy outcome in the HDSS on ANC-estimated gestational age.

parity were strongly associated with having a missing pregnancy outcome. Underreporting was more common among those making a single as opposed to multiple ANC visits. Women who were in the youngest and oldest age groups also had proportionately more missing outcomes than those aged 20-34 years. Women who had no previous births were more likely to have missing outcomes, as well as those for whom parity was not known.

The variables of gestational age at first ANC visit, HIV status, and household wealth had moderate to weak associations with HDSS pregnancy outcome reporting. Those making their first ANC visit during the first trimester were more likely to be missing pregnancy outcomes in the HDSS. The same was true for women who were HIV-positive or of unknown status. Women belonging to households in the top three wealth quintiles had a higher proportion of missing pregnancy outcome reports than those in the bottom quintiles, though there was not a clear pattern across quintiles. Duration of residency in the HDSS, education, and region of residency within the HDSS were not significantly associated with having a missing pregnancy outcome.

Characteristics associated with HDSS pregnancy outcome reporting were investigated further using logistic regression. The results of the full (1) and final (2) models are shown in Table 5.3. Controlling for all other variables in the full model, age was not significantly associated with having a missing pregnancy outcome, and was excluded from the final model. In the case of parity, women with an unknown number of previous births had 5.82 (95% CI $1.84 - 17.37$) times the odds of HDSS pregnancy outcome underreporting compared to those with one previous birth. There were not significant differences in the odds of pregnancy outcome reporting for individuals of known parity, and this variable was also left out of the final model.

In the final model, a one visit increase in an individual's total number of ANC visits for a given pregnancy was associated with 30% lower odds of pregnancy outcome underreporting (odds ratio [OR] 0.70, 95% CI $0.60 - 0.80$). Additionally, women making their first ANC visit during the first trimester had 2.64 (95% CI $1.52 - 4.51$) times the odds of pregnancy outcome underreporting compared to those making their first visit in the second trimester. Those testing HIV-positive during an ANC clinic visit had 86% higher odds of having a missing pregnancy outcome in the HDSS compared to those that tested HIV-negative (OR 1.86, 95% CI $1.22 - 2.78$). Compared to those in the median household wealth quintile, the odds of having a missing pregnancy outcome were 79% lower for women in the fourth quintile (OR 0.21, 95% CI $0.05 - 0.73$). In the first model, the odds of pregnancy underreporting were 54% higher for unmarried women compared to those in formal union (OR 1.54, 95% CI $0.91 - 2.60$). This increased to an odds ratio of 2.10 (95% CI $1.45 - 3.04$) in the final model.

TABLE 5.2: Characteristics of linked women that were missing pregnancy outcome reports in the HDSS. Table provides row percentages, denoting the share of pregnancies observed in the ANC records that were missing outcomes in the HDSS out of the total number that were assessed for reporting.

		Missing pregnancy outcomes		
Variable	Value	n	(%)	p-value
ANC clinic visits	1	57	(17.2)	<0.01
	2	27	(6.1)	
	3	22	(4.4)	
	4+	35	(3.3)	
Age	10-19	34	(9.4)	0.01
	20-24	36	(6.6)	
	25-29	25	(4.6)	
	30-34	20	(4.0)	
	35+	26	(6.4)	
Duration of residency in HDSS	<2 years	27	(6.7)	0.59
	2+ years	114	(5.9)	
Education	None	1	(5.3)	0.99
	Primary	102	(6.0)	
	Secondary/Higher	37	(5.9)	
	Unknown	1	(11.1)	
Gestational age at 1st ANC visit	1st trimester	24	(10.3)	<0.01
	2nd trimester	44	(4.9)	
	3rd trimester	18	(5.7)	
	Unknown	55	(6.1)	
HDSS region	Asembo	11	(6.2)	0.98
	Gem	126	(6.0)	
	Karemo	4	(5.6)	
HIV status	Negative	100	(5.4)	0.06
	Positive	36	(8.1)	
	Unknown	5	(13.2)	
Household wealth quintile	1	21	(6.1)	0.02
	2	17	(11.1)	
	3	10	(8.7)	
	4	3	(2.2)	
	5	6	(4.4)	
	Unknown	84	(5.7)	
Marital status	In union	83	(4.7)	<0.01
	Not in union	58	(9.8)	
Parity	0	46	(8.7)	<0.01
	1	22	(6.1)	
	2	66	(4.6)	
	Unknown	7	(35.0)	
Total		141	(6.0)	

Notes:
P-values display the results of Chi-squared tests of independence. Tests were performed on complete case data, with "Unknown" values excluded.

TABLE 5.3: Logistic regression results for having a missing pregnancy outcome in the HDSS. Sample only includes individuals with missing outcomes who were resident in the HDSS for the first household interview following their EDD.

	Dependent variable: Missing pregnancy outcome			
	(1)		(2)	
Variable	OR	(95% CI)	OR	(95% CI)
Age				
10-19	1.26	(0.72, 2.24)		
20-24	1			
25-29	0.93	(0.50, 1.71)		
30-34	0.79	(0.39, 1.57)		
35+	1.11	(0.57, 2.20)		
ANC clinic visits	0.71	(0.62, 0.82)	0.70	(0.60, 0.80)
Gestational age at 1st ANC visit				
1st trimester	2.70	(1.55, 4.62)	2.64	(1.52, 4.51)
2nd trimester	1		1	
3rd trimester	0.89	(0.49, 1.58)	0.87	(0.48, 1.53)
Unknown	0.85	(0.55, 1.32)	0.94	(0.61, 1.45)
HIV status				
Negative	1		1	
Positive	1.89	(1.20, 2.94)	1.86	(1.22, 2.78)
Unknown	2.39	(0.78, 6.00)	2.19	(0.72, 5.45)
Household wealth quintile				
1	0.60	(0.27, 1.40)	0.58	(0.27, 1.35)
2	1.27	(0.55, 3.08)	1.18	(0.51, 2.83)
3	1		1	
4	0.22	(0.05, 0.75)	0.21	(0.05, 0.73)
5	0.48	(0.16, 1.37)	0.47	(0.15, 1.35)
Unknown	0.67	(0.34, 1.46)	0.66	(0.34, 1.41)
Marital status				
In union	1		1	
Not in union	1.54	(0.91, 2.60)	2.10	(1.45, 3.04)
Parity				
0	1.08	(0.56, 2.11)		
1	1			
2	0.83	(0.46, 1.55)		
Unknown	5.82	(1.84, 17.37)		
Observations	2,348		2,348	
Log Likelihood	−487.66		−495.23	
Akaike Inf. Crit.	1,015.32		1,016.46	

Notes:
OR - odds ratio, CI - confidence interval.

5.5 Discussion

In this work, we compared individually-linked ANC clinic data with records of pregnancies and their outcomes in Siaya HDSS. HDSS are often set up in regions where availability of other population health data is limited (Sankoh and Byass, 2012) and this offered a unique opportunity to externally validate HDSS pregnancy reporting completeness, identify predictors of underreporting, and examine potential bias in the reporting of adverse outcomes. Linkage with ANC is a particularly high value extension of HDSS data, given the almost universal coverage of ANC services in many parts of sub-Saharan Africa. It has been estimated that close to 90% of women in sub-Saharan Africa and 94% in Siaya HDSS receive ANC at least once during pregnancy (Adedokun and Yaya, 2020; Amek et al., 2015).

Of the 2,475 pregnancies observed in ANC records that were included in the analysis, there were 268 that did not have a corresponding pregnancy outcome report in Siaya HDSS. Close to 36% of these missing outcome reports were attributable to the individual exiting the study area or being lost to follow-up prior to the next HDSS data collection round. A household interview had not yet taken place in another 11% of cases. The remaining 53% did not have a corresponding pregnancy outcome report in the HDSS, despite the occurrence of a household interview following their expected delivery. This subset of missing outcomes is cause for concern, as such underreporting is more likely when an adverse pregnancy outcome or early death has occurred.

It is worth noting that outcomes which were not reported at the first interview following delivery could still be reported in later data collection rounds. However, delayed reports are only likely for pregnancies that were registered or resulted in a live birth. In the case of the former, the HDSS fieldworker will be prompted to inquire about the status of the pregnancy until an outcome is recorded or the individual is no longer resident in the study site. Alternatively, in the case of a live birth, the presence of a new child in the household helps to ensure that the birth is recorded when the child is enumerated in the HDSS. Otherwise, the likelihood that adverse outcomes are reported diminishes with time, amplifying the downward bias in measurements of perinatal and neonatal mortality.

Our results suggest that pregnancy registration can improve ascertainment of adverse outcomes that are otherwise vulnerable to underreporting. The observed rate of stillbirths for pregnancies that were registered in the HDSS was 12.6 (95% CI 6.3 − 18.9), while zero stillbirths were reported among those lacking pregnancy registration. The observed neonatal mortality for registered and unregistered pregnancies was similar, though significant differences in neonatal mortality by pregnancy registration status have been found in other research (Eilerts et al., 2021b; Nareeba et al., 2021). If such differences were attributable to differential usage of ANC services, this could explain the lack of significant difference in our own analysis, where the sample was entirely composed of pregnancies for which the mother had accessed ANC. It is

also possible that the lack of a significant difference in neonatal mortality in this work was related to the relatively small number of observations.

We found that women who began ANC in the first trimester and had fewer visits overall were more likely to be missing pregnancy outcomes in the HDSS. It is possible that some of these pregnancies were not carried to full term, and such women made fewer ANC visits as a result. This proposition is further supported by the finding that women who were unmarried and HIV-positive were more likely to have missing pregnancy outcomes. In both cases, these characteristics are independently associated with elevated risk of perinatal mortality (Aminu et al., 2014; Kupka et al., 2009), as well as under-5 mortality (Clark and Hamplová, 2013; Marston et al., 2011; Zaba et al., 2003). However, it could also be the case that women with fewer clinic visits out-migrated from the study area midway through pregnancy, and were thus also less likely to have reported pregnancy outcomes. While our evaluation was restricted to women who were resident in the HDSS during their ANC usage and for the first household interview following their EDD, some absences from the study area may not have been accurately reflected in HDSS records.

Tracking frequent in- and out-migrations that vary in their destination and duration is a complex task in HDSS (Sankoh and Byass, 2012). Pregnancy and the postpartum period can be a time of increased mobility, when women often travel to seek medical care or the support of family members (Clouse et al., 2013; Kaplan et al., 2008; Wang et al., 2011). If such migrations were not captured by the HDSS, it may have appeared some individuals were present in the site for the data collection round following their delivery when they were in fact not. As this relates to our results, it is possible that the share of missing pregnancy outcomes attributable to migration was larger than observed, and the share that were unreported adverse outcomes was smaller.

Nevertheless, our results are consistent with previous findings of downward bias in HDSS estimates of early mortality. The tendency for deaths occurring soon after birth to be under-counted in population-based surveillance data was noted as early as the 1950s and 60s (Billewicz and McGregor, 1981; Cantrelle, 1969). More recently, rates of neonatal mortality in African HDSS have been found to be lower than corresponding estimates from retrospective household surveys such as DHS and MICS (Eilerts et al., 2021a; Waiswa et al., 2019). Additionally, while both sources exhibit higher levels of child mortality (deaths between one and four years) than found in the historic record of high-quality data, HDSS are alone in their downward deviation at early ages (Verhulst et al., 2021). Reliably monitoring pregnancy outcomes and early mortality has been identified as one of the most serious challenges faced by HDSS (Sankoh and Byass, 2012).

The magnitude of downward bias in HDSS estimates of adverse pregnancy outcomes and early mortality ultimately depends on the rate of such events among missing pregnancy outcomes, which cannot be determined without additional data collection. This is an area in which record linkage with ANC data could be very useful. Leveraging information on ANC usage to organize follow-up interviews with those who are missing pregnancy outcome reports is an important avenue for future research. Such work has been piloted using record-linked ANC data in HDSS in The Gambia, where it identified elevated rates of perinatal and neonatal mortality among pregnancies that were missing outcome reports in the HDSS (Rerimoi, 2019).

Ongoing record linkage between HDSS and ANC clinics could allow for earlier pregnancy detection and more accurate outcome reporting. We found that individuals in Siaya HDSS had visited an ANC clinic prior to registering pregnancies in the site in approximately 77% of cases. Additionally, ANC clinic information on gestational age indicated that half of reported miscarriages were potentially misclassified stillbirths. However, it is worth noting that the assessment of misclassification between miscarriage and stillbirth in the HDSS relied on the accurate estimation of gestational age at the ANC clinics. Given the uncertainty of estimating gestational age through fundal height palpitation and reported LMP (White et al., 2012), it is important to interpret these results cautiously.

This study is subject to a few important limitations. First, it would be beneficial to conduct follow-up data collection on pregnancies that were observed in the ANC but missing outcome reports in the HDSS. This analysis identified pregnancies that were likely to have ended in adverse outcomes given that they were not reported in a data collection round occurring after the EDD. This shed light on the potential for bias in HDSS estimates of adverse pregnancy outcomes and early mortality, however, additional information is needed in order to provide a more precise estimate of its magnitude. It is also important to acknowledge that close to 45% of women seeking ANC who self-reported residence in the HDSS were not successfully linked to their record. While some characteristics associated with lower HDSS-ANC match rates were also associated with having a missing pregnancy outcome in the regression analysis (e.g. being younger, seeking ANC services in the first trimester), others were positively associated with pregnancy outcome reporting (e.g. being HIV negative, in union). It is thus unclear how an improvement in linkage rates would affect the proportion of pregnancy outcomes missing from HDSS records.

5.6 Conclusion

HDSS are valuable sources of empirical demographic and epidemiological data for much of sub-Saharan Africa where current systems of CRVS and HMIS are deficient. However, HDSS data on pregnancies, pregnancy outcomes and neonatal mortality is

often incomplete. This research demonstrates the potential of using record linkage with ANC clinics to evaluate pregnancy reporting completeness in HDSS and investigate bias in estimates of adverse pregnancy outcomes and early mortality. Record linkage between HDSS and routine programme data is an efficient manner of augmenting population health information in sub-Saharan Africa, and addressing the lack of good quality data on pregnancy and early mortality. Such efforts have the potential to both improve our understanding of population health and our ability to accurately measure it.

Chapter 6

HIV-infected children and missing mothers: an empirical investigation of bias in under-5 mortality estimates from birth history survey data

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1600573	**Title**	Ms
First Name(s)	Hallie		
Surname/Family Name	Eilerts		
Thesis Title	Under-five mortality estimates for sub-Saharan Africa: an inquiry into data sources and estimation methods		
Primary Supervisor	Georges Reniers		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Undecided
Please list the paper's authors in the intended authorship order:	Eilerts, Hallie; Eaton, Jeffrey, W.; Romero Prieto, Julio; Reniers, Georges
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the study in collaboration with Jeffrey W. Eaton. I conducted all data analysis, interpreted findings, and wrote the first draft of the manuscript. All co-authors have provided input on the analysis and comments on the manuscript.

SECTION E

Student Signature	Hallie Eilerts
Date	21 March 2022

Supervisor Signature	Georges Reniers
Date	21 March 2022

6.1 Abstract

INTRODUCTION

Birth history questionnaires in large surveys serve as the foundation for under-5 mortality (U5M) estimation in many low- and lower-middle-income countries. In high HIV settings, such estimates can be biased by the correlation of HIV-related mortality of mothers and their children. Little is known about how such bias has evolved over the course of the HIV epidemic and the rollout of programs to mitigate its mortality impact.

OBJECTIVE

We first assessed the link between maternal HIV status/antiretroviral treatment (ART) usage and child survival over time in Somkhele Health and Demographic Surveillance System (HDSS) in South Africa. We then empirically investigated bias from missing mothers from 1998-2017 using data from 12 HDSS in the International Network for the Demographic Evaluation of Populations and their Health (INDEPTH Network).

METHODS

In Somkhele HDSS, we used survival analysis and demographic decomposition techniques to assess child survival by maternal HIV/ART status. We simulated a series of cross-sectional surveys in the INDEPTH Network HDSS, and calculated estimates of U5M for the children of resident mothers. These were compared to the U5M of all children to gain insight on bias in estimates of U5M derived from birth history data. Additional comparisons were made with U5M estimates for children whose mothers were not present due to death or out-migration.

RESULTS

Children born to HIV-positive mothers in Somkhele HDSS had a significantly elevated hazard of U5M compared to those with HIV-negative mothers (hazard ratio 3.01, 95% confidence interval $2.25 - 4.03$), which was even higher for those born in [2003, 2007) compared to [2007, 2011). There was no significant difference in survival between children with HIV-negative mothers and HIV-positive mothers on ART. Bias in U5M estimates from simulated surveys in the INDEPTH Network HDSS was highest for retrospective periods of 0-4 years in sites in South Africa (median = 18%), followed by eastern and western Africa (medians of 15% and 8%, respectively). Bias tended to be higher for surveys prior to 2011 in eastern and southern Africa, as well as more distant retrospective periods in the latter region. The death of a mother within the first five years of life resulted in increased risk of U5M in all sites.

CONCLUSIONS

The association between maternal HIV status and child survival has weakened in Somkhele HDSS in the ART era. In high HIV settings, bias from missing mothers in survey derived estimates of U5M has likely declined in the past decade. However, bias correction remains necessary and estimates should be limited to short periods of retrospection.

6.2 Introduction

The United Nations Sustainable Development Goals (SDGs) call for ending preventable deaths of infants and children under five years of age by 2030 (United Nations, 2017). Countries facing the highest burden of under-5 mortality (U5M) often lack the high-quality data necessary for tracking and accelerating progress towards its reduction. Civil registration and vital statistics (CRVS) systems are incomplete throughout sub-Saharan Africa, the region where the majority of global under-5 deaths take place. In such settings, retrospective birth history questionnaires have been instrumental to monitoring levels and trends in U5M.

Full birth history (FBH) questionnaires collected in nationally representative surveys such as the Demographic and Health Surveys (DHS) or Multiple Indicator Cluster Surveys (MICS) are the primary data underpinning mortality estimates for African countries (UN IGME, 2021; Wang et al., 2014). In a FBH, women of reproductive age are asked to report on all live births with information regarding the date of the birth, the survival status of children, and if applicable, the child's age at death. This retrospective information can be used to directly calculate mortality estimates using the dates of deaths and the time that children were exposed to the risk of dying.

In order for U5M estimates from FBHs to be unbiased, the births on which data are collected must be representative of all births in the period of interest (Rutstein and Rojas, 2006; Walker, Hill, and Zhao, 2012). Though this condition can never be perfectly met, as the births of mothers who have died or migrated are not included, any bias introduced by this is typically assumed to be negligible (Walker, Hill, and Zhao, 2012). However, this assumption is less tenable in settings with generalized HIV epidemics. If reproductive age mortality from HIV/AIDS is high, surviving HIV-negative women will be over-represented in survey samples. Additionally, vertical transmission of HIV during pregnancy, delivery, and breastfeeding diminishes the survival prospects of children born to HIV-positive women. Thus, HIV-positive children will be more likely to die than other children, and their deaths will also be less likely to be reported given the elevated mortality risks of their mothers

Most research investigating HIV-related bias in U5M estimates from retrospective birth histories has been conducted using mathematical models that rely on predicted trajectories of HIV incidence, fertility and mortality rates, and more recently, antiretroviral treatment (ART) coverage (Artzrouni and Zaba, 2003; Zaba, Marston, and Floyd, 2003; Mutemaringa, 2011; Quattrochi et al., 2019). These time-varying and inter-dependent dynamics are difficult to estimate, and many settings lack sufficient empirical data to inform these models. To our knowledge, the only empirical investigation of HIV-related bias in U5M estimates from FBHs was conducted using data from Zimbabwe in 1998-2005 (Hallett et al., 2010). This study found downward bias of approximately 10% from missing mothers, and served as the basis for the UN IGME bias adjustment procedure for U5M estimates arising from FBHs in high HIV

settings (Walker, Hill, and Zhao, 2012). Substantial evolution of the HIV epidemic and treatment options in the past two decades means that the nature of such bias has likely evolved since this time. Indeed, in 2020, UN IGME updated its bias adjustment procedure to account for the impact of ART on survival and vertical transmission (Johnson, Mizoguchi, and Pantazis, 2020).

Given the continued centrality of birth history questionnaires to U5M estimation in sub-Saharan Africa, accurate adjustment for HIV-related bias is of the utmost importance. Health and Demographic Surveillance Systems (HDSS) are valuable sources of longitudinal population-based data for investigating such bias empirically. HDSS collect detailed health and population dynamics data for circumscribed populations throughout sub-Saharan Africa. Data collection is conducted on a longitudinal basis, typically through recurring household interview rounds (Sankoh and Byass, 2012). Founded in 1998, the International Network for the Demographic Evaluation of Populations and their Health (INDEPTH Network) is the largest group of affiliated HDSS, with 49 field sites (Ngom, 2001; INDEPTH Network, 2021). HDSS have generated essential data on HIV incidence, ART scale-up, and mortality patterns over the course of the epidemic (Bor et al., 2013; Sankoh, 2017) which can shed light on how such factors impact U5M estimation in FBHs.

In this research, we leverage HDSS data to empirically investigate mother survivorship bias in estimates of U5M from retrospective birth histories in high HIV settings. The first part of the analysis utilizes demographic and HIV surveillance data from the Africa Health Research Institute's (AHRI; formerly known as the Africa Centre) Somkhele HDSS in South Africa. We assess the link between maternal HIV status and child survival over the course of the HIV epidemic and with the rollout of ARTs, and decompose changes in overall U5M into parts attributable to the changing composition of HIV-positive mothers and changes in child survival by maternal HIV status. We then bring together data from 11 additional HDSS sites throughout sub-Saharan Africa. We simulate cross-sectional surveys in the HDSS and calculate estimates of U5M for the children of surviving mothers. These are compared to estimates of U5M for all children residing in the HDSS to shed light on bias from missing mothers in high HIV settings. Additionally, we contrast the survival of children whose mothers were not present in the site due to death or out-migration.

This unique contribution provides new empirical evidence regarding how HIV-induced bias in U5M estimates has evolved over the course of the HIV epidemic and in the era of ARTs. Furthering understanding in this regard is crucial to informing bias-correction techniques applied to birth histories, which serve as the most widely used source of U5M data in many low- and lower-middle income countries (LLMICs).

6.3 Background

South Africa is one of the countries most heavily affected by HIV in the world, and was home to 20% of the total population estimated to be living with HIV in 2020 (UNAIDS, 2021). In 1990, South Africa's first national seroprevalence survey among pregnant women at antenatal clinics estimated HIV prevalence at 1% (Abdool Karim et al., 2009; Burton, Giddy, and Stinson, 2015). HIV infections increased exponentially in subsequent years, resulting in approximately 20% prevalence among pregnant women in 2000 and 30% in 2005 (Abdool Karim et al., 2009). The eastern provinces of the country have experienced the highest burden (Abdool Karim et al., 2009; Herbst et al., 2009; Welz et al., 2007). In Agincourt HDSS, located close to the Mozambican border in Mpumalanga Province, HIV/AIDS attributed mortality among young adults increased fivefold from the early 1990s to 2000s (Kahn et al., 2007). In Somkhele HDSS, located in KwaZulu-Natal province, close to half of all pregnant women aged 25-29 years were HIV-positive in 2003-2004 (Welz et al., 2007). According to recent estimates of seroprevalence among pregnant women, KwaZulu-Natal remains the province with the highest prevalence at approximately 41%, followed by Eastern Cape and Mpumalanga at 37% and 34%, respectively (Woldesenbet et al., 2021).

As the HIV epidemic escalated in many countries across sub-Saharan Africa in the 1990s, gains made in child mortality also stagnated (Adetunji, 2000; Ahmad, Lopez, and Inoue, 2000; Timaeus, 1998; Rutstein, 2000). A series of studies investigating the effect of maternal HIV on child survival found a strong correlation between maternal and child mortality (Cock et al., 1994; Nakiyingi et al., 2003; Ng'weshemi et al., 2003; Nicoll et al., 1994; Spira et al., 1999; Zaba et al., 2005). Maternal HIV status affects the survival of children through the vertical transmission of HIV from mother to child during pregnancy, delivery, and breastfeeding. Without treatment, the median survival of perinatally infected children has been estimated as two years (Mahy et al., 2017). Independent of increased mortality risk from mother-to-child transmission, children of HIV-positive mothers are also negatively affected by the indirect effects of maternal illness and death (Hunter, Twine, and Johnson, 2011; Nakiyingi et al., 2003; Ng'weshemi et al., 2003; Sartorius et al., 2011).

In the early 2000s, there was growing awareness of HIV-related bias in U5M estimates arising from birth histories due to high levels of reproductive age mortality and the strong correlation of maternal and child mortality. Initially, the magnitude of such bias was thought to be relatively small, especially in comparison to the prevailing uncertainty intervals of U5M estimates from DHS birth histories (Artzrouni and Zaba, 2003; Ng'weshemi et al., 2003). A systematic adjustment was considered unnecessary if direct estimates of U5M were calculated for recent retrospective periods and HIV prevalence of pregnant women was under 5% (Zaba, Marston, and Floyd, 2003; Mahy, 2003). Consensus regarding the magnitude of downward bias introduced by HIV shifted as the epidemic progressed (Hallett et al., 2010). By the late

2000s, summary birth histories (SBHs) relying on indirect estimation techniques were considered inappropriate for estimating U5M in high HIV settings (Ward and Zaba, 2008; Mutemaringa, 2011).

In 2010, Hallett et al. published an empirical investigation of HIV-related bias in estimates of U5M from FBHs which influenced UN IGME to adopt a correction factor for such estimates (Hallett et al., 2010; Walker, Hill, and Zhao, 2012). In the UN IGME adjustment procedure, a model was used to project the annual number of births occurring to HIV-positive and HIV-negative mothers at the national level, and calculate the proportion of children that would become infected with HIV. These projections took into account HIV prevalence, incidence, HIV fertility-reducing effects, and transmission rates from mother to child given breastfeeding practices and Prevention of Mother-to-Child Transmission (PMTCT) interventions (Stover, Brown, and Marston, 2012). The probability of HIV-negative children and their mothers surviving to a presupposed future survey date was then calculated through model life tables, while a mortality schedule previously derived from cohort studies was used for HIV-positive women and children (Walker, Hill, and Zhao, 2012). The ratio of under-5 deaths to births was calculated for mothers who survived to the date of the survey, and would have thus been able to report on their reproductive history. This quantity was compared to the true ratio of all under-5 deaths to births which would have been observed had none of the mothers died. The HIV-related bias for each period estimate was calculated as the ratio of *true* and *reported* ratios, and used to adjust estimates of U5M from FBHs in countries with generalized HIV epidemics (defined as prevalence of 5% or more among the adult population) (Walker, Hill, and Zhao, 2012). For DHS surveys taking place in southern Africa between 2004-2007, bias was estimated as 5-15% for the most recent five-year period, 11-27% for the 5-9 year window of retrospection, and 7-19% for the period 10-14 years prior (Walker, Hill, and Zhao, 2012). Bias for the 2003 DHS in Kenya was estimated as 7%, 10%, and 3% for the three retrospective periods, while the 2005 survey in Côte d'Ivoire had bias of 3% in the most recent period, and 5% and 2% for the periods 5-9 and 10-14 years prior, respectively.

As understanding of HIV-related bias in U5M estimates improved, HIV treatment options were also becoming more widespread. Botswana was the first African country to roll out a national HIV treatment program in 2002 (Farahani et al., 2014). Nationwide free provision of ART started in other countries in southern and eastern Africa between 2004 and 2007 (Houlihan et al., 2011; Wringe et al., 2012). Following the widespread availability of ART, there was a rapid reduction in the mortality of HIV-positive adults. In Somkhele HDSS, mortality among HIV-positive women fell 57% from 2003 to 2011 (Reniers et al., 2014). In 2017, 41% of HIV-infected adults in the site were using ART, and incidence of new HIV infections was estimated to have declined by 43% since 2012 (Vandormael et al., 2019). PMTCT programs consisting of a suite of HIV testing and ART protocols also produced substantial improvements in child

health. In South Africa, the vertical transmission rate was reduced from an estimated 25% in the absence of intervention to 2-3% (Goga et al., 2015).

There has thus been substantial evolution of the HIV epidemic and treatment options in sub-Saharan Africa in recent decades. Against this background, assessing HIV-related bias in U5M estimates from cross-sectional surveys is complex. While increases in HIV prevalence among reproductive age women likely exacerbated such bias in the pre-ART era, widespread treatment availability and the improved survival of HIV-positive women would later counteract bias by reducing the number of missing mothers. The reduction of vertical transmission through PMTCT would also reduce the excess mortality of children born to HIV-positive mothers. How these dynamics intersect and affect the overall magnitude of bias is unclear. Demographic and HIV surveillance data from HDSS throughout sub-Saharan Africa, in both high and low HIV settings, can serve as an invaluable resource for investigating these issues further.

6.4 Methods

6.4.1 Data

We first used data from Somkhele HDSS to assess the link between maternal HIV/ART status and child survival over the course of the HIV epidemic. In the second part of the analysis, this data was supplemented with data from other African HDSS to empirically investigate bias in U5M estimates from birth histories due to missing mothers.

Somkhele HDSS was founded in 2000, initially covering an area of 438 km^2 and population of 85,000 (Gareta et al., 2021). In 2017, the site was expanded to 845 km^2 and a population of 140,000 individuals. Since the site's inception, standard demographic information on the births, deaths, and migrations of all individuals residing in the HDSS has been collected through recurring household interviews. Interviews were conducted twice per year until 2011, and three times per year thereafter. The site began inviting eligible individuals to participate in annual anonymized HIV testing during household interviews in 2003. This initially included women 15-49 years and men 15-54 years, and was expanded to all residents aged 15 years or older in 2006. HIV testing has been conducted through the collection of a dried blood spot (Gareta et al., 2021). Data on self-reported ART usage for individuals testing positive, or who had previously tested positive, has been collected since 2006.

HDSS data for sites in the INDEPTH Network was primarily accessed through the iShare Data Repository[1]. We downloaded HDSS core datasets for sites in sub-Saharan Africa that were released between 2015 and 2019 on the iShare Data Repository. Sites were included in the analysis if they had more than 12 years of available data and a population of at least 10,000 individuals. The first condition was set to ensure that

[1] http://www.indepth-ishare.org

estimates of U5M could be calculated on a continuous basis for at least three five-year birth cohorts. Secondly, sites were excluded on the basis of population size given the relative rarity of mortality as an event, and the high variability of mortality estimates in small samples. In total, data was gathered from 11 sites in eight countries. This included Agincourt HDSS, which, together with Somkhele, were the only two sites in southern Africa. In East Africa, we included data from Siaya, Kenya; Karonga, Malawi; and Ifakara, Magu, and Rufiji in Tanzania. In West Africa, sites included Nouna, Burkina Faso; Navrongo, Ghana; Farafenni, The Gambia; and Bandafassi and Niakhar in Senegal. Data for Siaya HDSS was furnished independently by the Kenya Medical Research Institute.

6.4.2 Maternal HIV/ART status and child survival

In the Somkhele HDSS data, routine HIV testing was used to classify children by the HIV status of their mother at the time of birth. The mother was considered HIV-positive she ever tested positive up to six months postpartum. Mothers were labelled as HIV-negative if they had a negative test result from at least six months postpartum. In cases where mothers had an HIV-negative test result prior to delivery and an HIV-positive test result later than six months postpartum; we estimated that seroconversion took place at the midpoint of these two dates. If the estimated date of seroconversion was less than six months postpartum, the mother was considered HIV-positive at the time of birth. Maternal HIV status was considered unknown if it could not be determined from the available testing data. If maternal HIV status was positive, we used the available data to determine whether the mother was on ART at the time of birth. Mothers were classified as "On ART" if they self-reported to be on treatment from any point prior to delivery, continuing up to at least six months after delivery. Mothers who were HIV-positive but missing information on ART were assumed to not be on ART.

We present the distribution of births in Somkhele HDSS by maternal HIV and ART status in Figure 6.1. In the first year that HIV testing was conducted in the site, maternal HIV status was positive for approximately 4% of births, negative for 45%, and unknown for the remaining 50%. The share of HIV-positive mothers increased steadily to 28% in 2014, while those of unknown status decreased to 27%. The share of HIV-negative mothers remained relatively constant at 40-45% each year.

We investigated the relationship between maternal HIV status and U5M for children born between 2003 and 2015. The start of this period coincided with the first year of HIV testing in the HDSS, while the upper boundary limited the sample to children could be followed-up for five or more years in the latest available HDSS data. We calculated Kaplan-Meier survival curves from birth to age five for all children born in the HDSS to resident mothers. Survival curves were estimated separately for births taking place in the years [2003, 2007), [2007, 2011), and [2011, 2015); and by maternal

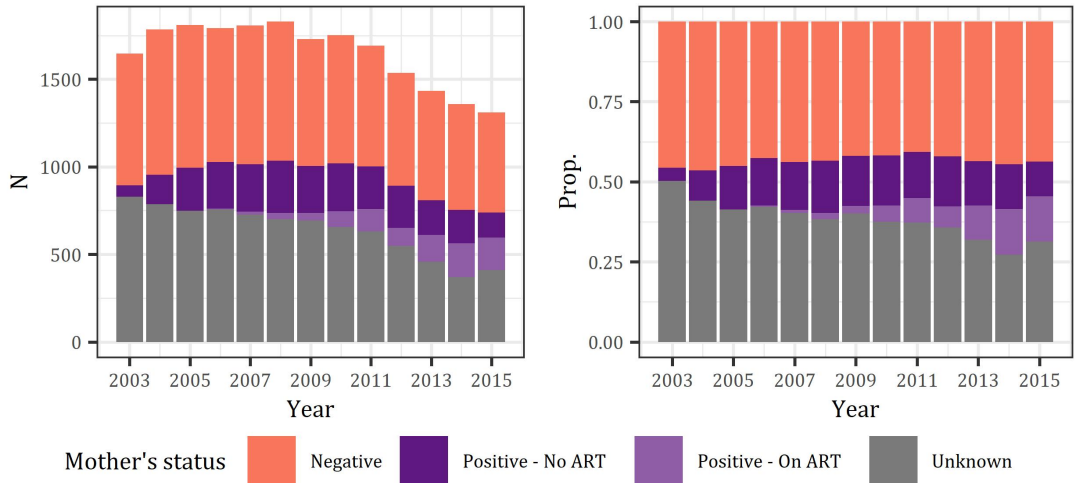


FIGURE 6.1: Annual births in Somkhele HDSS by the maternal HIV and ART status.

HIV status. Survival curves were also estimated by maternal ART status for HIV-positive mothers in the latter two cohorts. The sample of HIV-positive mothers on ART was not sufficient to produce estimates of child survival prior to this time. We fit Cox proportional hazards models regressing child survival on covariates for cohort of birth and maternal HIV/ART status. Survival time was measured in days, and observations were interval censored for temporary absences from the HDSS and right censored upon permanent out-migration or loss to follow-up.

We used individual-level records of births and deaths to estimate the crude probability of dying by age five for [2005, 2010) and [2010, 2015). The formula for the crude probability of dying, referred to as D, is shown in equation 6.1.

$$D = \frac{d}{n} \tag{6.1}$$

where d is the number of deaths and n is the number of individuals at risk.

The crude probability of dying can also be written as a weighted sum of age-specific probabilities of dying, by maternal HIV status (Bergeron-Boucher et al., 2019). This is shown in equation 6.2, with the variable i denoting maternal HIV status, and c_x^i being the share of children in age segment x for the ith category.

$$D = \sum_i \sum_x d_x^i \cdot c_x^i \tag{6.2}$$

We applied Kitagawa's decomposition method to this weighted expression to investigate changes in the crude probability of dying by age five between the two periods (Kitagawa, 1955). This approach allowed us separate the change in the crude probability of dying into two components: the part attributable to changes in the maternal HIV status composition of the population (composition effect), and the part

attributable to changes in composition-specific probabilities of dying (survival effect). The decomposition formula for the change in the crude probability of dying between two time points (t_1 and t_2) is shown in equation 6.3.

$$\Delta D = \sum_i \sum_x \underbrace{\left(\frac{d_x^i(t_2) + d_x^i(t_1)}{2}\right)\left(c_x^i(t_2) - c_x^i(t_1)\right)}_{\text{Composition effect}} + \underbrace{\left(\frac{c_x^i(t_2) + c_x^i(t_1)}{2}\right)\left(d_x^i(t_2) - d_x^i(t_1)\right)}_{\text{Survival effect}} \tag{6.3}$$

Maternal HIV status was categorized as positive or negative/unknown. The composition and survival effects were assessed for standard age segments under five years (0 to 1 month, 1 to 3 months, 3 to 6 months, 6 months to 1 year, and single years up to age 5), and summed to assess the total and relative contributions of each effect to the overall difference the crude probability of dying by age five.

6.4.3 Missing mothers and bias in estimates of U5M

In the next part of the analysis, data from 12 HDSS sites in the INDEPTH Network (including Somkhele HDSS) was used to investigate bias in estimates of U5M from missing mothers. This analysis was designed to shed light on bias in U5M estimates derived from FBHs in cross-sectional surveys. As such, we calculated estimates of U5M from the HDSS data using an approach that was analogous to that of DHS.

In each site, we simulated a set of continuous surveys taking place at the midpoint of every year. We identified all women of reproductive age who were present in the site on these dates, and thus would have been able to report their birth history in a cross-sectional survey. We identified the children of such women through mother-child links, and generated estimates of U5M for retrospective periods 0-4, 5-9, and 10-14 years prior to the simulated survey date.

Children were only included in U5M calculations if there were born in the HDSS site. The purpose of this was twofold. First, it addressed the issue of left-censoring of deaths among children who in-migrate to an HDSS after birth. Secondly, when U5M is estimated directly from FBHs, exposure to risk is calculated as the time since the child's date of birth up to their fifth birthday, death, or the date of the survey (whichever comes first). Restricting the sample to children born in the site ensured that observation began at birth, and made the U5M estimates more comparable with direct estimates calculated from FBHs.

U5M was calculated in the same manner as described in the previous section, though with the additional step of converting the mortality rate to the probability

of dying ($q(5y)$). The U5M estimate for children with mothers who were current residents represented that which would have been reported in a cross-sectional survey (i.e. *reported* $q(5y)$). We also calculated the *true* $q(5y)$ using the records of all children residing in the HDSS during the periods of interest, irrespective of their mothers' presence in the site. We calculated the ratio of true to reported $q(5y)$ to investigate bias from missing mothers. We also examined the relationship between true and reported mortality for infants (i.e. those aged from birth to 1 year; $q(1y)$) and children (i.e. those aged from one to five years; $q(1y, 5y)$) and by African region of the HDSS (i.e. eastern, southern, and western).

For children whose mothers were not present at the year midpoint, we assessed the manner of the mother's exit from the HDSS site. We differentiated between mothers that had died prior to the simulated survey date, versus those that had out-migrated from the HDSS area or been lost to follow-up. We calculated $q(5y)$ estimates for children of mothers who had died and compared them with the level of mortality reported by mothers who were current residents. As a robustness check, we also calculated $q(5y)$ for children whose mothers were not present due to out-migration as opposed to death. We also assessed differences in $q(1y)$ and $q(1y, 5y)$ by mother's exit type and region.

All statistical analysis was performed using R version 3.6.1.

6.5 Results

6.5.1 Survival analysis

The probabilities of survival up to age five years for children born in the three cohorts of interest in Somkhele HDSS are shown in Figure 6.2. Survival of children born to HIV-negative mothers has remained relatively constant across the cohorts. Those born to an HIV-negative mother in the years [2003, 2007) and [2007, 2011) had a probability of surviving to age five of 0.97 (95% confidence interval [CI] $0.97 - 0.98$), which increased slightly to 0.98 (95% CI $0.98 - 0.99$) for those born in [2011, 2015). There have been notable improvements in the survival of children born to HIV-positive mothers and mothers of unknown status over time. Children born to HIV-positive mothers in [2003, 2007) had a probability of surviving to age five of 0.87 (95% CI $0.84 - 0.89$), which increased to 0.92 (95% CI $0.91 - 0.94$) and 0.96 (95% CI $0.94 - 0.97$) for those born in [2007, 2011) and [2011, 2015), respectively. Children born in the years [2003, 2007) to a mother of unknown HIV status had a probability of surviving to age five of 0.91 (95% CI $0.90 - 0.92$). Their survival became more similar to that of children with HIV-negative mothers for the cohorts born in [2007, 2011) and [2011, 2015), where the proportions surviving to age five were 0.96 (95% CI $0.95 - 0.97$) and 0.97 (95% CI $0.97 - 0.98$), respectively.

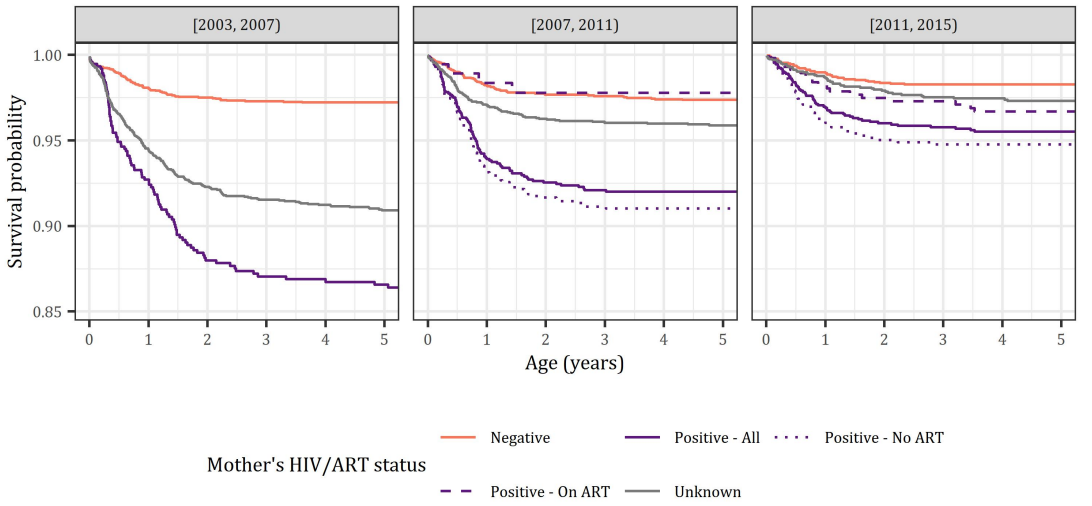


FIGURE 6.2: Kaplan-Meier survival curves for children born in three cohorts in Somkhele HDSS by the HIV/ART status of their mothers.

For children born in [2007, 2011) to HIV-positive mothers, approximately 14% (n = 184) of the mothers were reported to be on ART at the time of birth. Within this group, the probability of surviving to age five was 0.98 (95% CI 0.96 − 1.00). Survival was much lower for children of HIV-positive women who did not report ART usage at 0.91 (95% CI 0.89 − 0.93). Maternal ART usage was more widespread for the cohort born in [2011, 2015), with approximately 40% of HIV-positive mothers on ART at the time of birth. Children of mothers on ART had a probability of surviving to age five of 0.97 (95% CI 0.95 − 0.98), compared to 0.95 (95% CI 0.93 − 0.96) for children of mothers not on ART. Detailed results for the Kaplan-Meier survival curves are available in appendix Table E.1.

The results of the Cox regression analysis are shown in Table 6.1. The first model included covariates for birth cohort, maternal HIV status, and an interaction term for the two variables. Using the cohort born in [2007, 2011) as a reference and controlling for the other variables in the model, the hazard ratio for being born in [2003, 2007) was not significant (hazard ratio [HR] 1.14, 95% CI 0.85 − 1.52). On the other hand, those born in the latest period had lower mortality, with moderate significance (HR 0.65, 95% CI 0.45 − 0.94).

Children born to HIV-positive mothers had three times the hazard (HR 3.01, 95% CI 2.25 − 4.03) of U5M compared to children born to HIV-negative mothers. The interaction term showed increased mortality risk for those born to HIV-positive mothers in [2003, 2007) (HR 1.52, 95% CI 1.02 − 2.28). Taken together, the mortality hazard for children born in [2003, 2007) to HIV-positive mothers was around five times that of children born to HIV-negative mothers in [2007, 2011). The hazard ratio for children born to HIV-positive mothers in [2011, 2015) was reduced, but approximately 68%

TABLE 6.1: Mortality hazard ratios from Cox regression analysis.

		Model 1		*Model 2*	
		HR	95% CI	HR	95% CI
Cohort, ref. = [2007, 2011)	[2003, 2007)	1.14	(0.85, 1.52)		
	[2011, 2015)	0.65*	(0.45, 0.94)	0.66*	(0.46, 0.95)
Mother HIV status, ref. = Negative	Positive	3.01***	(2.25, 4.03)		
	Unknown	1.65***	(1.24, 2.19)	1.64***	(1.24, 2.18)
	Positive - No ART			3.35***	(2.49, 4.50)
	Positive - On ART			1.01	(0.41, 2.49)
Cohort : Mother HIV status	[2003, 2007) : Positive	1.52*	(1.02, 2.28)		
	[2003, 2007) : Unknown	1.81**	(1.26, 2.60)		
	[2011, 2015) : Positive	0.86	(0.53, 1.39)		
	[2011, 2015) : Unknown	0.90	(0.55, 1.49)	0.90	(0.55, 1.48)
	[2011, 2015) : Positive - No ART			0.91	(0.55, 1.52)
	[2011, 2015) : Positive - On ART			1.85	(0.64, 5.31)

Notes:
CI - confidence interval, HR - hazard ratio
*p<0.05; **p<0.01; ***p<0.001

higher than the baseline group. This indicated that having an HIV-positive mother had a more detrimental effect on survival in the earliest period than in the latest. A similar effect was seen for children born to mothers of unknown HIV status, who had an even higher hazard when born in [2003, 2007) (HR 1.81, 95% CI $1.26 - 2.60$) compared to [2011, 2015) (HR 0.90, 95% CI $0.55 - 1.49$).

In the second model, the maternal HIV status variable was re-categorized to include factor levels for whether HIV-positive mothers were on ART at the time of birth. As the HDSS only began collecting information on ART usage in 2006, the cohort born in [2003, 2007) was left out of the model. We found that children whose HIV-positive mothers were on ART did not have significantly different mortality from those with HIV-negative mothers (HR 1.01, 95% CI $0.41 - 2.49$). Alternatively, the hazard ratio for being born to an HIV-positive mother who was not on ART was significantly higher (HR 3.35, 95% CI $2.49 - 4.50$). The interaction term between birth cohort and maternal HIV status was not significant.

6.5.2 Decomposition of U5M

The crude probability of dying by age five years for [2005, 2010) was 21.8 per 1,000, compared to 8.6 per 1,000 for [2010, 2015). This resulted in a difference of -13.2 between the two periods, which we decomposed into composition and survival effects. The results of the decomposition analysis are displayed in Table 6.2.

The share of children born to HIV-positive mothers increased from [2005, 2010) to [2010, 2015). This contributed to a reduction in mortality from birth up to two years, and an increase in mortality from two to five years. This likely reflects the

TABLE 6.2: Decomposition of differences in the crude probability of dying by age five years. The contributions of the composition and survival effects to the change in the crude probability of dying are shown by detailed age segment. The absolute values of each effect were used to calculate relative contributions to the change in mortality.

Crude probability of dying by age 5				*Components (%)*		
[2005, 2010)	[2010, 2015)	Δ	Age	Composition	Rates	Total
21.8	8.6	-13.2	[0, 1m)	-0.2 (36.3)	-0.3 (63.7)	-0.5
			[1m, 3m)	-0.4 (24.4)	-1.2 (75.6)	-1.6
			[3m, 6m)	-0.7 (24.6)	-2.1 (75.4)	-2.8
			[6m, 1y)	-0.7 (21.6)	-2.7 (78.4)	-3.4
			[1y, 2y)	-0.2 (5.1)	-4.1 (94.9)	-4.3
			[2y, 3y)	0.2 (20.9)	-0.7 (79.1)	-0.5
			[3y, 4y)	0.1 (48.5)	-0.1 (51.5)	-0.0
			[4y, 5y)	0.1 (36.3)	-0.3 (63.7)	-0.1
			Total	-1.8 (13.4)	-11.4 (86.6)	-13.2

Notes:
Crude probabilities of dying displayed per 1,000

increasing usage of ARTs among HIV-positive mothers, which would yield the largest relative increases in survival at early ages. Overall, the composition effect contributed to reducing the crude probability of dying by age five by 1.8 per 1,000 between the two periods. Changes in the composition-specific probabilities of dying accounted for a reduction in the crude probability of dying by age five by 11.4 per 1,000. Reductions in the mortality of children aged three months to two years were particularly substantial.

6.5.3 Empirical investigation of bias from missing mothers

In this section, levels of true and reported infant, child, and under-5 mortality were examined for 12 HDSS sites in the INDEPTH Network. After this initial comparison, reported U5M was compared to estimates for children whose mothers were not present at the time of the simulated surveys. We distinguished between mothers who had died prior to the occurrence of the survey and those who had out-migrated from the HDSS. Details for all U5M estimates are provided in Table E.2. Tables E.3 and E.4 contain information on infant and child mortality estimates, respectively.

True and reported U5M

In Figure 6.3, the ratio of true to reported $q(5y)$ in each HDSS site is displayed by the year of the simulated survey (i.e. "simulated interview year") and the period of the estimate. As the calculations were limited to children who were born in the sites, the earliest period estimates were five years after the initiation of data collection in each HDSS.

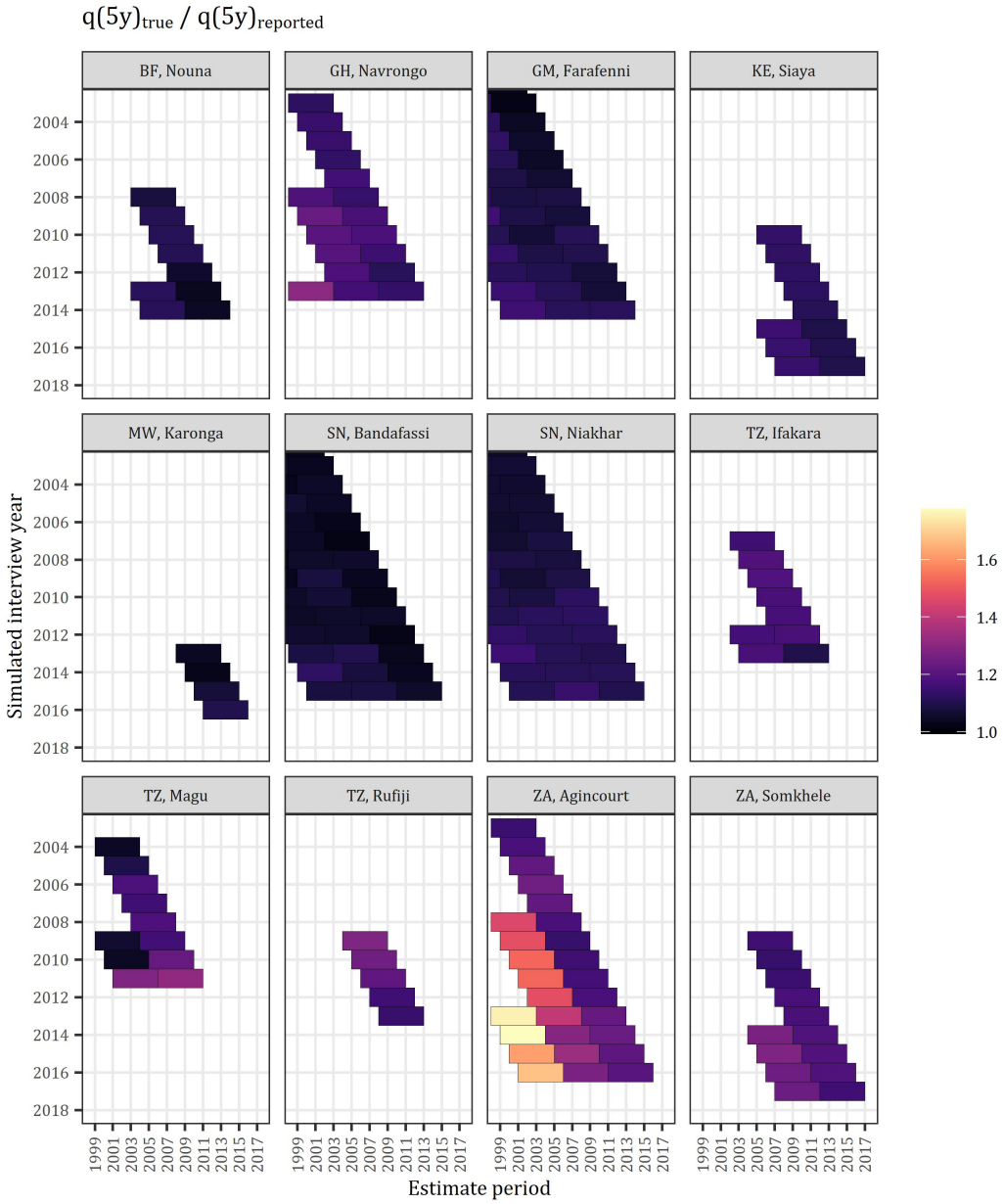
The highest ratios of true to reported $q(5y)$ were observed in estimates for 10-14 year retrospective periods in Agincourt HDSS. U5M estimates derived from the experience of women present in the site in 2013 and 2014 for periods 10-14 years prior were 76-78% lower than true $q(5y)$. The ratio of true to reported U5M declined slightly to 1.62 and 1.68 for estimates 10-14 years prior to surveys simulated in 2015 and 2016, respectively. Ratios for periods 5-9 years prior to surveys in 2008-2013 were also among the highest observed in any site, ranging between 1.41-1.53.

In this analysis, the first overlapping five-year period of U5M estimates for Agincourt and Somkhele HDSS was [2004, 2009). As estimated from women present in the site in 2009, the ratio of true to reported $q(5y)$ in Somkhele was 1.16, compared to 1.14 in Agincourt. The ratios for the same period, but as estimated from 2014, were also similar at 1.27 in Somkhele and 1.29 in Agincourt. In estimates for retrospective windows 0-4 years prior to surveys simulated between 2010 and 2017 in Somkhele, reported $q(5y)$ was downward biased by 14-19%. Corresponding estimates of bias in Agincourt ranged from 16-24%.

Compared to the South African sites, the difference between true and reported $q(5y)$ was smaller in the West African HDSS of Burkina Faso, The Gambia, and Senegal. Ratios of true to reported $q(5y)$ for all periods were less than 1.16, with 75% less than 1.11. Navrongo HDSS, located in rural northern Ghana, was notable for being the only site in West Africa with more substantial differences in true and reported $q(5y)$. True $q(5y)$ was 31% higher than the reported value for [1998, 2003), as estimated from 2013. Ratios for periods 5-9 years prior to 2008-2013 ranged from 1.16-1.23, while the retrospective 0-4 year period estimates had an average ratio of 1.14.

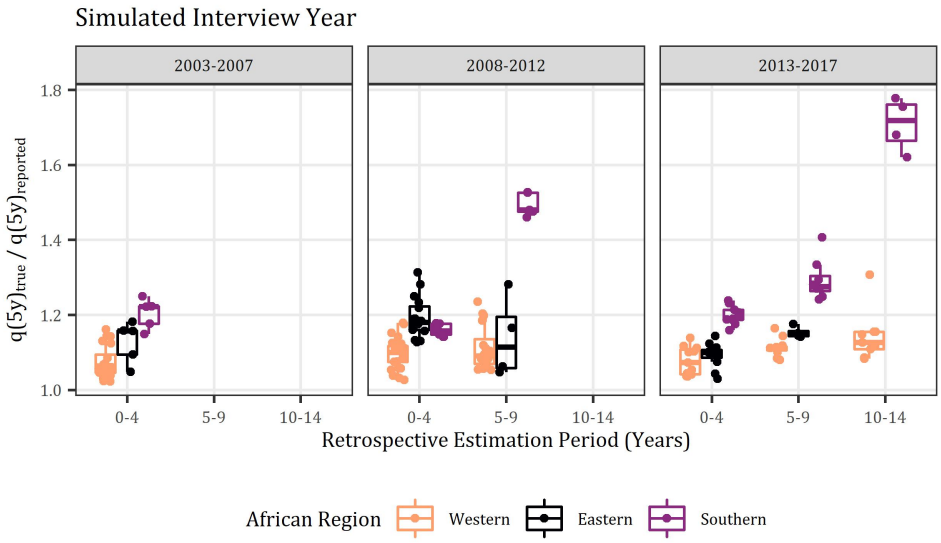
Bias for U5M estimates from sites in the eastern African countries of Kenya, Malawi, and Tanzania tended to fall between that of sites in the southern and western regions. Ratios were lowest in Karonga HDSS of northern Malawi. Reported $q(5y)$ was 4% lower than true $q(5y)$ in [2008, 2013) and increased steadily to 10% in [2011, 2016). The HDSS in Siaya, Kenya had bias of 10-13% for all 0-4 year period estimates from surveys simulated 2010-2017. The estimates for 5-9 year windows of retrospection prior to 2015-2017 had slightly higher bias at 14-15%. Magu and Rufiji HDSS in Tanzania were notable for having the highest ratios of true to reported $q(5y)$ for retrospective periods of 0-4 years in any site. Estimates of true $q(5y)$ generated from 2010-2011 were 23-31% higher than reported $q(5y)$ in Magu, while those from 2009-2011 were 22-28% higher in Rufiji. In the third Tanzanian site, Ifakara HDSS, ratios of true to reported $q(5y)$ ranged from 1.10 to 1.19 for all periods.

Figure 6.4 provides a summary of the ratios of true to reported $q(5y)$ in all HDSS by African region and years of simulated survey interviews. Bias was lowest overall in the West African sites. In the HDSS of East Africa, there was a median bias of 16% in 0-4 year period estimates from 2003-2007. This rose to 18% for interviews taking place 2008-2012, before declining to 10% for interviews in 2013-2017. The median bias



Note: Only children born in the HDSS site were included in estimation of $q(5y)$. The ratio of true to reported $q(5)$ is shown by year of the simulated survey and estimate period.

FIGURE 6.3: Ratios of true to reported $q(5y)$ from surveys simulated in HDSS.



Note: Logarithmic scale. The ratio of true to reported $q(5)$ is shown by year of the simulated survey and estimate period.

FIGURE 6.4: Ratios of true to reported $q(5y)$ by African region of HDSS.

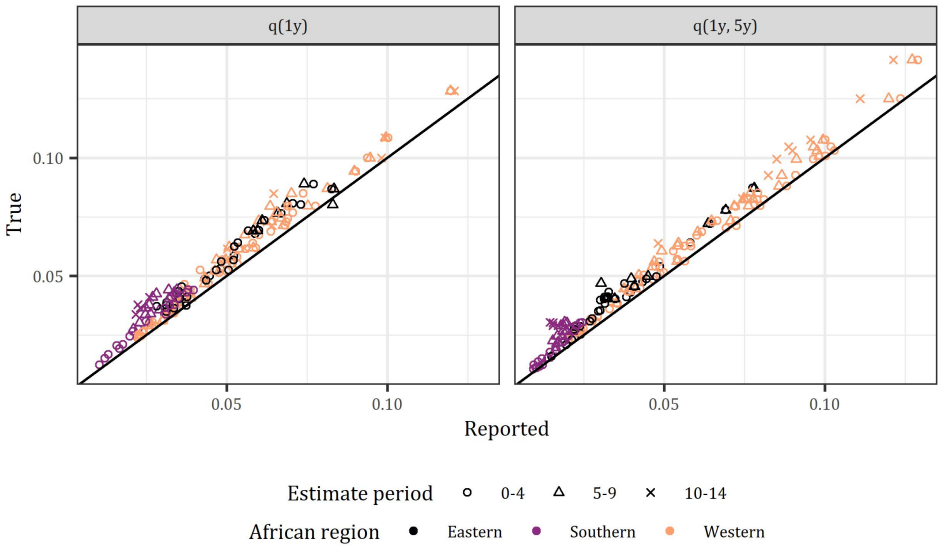
for 5-9 year period estimates was 11% and 15% for surveys simulated 2008-2012 and 2013-2017, respectively. Sites in this region had shorter time series, and there were no data points for retrospective periods 10-14 years prior to simulated surveys.

For Agincourt and Somkhele HDSS in South Africa, median bias for 0-4 year period estimates was 22% for interviews taking place 2003-2007, 16% for 2008-2012, and 19% for 2013-2017. A more substantial change was observed in the ratio of true to reported $q(5y)$ for periods 5-9 years prior to the interview. The median ratio of such estimates from interviews in 2008-2012 was 1.48, compared to 1.28 for 2013-2017. The median bias for periods 10-14 years prior to surveys simulated in Agincourt was 72%.

Figure 6.5 shows the levels of true and reported mortality in all HDSS broken down by infant and child ages. Infant mortality was lowest in the South African HDSS, followed by those in East and West Africa. Levels of $q(1y, 5y)$ in the West African sites were much higher than those found in the other two regions. The absolute difference between true and reported $q(1y, 5y)$ in the western region also increased slightly with the overall level of mortality. The South African sites had relatively larger differences between true and reported $q(1y)$ and $q(1y, 5y)$ at lower levels of mortality.

U5M for children of mothers who died

In Figure 6.6, reported $q(5y)$ was compared that of children whose mothers had died prior to the date of the simulated survey. It is important to note that the timing of the mother's death was only taken into account insofar as it occurred after the child's birth and before the date of the simulated survey. Thus, for children born in the 5-9 or 10-14 years prior to a survey, it was possible that the mother's death took place after



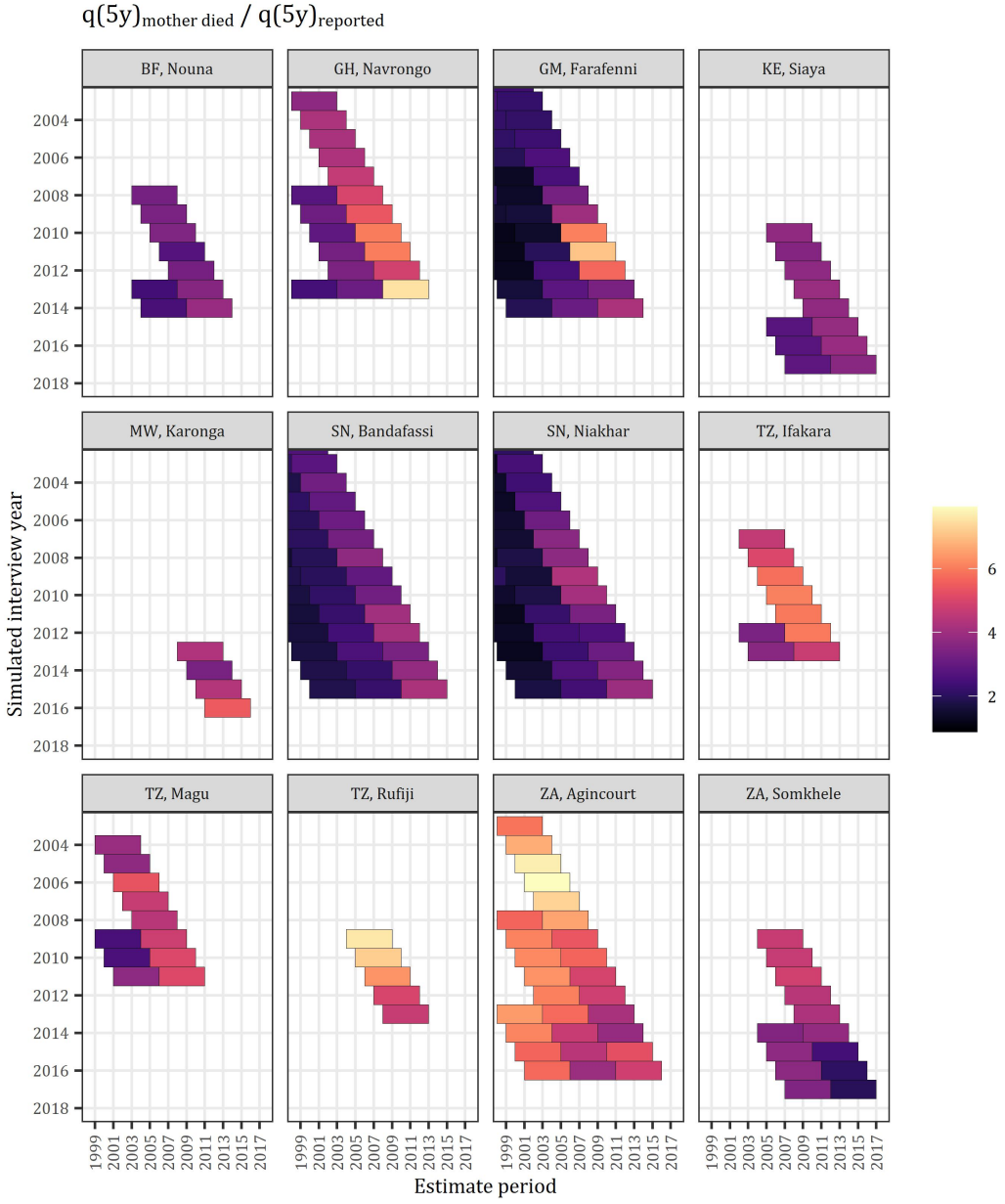
Note: Plots include estimates for periods 0-4, 5-9, and 10-14 years prior to surveys simulated in HDSS between 2003 and 2017.

FIGURE 6.5: Comparison of true and reported infant and child mortality from simulated surveys by African region of HDSS

the child's fifth birthday. In general, such deaths would not be expected to have as strong of an impact on child survival as ones taking place when the child was very young. Indeed, this appears to have been the case in Farafenni, Bandafassi, and Niakhar HDSS. For 10-14 year period estimates, $q(5y)$ for children whose mothers did not survive to the time of the survey was the most similar to reported $q(5y)$, with ratios ranging from 1.28-1.92. On the other hand, ratios for 10-14 year period estimates were substantially higher in Agincourt HDSS, ranging between 5.64-6.46.

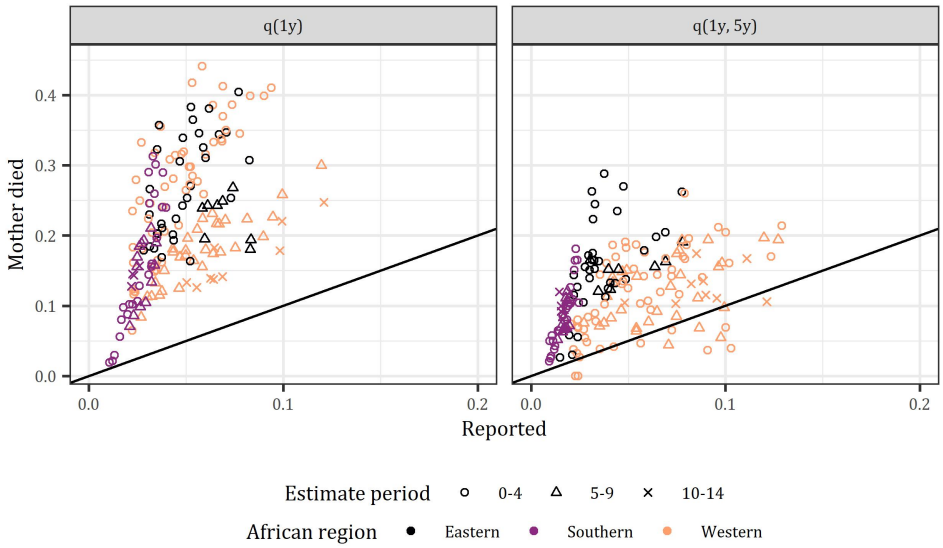
Agincourt and Somkhele HDSS were the only sites where the mortality disadvantage of children with non-surviving mothers was sometimes more pronounced in the more distant retrospective estimation periods for the same interview year. In Agincourt, U5M for children whose mothers had died averaged six times higher than reported U5M for periods 5-9 years prior to 2008-2013, compared to 5.26 times higher in 0-4 year period estimates. In Somkhele, the average ratio for surveys taking place in 2014-2017 was 3.61 for the 5-9 year windows of retrospection, compared to 2.58 for 0-4 years preceding the survey. In general, the survival disadvantage of children with deceased mothers declined since the earliest simulated surveys in both sites.

After Agincourt, the highest ratios of $q(5y)$ for children with deceased mothers compared to those with mothers present in the site were found in 0-4 period estimates from Navrongo, Farafenni, and Rufiji HDSS. These included ratios of 7.52 in Navrongo in [2008, 2013), 7.07 in Farafenni in [2006, 2011), and 7.60 in Rufiji in [2004, 2009). Ratios in Navrongo HDSS were substantially higher than other sites in West Africa. In the 0-4 years prior to simulated surveys, children whose mothers had died in Navrongo had $q(5y)$ that was on average 5.08 times higher than reported levels, compared to



Note: Only children born in the HDSS site were included in estimation of $q(5y)$. The ratio of $q(5)$ for children whose mothers had died compared to those with mothers present in the site is shown by year of interview for the simulated survey and estimate period.

FIGURE 6.6: Ratios of $q(5y)$ for children whose mothers had died prior to the date of the simulated survey in the HDSS compared to reported $q(5y)$.



Note: Plots include estimates for periods 0-4, 5-9, and 10-14 years prior to surveys simulated in HDSS between 2003 and 2017.

FIGURE 6.7: Comparison of reported infant and child mortality from simulated surveys with that of children whose mothers had died, by African region of HDSS.

3.49 for sites in Burkina Faso, The Gambia, and Senegal. Ratios were also relatively high in 0-4 year period estimates for Ifakara HDSS, averaging 5.03. They were slightly lower in Karonga and Magu, which had respective averages of 4.37 and 4.17. In Siaya, $q(5y)$ for children whose mothers had died averaged 3.76 times that of children with surviving mothers in 0-4 year period estimates, and 2.82 in 5-9 year period estimates.

In Figure 6.7, estimates of $q(1y)$ and $q(1y, 5y)$ for children whose mothers had died were plotted against reported $q(x)$. Across all regions, levels of $q(1y)$ among children with deceased mothers were consistently higher than those whose mothers were still present in the site at the time of the interview. For sites in eastern and western Africa, this survival disadvantage was almost always larger in estimates 0-4 years prior to the simulated survey compared to those 5-9 and 10-14 years prior. Regional differences in mortality by mother survival were more evident between ages one and five years. In HDSS located in southern and eastern Africa, $q(1y, 5y)$ among children with deceased mothers was always higher than reported $q(1y, 5y)$. In West Africa, $q(1y, 5y)_{mother\ died}$ tended to be higher than $q(1y, 5y)_{reported}$, though the difference between estimates was smaller. Reported $q(1y, 5y)$ was higher than that of children whose mothers had died in 13% of all period estimates from West African sites.

Robustness check: U5M for children of mothers who out-migrated

Reported $q(5y)$ was also compared to that of children whose mothers had out-migrated prior to the date of the simulated survey. The survival disadvantage from having a mother who was not present due to out-migration was much smaller than having a

mother who was not present due to death. The highest ratio of $q(5y)$ for children whose mothers had out-migrated compared to those present in the site was 2.11, observed in [2009, 2014) in Agincourt HDSS. Ratios of less than one were observed in certain periods in Farafenni, Bandafassi, and Niakhar.

Ratios of $q(5y)_{mother\ out-migrated}$ to $q(5y)_{reported}$ were highest in 0-4 year windows of retrospection. Again, this was likely due to the mother's out-migration occurring while the child was under five years of age, as opposed to potentially occurring at a later age, while still prior to the simulated survey date. Agincourt HDSS was the only site that had higher ratios for more distant retrospective periods compared to periods closer to the date of the simulated survey. For example, the average ratio for periods 10-14 years prior to surveys taking place 2013-2016 was 1.78 compared to 1.44 for 5-9 year periods. In Navrongo, children whose mothers had out-migrated averaged 1.61 times the mortality of those with surviving mothers in 0-4 year period estimates. This was higher than other sites in West Africa, which had an average ratio of 1.30. More detail on these results is provided in the appendix in Tables E.2-E.4 and Figure E.1.

6.6 Discussion

The share of children born to HIV-positive mothers in Somkhele HDSS steadily increased from 2003 to 2014. After 2014, it remained relatively constant at 25-30%. A larger share of HIV-positive mothers would be expected to have negative consequences on child health, given potential vertical transmission from mother to child and increased risk from living in an HIV-affected household (Slogrove et al., 2020). However, U5M declined substantially during the same time frame. This fits with the broader nationwide pattern of dramatic reductions in U5M from the early 2000s to the late 2010s (Johnson et al., 2020).

In Somkhele, the overall decline in U5M during this period appears largely due to reductions in the mortality of children born to HIV-positive women on ART. Information on ART uptake among HIV-positive children was not available in the site, though national estimates suggest that coverage is lower than that of adults. Approximately half of children diagnosed with HIV were estimated to be on ART in 2018 (Johnson et al., 2020). The improvement in survival is thus more likely attributable to the prevention of vertical transmission. Scale-up of PMTCT in South Africa started slow with approximately 3% coverage of pregnant women from 2002-2004, before increasing to approximately 30% in 2005 (Chigwedere et al., 2008), 60% in 2010, and close to 90% in 2018 (Johnson et al., 2020).

Yet, despite important reductions in the excess mortality of children born to HIV-positive mothers in South Africa from 2003-2017, bias in survey estimates of U5M was substantial. The median bias in $q(5y)$ estimates for periods 0-4 years prior to simulated surveys in Agincourt and Somkhele HDSS was 22% in 2003-2007, 16% for

2008-2012, and 19% for 2013-2016. Such values were all higher than the 11-15% used in the UN IGME bias adjustment procedure for surveys conducted in the neighbouring countries of Lesotho, Nambia, and Zimbabwe from 2004-2007 (Walker, Hill, and Zhao, 2012). Our estimates of bias for more distant retrospective periods were also substantially larger, at 48% compared to 27% for 5-9 year windows of retrospection, and 72% compared to 19% for periods 10-14 years prior (Walker, Hill, and Zhao, 2012).

For the 2003 DHS in Kenya, the UN IGME approach estimated downward bias of 7% for the 0-4 year period estimate of U5M (Walker, Hill, and Zhao, 2012). In this analysis, the first simulated survey in an East African HDSS was in 2004 in Magu, Tanzania, where downward bias for the 0-4 period estimate was 5%. It rose to 9% for the 0-4 years prior to 2005, held steady between 16-18% for surveys 2006-2009, and was 31% in the survey simulated in 2011. This increase in bias may have been influenced by the site's relatively flat incidence curve and modest declines in the mortality of HIV-positive adults between 2005 and 2010 (Marston et al., 2012; Reniers et al., 2014; Risher et al., 2021).

After 2011, bias declined in estimates for retrospective windows of 0-4 years for surveys simulated in other East African HDSS. Simulated surveys in sites in Kenya and Tanzania had median bias of 12% for 0-4 year period estimates. Bias was lowest in surveys simulated for 2013-2016 in Karonga HDSS, located in northern Malawi. HIV incidence has been declining in this site since 2000 (Risher et al., 2021), and dramatic reductions in adult mortality have been observed since the rollout of ARTs in 2005 (Floyd et al., 2010; Reniers et al., 2014). Overall, our results indicated an upward trajectory of bias from missing mothers in East Africa for surveys simulated in the mid-2000s to around 2010, afterwhich bias began to decline.

Bias from missing mothers was smallest in the West African sites, in keeping with the region's much lower burden of HIV (Streatfield et al., 2014b). However, Navrongo HDSS stood out for relatively large differences between true and reported $q(5y)$. This may indicate that there was a larger proportion of missing mothers, relatively higher risk for children whose mothers were missing, or some combination of these factors. With respect to the number of missing mothers, adults aged 15 or over in Navrongo have been found to have higher rates of mortality from non-communicable diseases compared to those residing in other West African sites such as Farafenni, Niakhar, and Nouna (Streatfield et al., 2014a). Rates of out-migration among young adults have also been noted to be especially high in Navrongo HDSS (Binka et al., 1999; Oduro et al., 2012). However, a cross-site comparison of adult mortality and migration was beyond the scope of this analysis. Further investigation is needed to clarify if the absence of mothers in Navrongo HDSS, whether due to death or out-migration, has a more negative impact on child survival than in other epidemiologically similar sites in West Africa.

The death of a mother has been shown to be an important risk factor for U5M,

independent of the mother's HIV status (Hammer et al., 2006; Nakiyingi et al., 2003; Slogrove et al., 2020; Zaba et al., 2005). A mother's death can affect children directly through the loss of their primary care giver or forced weaning, as well as indirectly through the redirection of household resources to the care of the mother or a reduction in household income due to the loss of the mother's wages (Bocquier et al., 2021; Hunter, Twine, and Johnson, 2011; Kadobera et al., 2012; Nakiyingi et al., 2003; Sartorius et al., 2011). In all sites, the ratio of U5M for children of deceased mothers compared to those with mothers present in the site was highest for 0-4 year period estimates. Ratios were generally lower for 5-9 and 10-14 year periods prior to simulated surveys, as the sample of children with deceased mothers included a combination of those whose mothers who died while they were under age five, as well as those whose mothers died after their fifth birthday; the latter scenario having less severe consequences for child health. In Agincourt HDSS, increased risk from a mother dying was highest in retrospective period estimates of 0-4 years between 1999 and 2007. However, ratios of $q(5y)_{mother\ died}$ to $q(5y)_{reported}$ remained elevated for 5-9 and 10-14 year period estimates. This was also observed in Somkhele HDSS, though to a lesser extent.

We hypothesize that the elevated ratios of $q(5y)_{mother\ died}$ to $q(5y)_{reported}$ across retrospective estimation periods in the South African HDSS primarily reflects three factors. The first can be identified as a cohort effect among children born prior to the widespread availability of PMTCT. Within this cohort, children of mothers who died of HIV-related causes would be much more likely to have been infected through vertical transmission, and subsequently have elevated U5M risk regardless of the relative timing of the mother's death. The second factor relates to the well-established fertility reducing effects of HIV (Glynn et al., 2000; Chen and Walker, 2010; Marston, Zaba, and Eaton, 2017; Zaba and Gregson, 1998). The fertility of HIV-infected women declines with infection duration (Ross et al., 2004; Marston et al., 2017), although the relationship may be attenuated by ART usage (Marston, Zaba, and Eaton, 2017). Nevertheless, HIV-positive women are less likely to give birth in the advanced stages of illness and relatively few mothers die from HIV-related causes in the year following birth (Ng'weshemi et al., 2003). In the case that the mother died more than five years after the birth, the increased mortality risk borne by her children would be reflected in the more distant retrospective period estimates. Lastly, HIV-related mortality among adults is often preceded by a long period of illness that may compromise the ability of the mother to attend to the health of her children (Houle et al., 2015; Todd et al., 2007). Thus, even in the case that vertical transmission did not take place, children of HIV-positive mothers may have been subject to increased risk of death from many years prior to the mother's death.

It is also important to note regional differences in the age pattern of U5M for children whose mothers had died. In western Africa, the survival disadvantage of children with deceased mothers was more evident in the first year of life compared to the following four. This was not the case in the eastern and southern regions, where children with deceased mothers had higher mortality across both age groups. The death of a mother from HIV-related causes has been previously found to be an important risk factor for mortality between ages one to three years in Agincourt HDSS (Sartorius et al., 2011). This primarily reflects the increased risk of childhood mortality for those vertically infected with HDSS. Indeed, in the decomposition of the crude probability of dying by age five in Somkhele HDSS, reductions in the mortality of children aged one to two years were the most substantial following improved coverage of ART and PMTCT programs.

This analysis was subject to several important limitations. First, the UN IGME bias correction for FBHs is only oriented towards adjusting for missing mothers who have died from HIV-related causes (Walker, Hill, and Zhao, 2012). In this analysis, the difference in the samples of children included in the true and reported $q(5y)$ calculations was not entirely attributable to HIV-related mortality on the part of the mother. There were also mothers who were not present due to out-migration, as well as those who had died from other causes. However, we do not think that this substantially compromised our assessment of HIV-related bias. Though we found that children whose mothers had out-migrated also had elevated risk of U5M compared to those with present mothers, the effect was negligible when compared to the death of a mother. HIV/AIDS is also the leading cause of death among adults in high HIV settings in southern and eastern Africa (Adjuik et al., 2006; Herbst et al., 2009; Jahn et al., 2008). At a minimum, the U5M risk associated with having a mother who out-migrated or died of a non-HIV-related cause was less than that of having an HIV-positive mother. As such, including the children of all mothers, regardless of cause of death or migration status was not likely to substantially increase the true $q(5y)$, or the estimated bias in reported $q(5y)$.

It should be emphasized that this analysis was not designed to estimate the effect of a mother's death or out-migration on child survival. We were primarily interested in the impact of missing mothers on retrospective birth history estimates of U5M. We thus did not control for many factors that would mediate the impact of a mother's absence on child survival, nor did we take account of the timing and duration of the absence relative to the child's exposure to the risk of dying. It is also important to note that in investigating the ratios of $q(5y)_{mother\ out-migrated}$ to $q(5y)_{reported}$, we did not account for flows of in-migration. Reported $q(5y)$ was calculated from the experience of all children of current HDSS residents, including those whose mothers who had recently in-migrated. There is abundant evidence for an association between U5M and mother's migration status (Brockerhoff, 1990; Bocquier, Madise, and Zulu, 2011;

Issaka, Agho, and Renzaho, 2016). To elucidate the effect of a mother's out-migration on child survival, it might be more useful to use the children of long-term residents as a comparison group. Additionally, we could not control for the fact that some of the mothers who out-migrated from the HDSS sites might have died. The estimates of $q(5y)_{mother\ out-migrated}$ thus likely reflected a combination of the risk to child survival imparted by a mother's out-migration as well as death.

Finally, while HDSS data can be leveraged to provide useful insight into bias from missing mothers in U5M estimates from FBHs, our findings are not directly applicable to estimates generated from surveys such as DHS or MICS. There are key differences in the coverage, data collection, and data quality of HDSS and surveys that warrant the cautious interpretation of results. First, populations residing in HDSS are generally much smaller and more homogenous than the sample which would be included in a large cross-sectional survey. The sites included in this analysis were predominantly rural, and some were deliberately established in highly HIV-affected areas. Whether inferences from HDSS data can be extrapolated to the wider region is a recurring question (Sankoh and Byass, 2012). With respect to data collection, HDSS typically only collect information on births occurring within the site, while FBHs gather information on all previous births. Mothers are tasked with reporting on their own children in FBHs. In contrast, HDSS often collect such information from a household representative. Routine data collection rounds are conducted at least once per year in HDSS, whereas information collected in FBHs is more vulnerable to recall bias due to the longer retrospective reporting period. Taken together, these factors may lead to systematic differences in estimates from HDSS and FBHs. With respect to data quality issues, of particular concern is the well-documented under-reporting of early mortality in HDSS (Billewicz and McGregor, 1981; Cantrelle, 1969; Eilerts et al., 2021a; Garenne, 1981; Pison and Langaney, 1985; Verhulst et al., 2021; Waiswa et al., 2019). If early mortality was underestimated in HDSS, the simulation analysis may underestimate the full extent of downward bias in FBH estimates of U5M from missing mothers.

6.7 Conclusion

Our research indicates that HIV-related bias in U5M estimates from missing mothers has declined in the past decade. In the period under study, there were dramatic improvements in child survival in the Somkhele HDSS. These appear to have been primarily driven by reductions in the mortality of children born to HIV-positive mothers. However, important bias persists in U5M estimates derived from the experience of surviving mothers in high HIV settings in eastern and southern Africa. We find that bias adjustment remains necessary, though the dynamics of ART and PMTCT uptake should be taken into account in order to avoid overcorrection.

Chapter 7

Discussion

7.1 Overview

This thesis has investigated bias and uncertainty in the level and age pattern of under-5 mortality (U5M) in sub-Saharan Africa using a variety of different data sources and inquiry methods. Of primary interest was the potential to improve understanding via data from Health and Demographic Surveillance Systems (HDSS). In this final chapter, I consolidate and reflect on the main findings. I synthesize the key results from each chapter in relation to the overall aim of this dissertation and discuss broader implications. This is followed by a discussion of strengths and weaknesses of this work, recommended avenues of future research, and concluding remarks.

7.2 Synthesis of findings

The overarching aim of this thesis was realised through four specific objectives which are listed below and provide the format for this section.

1. To compare estimates of neonatal, postneonatal, infant, and child mortality from HDSS with household surveys in sub-Saharan Africa and the historic record, and evaluate whether disparities between data sources are correlated with data quality or other attributes of the data source or population.

2. To assess whether pregnancy registration improves measurement of early mortality in HDSS.

3. To utilize linked antenatal clinic data to investigate HDSS pregnancy reporting completeness and bias in the measurement of adverse pregnancy outcomes and early mortality.

4. To leverage HDSS data to shed light on bias in birth history estimates of U5M arising from missing mothers in high-HIV settings.

7.2.1 Objective 1: To compare estimates of neonatal, postneonatal, infant, and child mortality from HDSS with household surveys in sub-Saharan Africa and the historic record, and evaluate whether disparities between data sources are correlated with data quality or other attributes of the data source or population.

The shortage of high-quality population health data has led to much uncertainty surrounding the age pattern of U5M in sub-Saharan Africa. This lack of clarity is compounded by evidence for a unique age pattern of U5M in certain populations—with the word, "unique", referring to a pattern that deviates from the historic record of high-quality data (Billewicz and McGregor, 1981; Blacker, Hill, and Timaeus, 1985; Cantrelle, 1969; Garenne, 1981; Guillot et al., 2012; Hill and Amouzou, 2006; Jasseh, 2003; McGregor, Billewicz, and Thomson, 1961; Pison and Langaney, 1985; Verhulst et al., 2021). This would undermine the practice of using conventional mortality models to examine the quality or consistency of estimates for African populations or derive estimates for ages with missing or unreliable data.

The extent to which African deviation from established age patterns is attributable to epidemiological factors or data quality issues is a subject of debate (Verhulst et al., 2021). Additionally, few studies have investigated the age pattern of U5M and issues of data quality side by side. In research focused on age patterns, issues of data quality are typically brought up to warrant the cautious interpretation of findings. On the other hand, detailed analyses of errors in African data sources do not always explore implications for the age pattern. In Chapter 3, these issues were investigated simultaneously to improve understanding of bias in U5M estimates from demographic surveillance and surveys in sub-Saharan Africa, as well as how such estimates differ systematically from one another and historical age patterns.

An atypical age pattern of infant and child mortality was observed across the contemporary data sources for sub-Saharan Africa included in the analysis, which consisted of HDSS, Demographic Health Surveys (DHS), and Multiple Indicator Cluster Surveys (MICS). These sources differ substantially in terms of their objectives, populations under study, organizational structure, resources, and perhaps mostly importantly, data collection methodologies. While the U5M estimates from HDSS were calculated from data collected longitudinally in recurring interview rounds, the survey estimates relied on information reported retrospectively in full birth history (FBH) questionnaires. As such, HDSS data collection may be less affected by recall errors. On the other hand, women report on their own reproductive experience in FBHs, while the frequent usage of proxy respondents is a systemic weakness of HDSS data. Given the many important differences between data sources, shared deviation from established age patterns of U5M lent strength to there being an epidemiological explanation, rather than one solely attributable to data quality issues.

This logic remained important when interpreting the results related to neonatal

and postneonatal mortality. While the HDSS yielded high estimates of postneonatal relative to neonatal mortality, those of DHS and MICS generally fell within the range of the historic record. There was considerable heterogeneity in HDSS estimates of neonatal mortality, with values often being unrealistically low. Lower estimates of neonatal mortality in HDSS were also associated with digit preference in the reported day of death. These findings signalled data quality problems. It also made it difficult to draw conclusions regarding the accuracy of early mortality estimates in FBHs, and more generally, the age pattern of mortality under-1 in sub-Saharan Africa.

It is worth noting that some of the extremely high ratios of child to infant mortality observed in HDSS in Chapter 3 were likely affected by under-registration of early deaths. Downward bias in HDSS measurement of neonatal mortality would manifest in infant mortality estimates, and contribute to higher than expected ratios of child to infant mortality. However, this did not invalidate the findings regarding relatively high child mortality in HDSS. HDSS estimates of child mortality were consistently higher than those of FBHs. Thus, even if atypically high ratios of child to infant mortality in HDSS were influenced by downward bias in neonatal mortality, they were not entirely explained by it. Indeed, higher estimates of child mortality in HDSS likely indicated downward bias on the part of FBHs.

The median HDSS estimate of child mortality was 6-7% higher than survey estimates for the same district or province. Child mortality estimates in DHS may be distorted by the tendency of interviews to shift births or deaths further back in time in order to avoid asking more detailed questions about child health (Pullum, 2006; Pullum and Becker, 2014; Pullum and Staveteig, 2017). If such transfers are more likely for individuals who have died, this practice can produce serious downward bias in mortality estimates for the most recent retrospective period prior to a survey (Pullum and Becker, 2014). Comparatively, the design of HDSS is well-suited to tracking child mortality. If a child survives long enough to be enumerated in the HDSS, they will be added to the household roster and followed up at each interview round. This in principle ensures data quality and accurate measurement of late child mortality. While issues with the omission of early deaths from FBHs have been well-documented (Pullum, 2006; Pullum and Becker, 2014), and served as the impetus for a variety of updates to the DHS questionnaire (Akuze et al., 2021), less attention has been called to data quality issues in FBH estimates of child mortality. This is perhaps due to such issues arising primarily from age or date reporting errors which can be indicative of larger data quality issues and thus difficult to diagnose (Pullum and Staveteig, 2017).

Some differences between HDSS and cross-sectional survey estimates of U5M may be attributable to the location of HDSS in especially high mortality populations. However, this critique is more applicable to national as opposed to sub-national level comparisons. Additionally, many sites included in the analysis of Chapter 3 were strategically established in areas deemed broadly representative of the surroundings (Ye et

al., 2012). It has also been suggested that HDSS populations may benefit from being a testing ground for public health interventions (Clark, 2021). Overall, the International Network for the Demographic Evaluation of Populations and their Health (INDEPTH Network) HDSS in sub-Saharan Africa have been found to provide relatively good coverage for a wide range of socio-economic conditions (Jia, Sankoh, and Tatem, 2015), ecological zones (Tatem, Snow, and Hay, 2006), and drivers of U5M. Though the ability to extrapolate findings from HDSS to broader areas remains unclear, this is perhaps not the most important question. Aggregate cross-site comparisons with survey data of the sort performed in Chapter 3 can further overall understanding of U5M, as well as how survey and HDSS data may be used in complement (Byass et al., 2007; Bocquier, Sankoh, and Byass, 2017).

More light was shed on potential downward bias in U5M estimates from FBHs in Chapter 6. The extreme heterogeneity and unrealistically low estimates of neonatal mortality in HDSS raised the question of whether such estimates could be improved by pregnancy registration, which was the remit of Objective 2.

7.2.2 Objective 2: To assess whether pregnancy registration improves measurement of early mortality in HDSS.

In population-based surveillance, pregnancy registration has been used as a strategy to improve monitoring of maternal health, adverse pregnancy outcomes, and births followed by early deaths. While pregnancy registration can be considered a supplemental activity to the tracking of key demographic events of births, deaths, and migrations in HDSS, it has become an important aspect of data collection in most sites (INDEPTH Network, 2008b). However, pregnancy surveillance methods are less standardized across HDSS sites than other core components, and it is often only a minority of births that are preceded by a pregnancy registration (Kwon et al., 2021; Waiswa et al., 2019).

In Chapter 4, I investigated whether U5M estimates for cohorts of births that were registered as pregnancies were more accurate than naïve estimates from HDSS. While there have been prior indications that following up on registered pregnancies can improve ascertainment of outcomes and the estimation of early mortality (Cantrelle, 1969; Cantrelle, 1974), the incompleteness of pregnancy registration (Garenne, 1981) has left open the possibility for confounding from shared risk factors for pregnancy registration and U5M. This justified a micro-level analysis to investigate concerns of selection bias. In this work, cohorts of births with and without a prior pregnancy registration (i.e. pregnancy and birth cohorts, henceforth) were balanced in terms of observable characteristics using inverse probability weighting. After this adjustment, the pregnancy cohort still had significantly higher neonatal mortality in two HDSS, and postneonatal mortality in the third site. These results led to the conclusion that

the higher observed mortality of the pregnancy cohorts was not likely due to selection bias, but rather improved ascertainment of early deaths.

The issue of misclassification of stillbirths and neonatal deaths is one that needs to be investigated further in HDSS. There is some possibility that the registration of pregnancies leads to proportionately more stillbirths wrongly reported as neonatal deaths, which would create upward bias in pregnancy cohort estimates of U5M. In Chapter 5, it was shown that stillbirths were much more likely to be reported in Siaya HDSS if the pregnancy was registered. However, this is not likely to entirely explain the higher U5M estimates for pregnancy cohorts. First, misclassified stillbirths would most likely be reported as very early deaths (Helleringer et al., 2020). As discussed in Chapter 4, pregnancy cohort U5M estimates were still higher after discounting all deaths from the first day of life. Secondly, to assert that the higher mortality of pregnancy cohorts is attributable to misclassified stillbirths, one has to also believe that pregnancy registration does not improve ascertainment of early deaths. This contention is unreasonable, unless one assumes that neonatal deaths are not subject to under-reporting in the HDSS, despite abundant evidence to the contrary.

While pregnancy registration was not strongly associated with observable characteristics of the household or mother in Chapter 4, this may not be the case in all HDSS sites. As demonstrated in Chapter 5, pregnancy registration was more likely to occur closer to the date of birth. Thus, the timing of the data collection round relative to gestational age was an important factor in pregnancy registration. It is therefore conceivable that mother-level characteristics would be better predictors of pregnancy registration in HDSS sites where data collection rounds were conducted more frequently, and a household interview was guaranteed to take place in the third trimester. The same is likely true for sites where information on pregnancy status was collected consistently, either by self-report or proxy.

The findings of Chapter 4 strongly indicated that pregnancy cohort estimates of U5M were less affected by the omission of early deaths than those of birth cohorts. However, it is possible that such estimates were still too low. Comparing the survival of children by pregnancy registration status does not take account of those who were never enumerated. If adverse events were more common among this group, U5M calculated for pregnancy cohorts could still be biased downwards. Previous research has shown that the motivations for pregnancy concealment include fear of gossip, superstition, the lack of spousal support, or being suspected of induced abortion if the pregnancy is lost (Haws et al., 2010; Kwesiga et al., 2021). Taken together, this literature illustrates the high degree of social risk faced by all women in such settings, and especially those who are socially vulnerable and may have more risk factors for U5M.

The question of whether pregnancy cohort estimates of U5M were subject to downward bias thus remained open. Objective 3 aimed to gain perspective on this issue

using an outside data source.

7.2.3 Objective 3: To utilize linked antenatal clinic data to investigate HDSS pregnancy reporting completeness and bias in the measurement of adverse pregnancy outcomes and early mortality.

Linkage with health facility records is of increasing importance to generating rich primary data on population health throughout sub-Saharan Africa (Marsh et al., 2020; Sankoh, 2015; Tollman et al., 2021). In the past decade, a number of HDSS have undertaken linkage with data from health facilities serving the study population (Reniers et al., 2016). Such data has proved valuable for investigating patterns of healthcare services utilisation in Tanzania (Cawley et al., 2015; Gourlay et al., 2015; Rentsch et al., 2018), HIV treatment outcomes in South Africa (Ambia et al., 2019; Hontelez et al., 2018; Kabudula et al., 2017b; Vandormael et al., 2014), and the burden of disease in Kenya (Etyang et al., 2014). While often not the primary objective of such studies, the applications of record linkage may be extended to augment and validate data collected in HDSS. In Chapter 5, linkage between antenatal care (ANC) clinics and an HDSS in Siaya, Kenya was used to characterize reporting of pregnancies and their outcomes and investigate potential bias in estimates of perinatal and neonatal mortality.

Chapter 5 built upon a preliminary analysis using linked ANC data from Basse HDSS in The Gambia (Rerimoi, 2019). In this study, Rerimoi organized follow-up data collection in Basse which revealed that instances of perinatal and neonatal mortality were more common for pregnancies that were missing outcomes in the HDSS. This work strongly indicated that naïve estimates of perinatal and neonatal mortality in the HDSS were biased downward, and demonstrated the feasibility of using linked ANC records to improve such data. However, the study was limited by the retrospective nature of record linkage, which relied entirely on identifiers that were routinely collected by the ANC clinics and resulted in relatively low linkage rates.

The data used in Chapter 5 arose from record linkage conducted in Siaya HDSS as part of an ongoing HIV incidence surveillance study among individuals visiting ANC clinics. The Point-of-contact Interactive Record Linkage (PIRL) approach was employed to link individuals to their record in the HDSS using probabilistic methods and a brief interview to resolve uncertainty in potential matches (Kabudula et al., 2017a; Rentsch et al., 2017a). Data from ANC registers was also digitized, enabling the investigation of clinical variables associated with pregnancy reporting, the timing of pregnancy reporting in HDSS relative to care seeking and gestational age, and misclassification between miscarriages and stillbirths. As part of my doctoral research, I had initially planned to organize follow-up interviews with record-linked women who were missing pregnancy outcome reports in Siaya HDSS. This component was abandoned due to disruptions caused by the COVID-19 pandemic—an issue I discuss later on in this chapter in the limitations section.

The linked ANC data from Siaya was incorporated into a careful analysis of the completeness of HDSS pregnancy reporting. This work required taking account of complicated dynamics related to individual residency and household data collection schedules in the HDSS, as well as information from ANC registers regarding the date of expected delivery. Triangulating this information allowed for a distinction to be made between missing pregnancy outcomes that were a likely source of downward bias in HDSS estimates of perinatal and neonatal mortality versus those that were of less concern. Belonging to the latter group included cases where the individual had exited the study area soon after seeking ANC. Such missing reports are to some extent inevitable in HDSS, as sites only strive to collect information on births occurring within the study area. On the other hand, the former group consisted of individuals who continued to reside in the HDSS for an extended period following their ANC usage and expected delivery date, and lacked pregnancy outcome reports despite the occurrence of a household interview during this time. A missing pregnancy outcome report of this nature indicated a potentially unreported instance of an adverse pregnancy outcome. A key benefit of record linkage with ANC is the ability to identify such cases, and arrange follow-up interviews to fill in gaps in data collection.

The analysis showed that the stillbirth rate was higher among cases with pregnancy registration in the HDSS. The proportion of pregnancies ending in unreported adverse outcomes was estimated as 6%, indicating potential for substantial downward bias in HDSS measurements. While the sample size of reported adverse outcomes was small, the results showed that half of reported miscarriages were likely misclassified stillbirths.

Misclassification of stillbirths and neonatal deaths is an issue of growing concern in reproductive history questionnaires (Helleringer et al., 2020; Liu et al., 2016). It was not possible to investigate this type of misclassification in this analysis, given that the ANC data did not contain information on pregnancy outcomes. However, in Siaya and elsewhere throughout sub-Saharan Africa, delivery services may be provided in the same healthcare facility compounds where ANC is accessed (Gabrysch and Campbell, 2009; Ouma et al., 2010). While the majority of births in Siaya were estimated to take place outside of health facilities in 2005 (Ouma et al., 2010), the proportion of facility births has been increasing in recent years (Doctor, Nkhana-Salimu, and Abdulsalam-Anibilowo, 2018; Kenya National Bureau of Statistics, 2015). This coincides with the aim of increasing the proportion of deliveries attended by skilled health personnel, which is codified in Sustainable Development Goal (SDG) 3.1.2 (United Nations, 2017). In the case that individuals linked while accessing ANC returned to the health facility to give birth, delivery services records could also be digitized to supplement HDSS data on pregnancy outcomes. This could be a potentially high value extension to record linkage with ANC data.

A recurring theme of this thesis has been the importance of making the most of

available data in settings lacking well-functioning civil registration and vital statistics (CRVS) systems. Bringing together such data can shed light on the strengths and weaknesses of different sources, and with careful consideration, facilitate insights that would not be possible with any single source. This chapter highlighted the benefits of record linkage between HDSS and ANC for improving data collected on pregnancy outcomes and early mortality. Under the next objective, HDSS data would be used in a unique capacity to gain insight on potential bias in U5M estimates from birth history survey data.

7.2.4 Objective 4: To leverage HDSS data to shed light on bias in birth history estimates of U5M arising from missing mothers in high-HIV settings.

Chapter 6 was justified not only by the centrality of FBHs to measuring U5M in sub-Saharan Africa, but also the severe and far-reaching impacts of the HIV epidemic which have caused gains in U5M to stagnate in recent decades (Wang et al., 2016). HDSS have played a crucial role in monitoring the progression and demographic impacts of the HIV epidemic in sub-Saharan Africa. In this chapter, such data were used to empirically investigate bias in U5M estimates derived from birth histories over the course of the HIV epidemic.

The first part of the analysis utilized data from demographic surveillance, HIV testing, and self-reporting of antiretroviral treatment (ART) usage in Somkhele HDSS. This HDSS was established in rural South Africa in 2000 for the purposes of monitoring the health impacts of the HIV epidemic in a highly affected region (Gareta et al., 2021). The increased risk of U5M associated with being born to an HIV-positive mother was found to have weakened since the early years of data collection in Somkhele. Additionally, survival differences between children born to HIV-positive and HIV-negative women were less significant if the mother was on ART around the time of delivery. These results indicated that HIV-related bias in estimates of U5M derived from birth histories has likely decreased in the era of ARTs—a proposition that was investigated empirically in the second part of the analysis.

Using data from 12 HDSS sites throughout sub-Saharan Africa, levels of true U5M were compared to that which would have been reported in a cross-sectional survey from the experience of surviving mothers. The results suggested that bias in reported U5M has generally declined since the mid-2000s. However, substantial levels of bias (>15%) persisted in the most recent five-year period estimates from surveys simulated after 2010 in sites in eastern and southern Africa. This finding was especially relevant in light of the updates that the United Nations Inter-agency Group for Child Mortality Estimation (UN IGME) has made to its bias adjustment procedure for U5M estimates from FBHs in high HIV settings (Johnson, Mizoguchi, and Pantazis, 2020).

The new correction factor was first used in the UN IGME estimates for 2020, primarily with the intention of accounting for the impact of widespread ART availability on bias reduction (UN IGME, 2020b). While the details of the adjustment procedure are still under review, preliminary descriptions indicate that it entails a substantial downward revision of the magnitude of bias in U5M estimates from surveys conducted in high HIV settings after 2000 (Johnson, Mizoguchi, and Pantazis, 2020). For example, the bias estimated for the period 5-9 years prior to the 2014 Lesotho DHS was altered from approximately 25-35% using the original adjustment procedure to less than 8% (Johnson, Mizoguchi, and Pantazis, 2020). In Chapter 6, the empirically estimated bias for corresponding estimates from the South African HDSS was found to be closer to the value of the original adjustment than the new. While the simulated bias from HDSS should not be considered directly applicable to estimates arising from national-level DHS, these results suggest that there is a risk of overstating the impact of recent gains in child survival from prevention of mother-to-child transmission (PMTCT) programs and ARTs on bias reduction in FBHs.

It is widely acknowledged that the death of a mother is an important risk factor for U5M. However, given the relative rarity of reproductive age mortality outside of high HIV settings, this is typically assumed to produce negligible bias in estimates of U5M from birth histories. The findings from this chapter suggest that such assumptions may warrant further consideration. In the West African HDSS, where reproductive age mortality from HIV-related causes was extremely low, downward bias in reported U5M for five-year periods immediately prior to simulated surveys regularly ranged from 5-10%. Additionally, children whose mothers were not present in the HDSS due to out-migration were also at higher risk of U5M, albeit to a much lesser extent than those whose mothers had died. This finding coincided with recent research which found higher mortality risk among children who were not residing with their mothers in HDSS throughout sub-Saharan Africa (Bocquier et al., 2021). Recalling the work of Chapter 3, sub-national DHS estimates of child mortality were found to be generally lower than those observed in HDSS and potentially affected by downward bias. This finding may be partly explained by the FBH reliance on mothers as respondents, and the negative consequences of a mother's absence for child survival, regardless of the cause.

Whether the higher mortality risk of children whose mothers had out-migrated would systematically bias estimates of U5M from FBHs is not clear. Women who have out-migrated would theoretically be included in the sampling frame of their new place of residence. U5M estimates calculated from birth history questionnaires are naïve to migration in the sense that there is an implicit assumption that births, deaths, and exposure to risk occurred in same place as reporting (Bocquier, Madise, and Zulu, 2011). Thus, the elevated mortality risk of children who are not coresiding with their mother

would factor into estimates for her location. This would be of concern if: (i) a substantial proportion of children were not living with their mothers, (ii) the flow of migration of mothers was primarily unidirectional, and (iii) such migrations straddled an administrative boundary that was of interest for U5M estimation. These factors may be of the most relevance for estimates of U5M by residence (i.e. urban/rural). If patterns of rapid urbanization (Collinson et al., 2014) and an urban U5M advantage hold (Issaka, Agho, and Renzaho, 2016), the increased mortality risk of children who remained behind in a rural setting while their mother left to pursue economic opportunities in an urban one would bias estimates upward in the latter area and downward in the former. However, this would not be likely to produce substantial bias if a large majority of children under five years of age live with their mothers (Bocquier, Madise, and Zulu, 2011). In any case, it may be worth elucidating the dynamics of bias introduced by "migrating mothers" into birth history estimates of U5M, in addition to "missing mothers".

7.3 Reflections

In this section, I reflect on the sum of the work contained in chapters 3-6 and the strengths and limitations of this thesis as whole.

7.3.1 Strengths

A key strength of this thesis lies in the analysis of unique primary datasets from HDSS sites operated by the Africa Health Research Institute, Kenya Medical Research Institute, and Medical Research Council Unit, The Gambia at LSHTM. In addition to standard core datasets with information on key demographic events, I used data on household visitation schedules, pregnancy registrations and outcomes, individual and household attributes, HIV testing and surveillance, and linked health records. Collaborators in each HDSS also provided insight on data collection process, protocols, and data systems. The work of chapters 4-6 would not have been possible without these inputs.

Chapter 4, in particular, was a unique contribution in that it was a cross-site analysis which also incorporated information on individual- and household-level attributes. A majority of analyses including data from multiple HDSS rely only on the core data files, as these are both the most easily accessed and standardized across sites. In Chapter 4, additional epidemiologic and socio-demographic variables were harmonized across sites and included in micro-level analyses of pregnancy reporting and child survival. Given differences in data collection and site context, this was not an entirely straightforward process. However, it was the work of this chapter to take such considerations into account in designing the analysis and interpreting the results. This

yielded useful insights into the ways in which such characteristics interact with data collection and pregnancy surveillance practices across sites.

The inclusion of a vast quantity of information from different data sources was another strength of this thesis. This was most evident in Chapter 3, which incorporated data from close to one hundred DHS and MICS surveys, a similar number of life tables representing high mortality contexts in the Human Mortality Database, and 30 HDSS sites. Insights from any single survey or surveillance study can be limited by small sample sizes, data quality issues, or lack of generalisability. These weaknesses are all mitigated in the kind of aggregate analysis performed in chapters 3 and 6, which were useful for identifying larger patterns in the age pattern of U5M and source-specific bias in estimates.

7.3.2 Limitations

The insights afforded by the analysis of linked HDSS-ANC data in Chapter 5 would have been substantially improved by follow-up data collection. As has been previously mentioned, planned fieldwork in the Siaya HDSS was a central component of this doctoral research that was cancelled with the onset of the COVID-19 pandemic. The primary aim of this study was to use information gathered in follow-up interviews on missing pregnancy outcomes to evaluate bias in the measurement of adverse pregnancy outcomes and neonatal mortality in the HDSS. As a secondary objective, follow-up interviews would have also been used to confirm the validity of linkages between the HDSS and ANC that were established through PIRL. These activities would have helped elucidate quality issues with HDSS measurement of adverse pregnancy outcomes and early mortality, as well as the feasibility of using PIRL and supplemental data collection to address them.

In lieu of this fieldwork, I conducted a largely descriptive analysis which highlighted the insights to be gained from linked ANC data and the potential for bias in HDSS estimates of stillbirths, perinatal, and neonatal mortality. However, a key limitation remains that I was only able to dimension the potential for bias in HDSS estimates rather than calculate it more precisely using information gathered in follow-up interviews. While this still contributed to furthering understanding of bias in HDSS data, it fell far short of what was initially planned.

The analysis in Chapter 5 would have been strengthened by the inclusion of linked ANC data from additional HDSS sites. Linked health facility records were available from Somkhele HDSS, however this data did not include the full slate of variables digitized from ANC registers in Siaya. The analyses performed using the Somkhele ANC data were thus subject to a much higher degree of uncertainty, and for this reason, were not included in the present version of Chapter 5. This was unfortunate for primarily two reasons. First, including data from multiple sites would have constituted a more robust analysis. Secondly, it appeared that levels of pregnancy registration

were overall much lower in Somkhele HDSS during the record linkage period than in Siaya. Thus, the potential for bias in estimates of perinatal and early mortality was likely even greater in this site, which would have more strongly demonstrated the value of augmenting HDSS with linked ANC data.

Aside from Chapter 6, this thesis did not delve deeply into substantive topics in U5M. As my doctoral research was primarily focused on uncertainty in measurement of U5M in relation to data sources and estimation methods, such analyses were deemed not essential. However, given that the level and age pattern of U5M are closely tied to the epidemiological environment and health determinants, so too are any attempts to accurately describe them. The lack of substantive investigations is a missed opportunity in this regard.

7.4 Recommendations

This section outlines recommendations for improving mortality metrics and future research.

7.4.1 Programme, policy, and practice

Phase out the use of traditional mortality models for inferring the distribution of U5M in populations in sub-Saharan Africa.

It has been consistently demonstrated that traditional model life tables, such as those produced by Coale and Demeny (1966) and the United Nations (1982), do not provide a good fit to the age pattern of U5M in many African populations. Much previous research has identified atypical age patterns of U5M in West Africa (Billewicz and McGregor, 1981; Blacker, Hill, and Timaeus, 1985; Cantrelle, 1969; Garenne, 1981; Hill and Amouzou, 2006; McGregor, Billewicz, and Thomson, 1961; Pison and Langaney, 1985). The findings in Chapter 3 can be considered part of a growing body of research which indicates that deviation is more widespread (Abdullah et al., 2007; Jasseh, 2003; Guillot et al., 2012; Verhulst et al., 2021).

There is a need to expand mortality models to reflect the diversity of age patterns observed in sub-Saharan Africa, and outside of the European experience more generally. There have been a number of recent improvements to model life tables that allow for more flexibility in estimation (Clark, 2019; Murray et al., 2003; Wilmoth et al., 2012). Further advances will require improved understanding of the distribution of risk for detailed ages under five years, beyond the standard breakdowns of neonatal, post-neonatal, and childhood age segments. A newly compiled database for U5M, bringing together high-quality data from HDSS, surveys, sample registration systems, and cohort studies in low-income and lower-middle-income countries (LLMICs), holds much promise for facilitating this work (Guillot et al., 2022; *Global Age Patterns of*

Under-5 Mortality 2022). Recent analysis of this data has identified unique age patterns of U5M that are typically masked by the use of aggregate indicators (Verhulst et al., 2021). This included patterns of higher than expected mortality for ages below four months and after six months in South Asia, upward deviations in mortality after six months throughout sub-Saharan Africa, and additional excess mortality in the first 28 days in West Africa (Verhulst et al., 2021). Research of this nature will be essential to identifying new age patterns and informing adjustments to mortality models in order to better describe them.

Utilize pregnancy cohorts to estimate U5M in HDSS.

Naïve estimates of U5M in HDSS are likely biased downwards by the omission of children whose births and deaths take place between two rounds of data collection. U5M estimates should be restricted to births that were first registered as pregnancies (i.e., pregnancy cohorts) so as to circumvent survival bias. In many HDSS, it is only a minority of births that are preceded by pregnancy registration. However, it does not appear that the higher mortality of pregnancy cohorts is attributable to confounding from U5M risk factors. The improved measurement of early mortality provided by pregnancy cohorts should outweigh concerns of selection bias, for which there is little evidence. Furthermore, the potential for selection can be reduced with the implementation of the subsequent recommendation.

Enhance and standardize pregnancy surveillance protocols in HDSS.

The findings of this thesis collectively suggest that pregnancy surveillance is essential to ensuring the quality of data collected on maternal health, pregnancy outcomes, and early child mortality in HDSS. While the INDEPTH Network requires that member sites collect data on pregnancies (Waiswa et al., 2019), the benchmarks for such data collection remain insufficiently established. Large heterogeneity persists in pregnancy surveillance protocols across HDSS sites (Kwon et al., 2021; Waiswa et al., 2019).

Pregnancies registered during one data collection round should serve as prompts to fieldworkers to follow up on their outcomes at subsequent household visits. Sites should avoid the usage of the so-called "pregnancy at birth" method, in which pregnancies are registered retroactively upon locating a newborn in the household (Waiswa et al., 2019). This practice does not confer the benefits of prospective follow-up, which should be the primary objective of pregnancy surveillance. Similarly, some data collection systems in HDSS require the registration of a pregnancy in order to proceed to recording a birth. This may result in pregnancy registrations being imputed, thus leading to the conflation of births registered as pregnancies with those that were reported retrospectively. It would be beneficial to eliminate such dependencies or add a mechanism to indicate imputation.

The usage of proxy respondents likely contributes to the under-reporting of pregnancies and their outcomes in HDSS. When possible, sites should strive to avoid collecting proxy responses for pregnancy-related information and pose questions directly to women of reproductive age. The usage of proxy respondents can be considered a systematic weakness of HDSS, though it is also worth noting that it is a necessary tool in some respects. Proxy respondents allow for information to be collected on individuals who have exited the study area. Indeed, this extends to the children of women who have died or out-migrated; a subgroup whose exclusion from maternity history questions potentially constitutes an important source of bias in U5M estimates, as explored in Chapter 6. Additionally, allowing proxy reports may provide significant cost savings for HDSS. Absent a proxy respondent, fieldworkers would be faced with making multiple return visits to the same household in order to secure interviews with all eligible subjects. Thus, reforms in this area may be constrained by the availability of site resources and must be carefully assessed. It may be worthwhile to evaluate the cost-effectiveness of using direct informants for pregnancy status information. Nevertheless, deficiencies caused by the collection of proxy reports can be mitigated by collecting detailed information on the identity of the proxy respondent. This would allow for such information to be taken into account in analyses where it is suspected of impacting results.

Treating topics surrounding pregnancy as sensitive, and implementing protocols designed to improve the reporting of such events would likely entail gains to the quality and completeness of pregnancy data (Haws et al., 2010). This includes ensuring that interviews are conducted in a private place and providing enhanced training for interviewers regarding empathy, rapport building, and how to elicit responses through follow-up questions (Haws et al., 2010; Kwesiga et al., 2021). Qualitative research in five HDSS sites in Africa and Asia indicated that respondents frequently considered questions regarding pregnancy and pregnancy loss to be intrusive and without clear relevance (Kwesiga et al., 2021). The strong sociocultural norms impacting pregnancy disclosure and psycho-social impacts of pregnancy loss are persistent barriers to collecting information on such events (Haws et al., 2010; Kiguli et al., 2015; Modiba and Nolte, 2007; Kwesiga et al., 2021). Explaining the purpose of such questions is an important duty of data collectors which would also likely improve reporting (Kwesiga et al., 2021).

Finally, pregnancy surveillance protocols should be rigorously documented and included in site profiles, or made available in the same locations as downloadable datasets. Such information is often not readily available despite its important influence on data quality and completeness. Making this information available would assist with assessing the comparability of data across sites, and serve as an important first step towards greater standardization of pregnancy surveillance protocols.

Expand electronic data collection in HDSS.

HDSS differ in whether routine data collection is conducted through paper-based or electronic methods. A study investigating pregnancy surveillance protocols in 31 INDEPTH Network HDSS in 2019 found that close to two-thirds of the sites collected data using paper-based systems (Waiswa et al., 2019). Collecting data electronically through handheld computers, or Personal Digital Assistants (PDAs), has many advantages and is recommended by the INDEPTH Network (2008a). More widespread adoption of electronic methods could benefit HDSS data in general, and pregnancy surveillance in particular.

Information collected on paper must be entered into HDSS databases at a later time. Data collection and entry are combined with the usage of handheld computers, reducing by half the amount of work and potential for the introduction of human error. Validations can also be pre-programmed into PDA applications, which provide warning messages to fill in missing values or correct inputs when values are outside the defined range. Electronic data collection has been shown to be effective at reducing the number of missing values in comparison with paper-based methods (Ley et al., 2019; McLean et al., 2017).

Providing cues for interviewers to follow up on previously registered pregnancy statuses is a key component of pregnancy surveillance in HDSS. In paper-based systems, this entails entering information collected during household interviews into the HDSS database in a timely manner, and ensuring that relevant information is prepopulated on data collection sheets for the next interview round. With the use of PDAs, fieldworkers can be prompted by the data collection software to inquire about a previously registered pregnancy at a predetermined point in the interview. Additionally, electronic data collection allows for improved monitoring of data collection progress and enables supervisors to provide quick feedback to fieldworkers (Thysen et al., 2021). This could be very useful for efficiently coordinating follow-up data collection if an outcome for a registered pregnancy was not ascertained during the first visit.

Finally, electronic data collection facilitates the collection of information related to the interview process, known as paradata. Analysis of paradata can be useful for monitoring fieldworkers, analysing corrections made to data inputs by the interviewer that shed light on problematic elements of questionnaires (Gordeev et al., 2021), and investigating interviewer effects (Harling et al., 2018). In the context of using HDSS data to measure U5M, it also has a more straightforward application that would improve the accuracy of estimates. When it comes to enumerating births in HDSS, it is important to accurately capture both the date of birth and the date of the household interview, sometimes referred to as the *observation date*. This allows for the calculation of U5M estimates using "delayed entry", in which children do not contribute exposure time to mortality rate denominators prior the date of their enumeration, accounting

for left-censoring in the data (Nareeba et al., 2021). Calculations using "delayed entry" were performed in Chapter 3, though only for a subset of the HDSS included in the analysis. Currently, information on the observation date of recorded events is not universally available in core datasets on the INDEPTH Network iShare Data Repository.[1] Including such information would allow for more accurate estimation of U5M, and also shed light on the duration between birth and enumeration. This is a problematic period from a data quality perspective, as deaths occurring prior to the child's enumeration are extremely vulnerable to omission.

Nevertheless, there are many challenges associated with conducting electronic data collection in settings with unreliable access to power and internet (Byass et al., 2008; Thysen et al., 2021). Adopting electronic data collection also requires significant training for HDSS staff and adaptations to data management systems (Thysen et al., 2021). These challenges should not be undertaken lightly, though the benefits reaped through improved data quality and more efficient allocation of time and resources would likely be substantial.

Expand record linkage between HDSS and health facilities.

Record linkage is a promising avenue for rapidly expanding the availability of detailed population health data in sub-Saharan Africa. HDSS are unique for their generation of extremely rich demographic, epidemiologic, and social dynamics data on a longitudinal basis. Linking HDSS with health records allows for the assessment of healthcare services utilisation, treatment outcomes, and the broader social determinants of health. As illustrated in Chapter 5, there is great potential for using triangulation with linked ANC records to validate and improve HDSS data collected on pregnancies and their outcomes.

The expansion of record linkage in HDSS has been proposed by the INDEPTH Network in the establishment of Comprehensive Health and Epidemiological Surveillance Systems (CHESS) (Sankoh, 2015). The CHESS initiative would aim to establish ongoing linkage between HDSS and health care facilities for the delivery of timely information on disease-specific morbidity and cause-specific mortality, and contribute to tracking and accelerating progress towards the SDGs.

Strengthen CRVS systems in LLMICs.

In this thesis, I have used a variety of data sources and inquiry methods to gain insight into bias in U5M estimates for sub-Saharan Africa. In the absence of reliable CRVS systems, data from demographic surveillance, surveys, and other alternative sources play an essential role in measuring mortality and serving as an evidence base for guiding population health policy in many LLMICs. It is important to leverage these sources

[1] https://www.indepth-ishare.org/index.php/catalog/central

to their full potential in order to fill existing data gaps, however, it must be remembered that they are not a substitute for CRVS data (AbouZahr et al., 2015b). Instead, these should be considered short- to medium-term measures, with the ultimate goal of establishing well-functioning CRVS systems (Hill et al., 2007; Ye et al., 2012).

High-quality CRVS systems generate timely data on births, deaths, and age- and cause-specific mortality (AbouZahr et al., 2015b). In LLMICs, the value of such information for resolving uncertainty regarding levels and age patterns of U5M cannot be understated. Improvements in CRVS have also been found to coincide with better health outcomes (Phillips et al., 2015). When public health programmes and policies are enacted, CRVS data are the preferred source for monitoring their effectiveness, and ensuring that targets are met (Mathers et al., 2005).

Maintaining CRVS systems requires sustained investment that is often challenging to secure in resource-constrained settings. Globally, there have been modest improvements in CRVS systems since 2000 (Mikkelsen et al., 2015). South Africa has been a notable exception, where vast improvements in birth and death registration were achieved with the establishment of mobile registration facilities in health centres and remote areas (Bah, 2009). Lessons learned in the South African context can be imparted to other countries in the region through collaborations between governments and statistical offices. The widespread usage of mobile phones also offers new ways of expanding birth and death registration to hard-to-reach remote areas (Chasukwa et al., 2022). In this era of renewed focus on CRVS, sustained prioritization from governments and global development partners will be essential to generating the data to meet health challenges of the coming decades (AbouZahr et al., 2015a; Oomman et al., 2013)

7.4.2 Future research

The findings presented in this thesis give rise to several additional questions which warrant further investigation. In this section, I propose topics for future research.

Investigate heterogeneity in HDSS estimates of neonatal mortality.

The analysis in Chapter 3 found that several HDSS reported unrealistically low levels of neonatal mortality, likely due to the under-registration of early deaths. Conducting more frequent interview rounds in HDSS is often seen as a pathway for improving the quality and completeness of data collected on births and early deaths. However, the size of the relative difference between neonatal mortality estimates for HDSS and corresponding DHS subnational regions was not strongly associated with the length of HDSS round intervals in this analysis. This suggested that there were other more important factors affecting the discrepancy. It is worth conducting an investigation

of data collection protocols which may explain variability in neonatal mortality estimates in HDSS. This would likely entail a detailed cross-site comparison of pregnancy surveillance protocols and the usage of proxy respondents.

Also in Chapter 3, there were substantial differences in the relationship between HDSS and DHS neonatal mortality estimates across African regions. Most notably, the HDSS in South Africa measured lower levels of neonatal mortality relative to DHS sub-national estimates than observed in corresponding comparisons for eastern and western Africa. This finding provides interesting context to the results from Chapter 4. Here, pregnancy cohorts in HDSS in The Gambia and Kenya had significantly higher observed neonatal mortality than the birth cohort. However, in the South African HDSS, the pregnancy cohort had significantly higher mortality in post-neonatal ages rather than the first month of life. Taken together, these unique results for South African HDSS sites raise many questions. Do data collection protocols in the South African HDSS sites account for neonatal mortality estimates that are relatively lower than those of DHS, and potentially less improved by pregnancy surveillance? Does the epidemiological environment of this region systematically bias estimates of neonatal mortality from either HDSS or DHS in a way that does not occur elsewhere? Investigating such issues would be worthwhile.

Downward bias in HDSS estimates of early mortality also has important implications for the results of Chapter 6. In the first part of this analysis, data from a South African HDSS were used to investigate the relationship between maternal HIV status and child survival. If early deaths in the HDSS were subject to under-reporting, the increased mortality risk of being born to an HIV-positive mother may have been underestimated. Similarly, in the simulations conducted in the second part of the analysis, under-ascertainment of early deaths in HDSS sites could lead to underestimation of the bias from missing mothers in birth history-derived estimates of U5M. It would be useful to conduct a thorough investigation of the quality of early mortality data in HDSS sites to assist with setting study inclusion criteria for similar analyses.

Organize follow-up interviews with record-linked women who received ANC but are lacking pregnancy outcomes in the HDSS.

Integrating records of ANC usage into HDSS data systems can allow for the active tracking of pregnancies, in which supplemental data collection is arranged for individuals who have received ANC but lack reported outcomes in the HDSS. This work was piloted by Rerimoi (2019) in Basse HDSS in The Gambia, and showed that adverse pregnancy outcomes and early deaths were more likely among unreported outcomes. These results warranted the conduct of additional research in order to determine the full extent of bias in perinatal and early mortality estimates. As has been previously stated, carrying out research of this nature in Siaya HDSS was planned as an important component of this dissertation. While it was not possible to conduct these activities

during my doctoral timeline due to disruptions caused by the COVID-19 pandemic, there is still much value to be gained from carrying out this work in Siaya and elsewhere.

Investigate effects of interviewers and proxy respondents in HDSS data collection.

Depending on the site, information regarding the identity and characteristics of HDSS fieldworkers may not be routinely collected and tracked from round to round. In sites where such information is available, it is worth investigating the effect of interviewers on data collection. Such analysis would be especially useful for sensitive events such as pregnancy and early mortality, where interviewer skills and characteristics are likely to have an even greater impact on reporting. Extending the analysis from Chapter 4, interviewer identity could be used as an instrumental variable to correct for potential selection effects from pregnancy registration on U5M estimates. The use of proxy respondents is another aspect of HDSS data collection which has important and understudied impacts on information gathered on pregnancies, births, and deaths. Further analysis in this regard is warranted, and should take account of individual characteristics of the proxy respondent, household composition, and the relationship between the proxy and the person on whose behalf they are reporting.

7.5 Conclusion

In sub-Saharan Africa, the lack of comprehensive CRVS systems complicates the monitoring of population health as well as efforts aimed at improving it. Alternative sources of information such as surveys and demographic surveillance have been used to measure maternal and child health outcomes and fill population health data gaps. There have been substantial gains in reducing the level of U5M in the past decades, and these data sources have been essential to their success. However, in many countries, the rate of reduction is currently insufficient for achieving the newborn and child health targets of the SDGs. In raw numbers, the loss remains enormous, with close to 2.7 million under-5 deaths annually in sub-Saharan Africa, constituting more than half of the global total (UN IGME, 2021). The substantial burden of stillbirths and the associated negative effects for the health and well-being of mothers and families has also been insufficiently recognized. There is an urgent need to increase both the quantity and quality of data collected in these areas. Such information provides the evidence necessary for setting public health priorities, formulating policies and evaluating their effectiveness, and ensuring accountability. In these terms, deficiencies of the existing data for sub-Saharan Africa which obscure the underlying characteristics of perinatal and U5M have become increasingly consequential.

This thesis aimed to investigate bias and uncertainty in estimates of the level and age pattern of U5M throughout sub-Saharan Africa. The analyses shed light on systematic measurement issues affecting estimates of U5M from HDSS and maternity history questionnaires collected in cross-sectional surveys. Comparisons between the two sources indicated underestimation of early mortality in the former, and childhood mortality in the latter. Other findings underscored the importance of pregnancy surveillance to improving monitoring of pregnancy outcomes and early mortality in HDSS, as well as the high potential of using record linkage with ANC clinics to enhance and validate such data. In the final investigation, data from HDSS was used to shed light on the limitations of estimating U5M from retrospective maternity history questionnaires.

In highlighting the challenges of accurately measuring U5M with current data sources, this work provides further support to calls to strengthen CRVS systems in sub-Saharan Africa. The comprehensive registration of every birth and death, and generation of timely age/cause-specific morbidity and mortality data represent crucial pathways to effectively tackling health and development challenges. At the same time, this cannot be achieved overnight. Thus, understanding the biases of data collected through surveys, surveillance, and other alternative methods is of the utmost importance to addressing urgent population health needs. It is my hope that the findings of this thesis motivate efforts to improve data collection strategies and estimation techniques so that maternal and child health can be more accurately measured. And while better data is essential for understanding the state of health, data alone does nothing to improve it. It must be made accessible to public health actors and decision-makers, and used to drive action and investment in policies and programmes in the highest burden settings. The devotion of resources to improving data is justified insofar as it leads to concrete results in service of this ultimate goal.

Bibliography

Abdool Karim, Salim S. et al. (2009). "HIV Infection and Tuberculosis in South Africa: An Urgent Need to Escalate the Public Health Response". In: *Lancet* 374.9693, pp. 921–933. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(09)60916-8.

Abdullah, Salim et al. (2007). "Patterns of Age-Specific Mortality in Children in Endemic Areas of Sub-Saharan Africa". In: *American Journal of Tropical Medicine and Hygiene* 77.6. DOI: 10.4269/ajtmh.77.6.suppl.99.

AbouZahr, Carla et al. (2015a). "Civil Registration and Vital Statistics: Progress in the Data Revolution for Counting and Accountability". In: *The Lancet* 386.10001, pp. 1373–1385. ISSN: 01406736. DOI: 10.1016/S0140-6736(15)60173-8.

AbouZahr, Carla et al. (2015b). "Towards Universal Civil Registration and Vital Statistics Systems: The Time Is Now". In: *The Lancet* 386.10001, pp. 1407–1418. ISSN: 01406736. DOI: 10.1016/S0140-6736(15)60170-2.

Adedokun, Sulaimon T. and Sanni Yaya (2020). "Correlates of Antenatal Care Utilization among Women of Reproductive Age in Sub-Saharan Africa: Evidence from Multinomial Analysis of Demographic and Health Surveys (2010-2018) from 31 Countries". In: *Archives of Public Health* 78, p. 134. ISSN: 0778-7367. DOI: 10.1186/s13690-020-00516-w.

Adetunji, Jacob (2000). "Trends in Under-5 Mortality Rates and the HIV/AIDS Epidemic". In: *Bulletin of the World Health Organization* 78.10, pp. 1200–1206. ISSN: 0042-9686. DOI: 10.1590/S0042-96862000001000006.

Adjuik, Martin et al. (2006). "Cause-Specific Mortality Rates in Sub-Saharan Africa and Bangladesh". In: *Bulletin of the World Health Organization* 84.3, pp. 181–188. ISSN: 0042-9686. DOI: 10.2471/blt.05.026492.

Ahmad, Omar B, Alan D Lopez, and Mie Inoue (2000). "The Decline in Child Mortality: A Reappraisal". In: *Bulletin of the World Health Organization* 78.10, pp. 1175–1191. ISSN: 0042-9686. DOI: 10.1590/S0042-96862000001000004.

AHRI (2021). *AHRI | About Us | Working towards the Elimination of HIV and TB Disease*. Africa Health Research Institute. URL: https://www.ahri.org/about/ (visited on 06/30/2021).

Akuze, Joseph et al. (2020). "Randomised Comparison of Two Household Survey Modules for Measuring Stillbirths and Neonatal Deaths in Five Countries: The Every Newborn-INDEPTH Study". In: *The Lancet Global Health* 8.4, e555–e566. ISSN: 2214109X. DOI: 10.1016/S2214-109X(20)30044-9.

Akuze, Joseph et al. (2021). "Four Decades of Measuring Stillbirths and Neonatal Deaths in Demographic and Health Surveys: Historical Review". In: *Population Health Metrics* 19.S1, p. 8. ISSN: 1478-7954. DOI: 10.1186/s12963-020-00225-0.

Alabi, Olatunji et al. (2014a). "Health & Demographic Surveillance System Profile: The Nahuche Health and Demographic Surveillance System, Northern Nigeria (Nahuche HDSS)". In: *International Journal of Epidemiology* 43.6, pp. 1770–1780. ISSN: 1464-3685, 0300-5771. DOI: 10.1093/ije/dyu197.

Alabi, Olatunji et al. (2014b). "Lessons Learned from Setting up the Nahuche Health and Demographic Surveillance System in the Resource-Constrained Context of Northern Nigeria". In: *Global Health Action* 7. ISSN: 1654-9716. DOI: 10.3402/gha.v7.23368.

Alexander, Monica and Leontine Alkema (2018). "Global Estimation of Neonatal Mortality Using a Bayesian Hierarchical Splines Regression Model". In: *Demographic Research* 38, pp. 335–372. ISSN: 1435-9871. DOI: 10.4054/DemRes.2018.38.15.

Alkema, Leontine and Jin Rou New (2014). "Global Estimation of Child Mortality Using a Bayesian B-spline Bias-reduction Model". In: *The Annals of Applied Statistics* 8.4, pp. 2122–2149. ISSN: 1932-6157. DOI: 10.1214/14-AOAS768.

Ambia, Julie et al. (2019). "Outcomes of Patients Lost to Followup after Antiretroviral Therapy Initiation in Rural Northeastern South Africa". In: *Tropical Medicine & International Health* 24.6, pp. 747–756. ISSN: 1360-2276, 1365-3156. DOI: 10.1111/tmi.13236.

Amek, Nyaguara et al. (2015). "Using Health and Demographic Surveillance System (HDSS) Data to Analyze Geographical Distribution of Socio-Economic Status; an Experience from KEMRI/CDC HDSS". In: *Acta Tropica* 144, pp. 24–30. ISSN: 0001706X. DOI: 10.1016/j.actatropica.2015.01.006.

Aminu, M et al. (2014). "Causes of and Factors Associated with Stillbirth in Low- and Middle-Income Countries: A Systematic Literature Review". In: *BJOG: An International Journal of Obstetrics & Gynaecology* 121.s4, pp. 141–153. ISSN: 1471-0528. DOI: 10.1111/1471-0528.12995.

Ancel, Pierre-Yves et al. (2015). "Survival and Morbidity of Preterm Children Born at 22 Through 34 Weeks Gestation in France in 2011: Results of the EPIPAGE-2 Cohort Study". In: *JAMA Pediatrics* 169.3, p. 230. ISSN: 2168-6203. DOI: 10.1001/jamapediatrics.2014.3351.

Armstrong, David (1986). "The Invention of Infant Mortality". In: *Sociology of Health & Illness* 8.3, pp. 211–232. ISSN: 1467-9566. DOI: 10.1111/j.1467-9566.1986.tb00298.x.

Artzrouni, Marc and Basia Zaba (Mar. 2003). "HIV-induced Bias in the Estimation of Child Mortality Using Birth History Reports." Technical Meeting on HIV and Child Mortality (London).

Assefa, Nega et al. (2016). "HDSS Profile: The Kersa Health and Demographic Surveillance System". In: *International Journal of Epidemiology* 45.1, pp. 94–101. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dyv284.

Bah, Sulaiman (2009). "Multiple Forces Working in Unison: The Case of Rapid Improvement of Vital Statistics in South Africa Post-1996". In: *World Health & Population* 11.4, pp. 50–59. ISSN: 19296541. DOI: 10.12927/whp.2013.21017.

Bergeron-Boucher, Marie-Pier et al. (2019). "Understanding Differences in Cancer Survival between Populations: A New Approach and Application to Breast Cancer Survival Differentials between Danish Regions". In: *International Journal of Environmental Research and Public Health* 16.17, p. 3093. ISSN: 1660-4601. DOI: 10.3390/ijerph16173093.

Billewicz, W. Z. and I. A. McGregor (1981). "The Demography of Two West African (Gambian) Villages, 195175". In: *Journal of Biosocial Science* 13.2, pp. 219–240. ISSN: 1469-7599, 0021-9320. DOI: 10.1017/S0021932000013390.

Binka, Fred N. et al. (1999). "Assessing Population Dynamics in a Rural African Society: The Navrongo Demographic Surveillance System". In: *Journal of Biosocial Science* 31.3, pp. 375–391. ISSN: 0021-9320, 1469-7599.

Blacker, John, Allan Hill, and Ian Timaeus (June 5–12, 1985). "Age Patterns of Mortality in Africa: An Examination of Recent Evidence". In: *International Union for the Scientific Study of Population*. Florence International Population Conference.

Blencowe, Hannah (2020). "Counting the Smallest: Data to Estimate Global Stillbirth, Preterm Birth and Low Birthweight Rates". PhD thesis. London School of Hygiene & Tropical Medicine.

Blencowe, Hannah et al. (2016). "National, Regional, and Worldwide Estimates of Stillbirth Rates in 2015, with Trends from 2000: A Systematic Analysis". In: *The Lancet Global Health* 4.2, e98–e108. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(15)00275-2.

Bocquier, Philippe, Nyovani Janet Madise, and Eliya Msiyaphazi Zulu (2011). "Is There an Urban Advantage in Child Survival in Sub-Saharan Africa? Evidence From 18 Countries in the 1990s". In: *Demography* 48.2, pp. 531–558. ISSN: 0070-3370, 1533-7790. DOI: 10.1007/s13524-011-0019-2.

Bocquier, Philippe, Osman Sankoh, and Peter Byass (2017). "Are Health and Demographic Surveillance System Estimates Sufficiently Generalisable?" In: *Global Health Action* 10.1, p. 1356621. ISSN: 1654-9716, 1654-9880. DOI: 10.1080/16549716.2017.1356621.

Bocquier, Philippe et al. (2021). "The Crucial Role of Mothers and Siblings in Child Survival: Evidence From 29 Health and Demographic Surveillance Systems in Sub-Saharan Africa". In: *Demography* 58.5, pp. 1687–1713. ISSN: 0070-3370, 1533-7790. DOI: 10.1215/00703370-9429499.

Boerma, J. T., A. J. Nunn, and J. A. Whitworth (1998). "Mortality Impact of the AIDS Epidemic: Evidence from Community Studies in Less Developed Countries". In: *AIDS (London, England)* 12 Suppl 1, S3–14. ISSN: 0269-9370.

Bongaarts, John and Susan Cotts Watkins (1996). "Social Interactions and Contemporary Fertility Transitions". In: *Population and Development Review* 22.4, p. 639. ISSN: 00987921. DOI: 10.2307/2137804.

Bor, Jacob et al. (2013). "Increases in Adult Life Expectancy in Rural South Africa: Valuing the Scale-Up of HIV Treatment". In: *Science* 339.6122, pp. 961–965. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1230413.

Bradley, Sarah E.K., William Winfrey, and Trevor N. Croft (2015). *Contraceptive Use and Perinatal Mortality in the DHS: An Assessment of the Quality and Consistency of Calendars and Histories*. DHS Methodological Reports 17. Rockville, Maryland, USA: ICF International.

Brass, W. and S. Macrae (1984). "Childhood Mortality Estimated from Reports on Previous Births given by Mothers at the Time of a Maternity: Preceding-Births Technique". In: *Asian and Pacific Census Forum* 11.2, pp. 5–8. ISSN: 0732-0515.

Brockerhoff, Martin (1990). "Rural-to-Urban Migration and Child Survival in Senegal". In: *Demography* 27.4, pp. 601–616. ISSN: 0070-3370, 1533-7790. DOI: 10.2307/2061573.

Buchanan, Ashley L. et al. (2014). "Worth the Weight: Using Inverse Probability Weighted Cox Models in AIDS Research". In: *AIDS Research and Human Retroviruses* 30.12, pp. 1170–1177. ISSN: 0889-2229. DOI: 10.1089/aid.2014.0037.

Burton, Rosie, Janet Giddy, and Kathryn Stinson (2015). "Prevention of Mother-to-Child Transmission in South Africa: An Ever-Changing Landscape". In: *Obstetric Medicine* 8.1, pp. 5–12. ISSN: 1753-495X, 1753-4968. DOI: 10.1177/1753495X15570994.

Byass, Peter (2009). "The Unequal World of Health Data". In: *PLOS Medicine* 6.11, e1000155. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1000155.

Byass, Peter et al. (2007). "DSS and DHS: Longitudinal and Cross-Sectional Viewpoints on Child and Adolescent Mortality in Ethiopia". In: *Population Health Metrics* 5.1, p. 12. ISSN: 1478-7954. DOI: 10.1186/1478-7954-5-12.

Byass, Peter et al. (2008). "Direct Data Capture Using Hand-Held Computers in Rural Burkina Faso: Experiences, Benefits and Lessons Learnt". In: *Tropical Medicine & International Health* 13.s1, pp. 25–30. ISSN: 1365-3156. DOI: 10.1111/j.1365-3156.2008.02084.x.

Byass, Peter et al. (2015). "A Successful Failure: Missing the MDG4 Target for Under-Five Mortality in South Africa". In: *PLOS Medicine* 12.12, e1001926. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001926.

Cantrelle, P. and H. Leridon (1971). "Breast Feeding, Mortality in Childhood and Fertility in a Rural Zone of Senegal". In: *Population Studies* 25.3, pp. 505–533. ISSN: 0032-4728. DOI: 10.1080/00324728.1971.10405821.

Cantrelle, Pierre (1969). "Etude Démographique Dans La Région Du Sine-Saloum (Sénégal): État Civil et Observation Démographique 1963-1965". In: *Travaux et Documents de l'O.R.S.T.O.M.*

— (1974). "Is There a Standard Pattern of Tropical Mortality?" In: *Population in African Development*. Ed. by Pierre Cantrelle et al. International Union for the Scientific Study of Population (IUSSP), pp. 33–42.

Cawley, Caoimhe et al. (2015). "Risk Factors for Service Use and Trends in Coverage of Different HIV Testing and Counselling Models in Northwest Tanzania between 2003 and 2010". In: *Tropical Medicine & International Health* 20.11, pp. 1473–1487. ISSN: 13602276. DOI: 10.1111/tmi.12578.

Chasukwa, Michael et al. (2022). *Collecting Mortality Data via Mobile Phone Surveys: A Non-Inferiority Randomized Trial in Malawi*. preprint. Infectious Diseases (except HIV/AIDS). DOI: 10.1101/2022.03.02.22271441.

Chen, J. et al. (1998a). "Maternal Education and Fetal and Infant Mortality in Quebec. Fetal and Infant Mortality Study Group of the Canadian Perinatal Surveillance System." In: *Health reports* 10.2, pp. 53–64. ISSN: 0840-6529.

Chen, L. M. et al. (1998b). "Underregistration of Neonatal Deaths: An Empirical Study of the Accuracy of Infantile Vital Statistics in Taiwan". In: *Journal of Epidemiology and Community Health* 52.5, pp. 289–292. ISSN: 0143-005X.

Chen, W.-J. and N. Walker (2010). "Fertility of HIV-infected Women: Insights from Demographic and Health Surveys". In: *Sexually Transmitted Infections* 86 (Suppl 2), pp. ii22–ii27. ISSN: 1368-4973. DOI: 10.1136/sti.2010.043620.

Chigwedere, Pride et al. (2008). "Estimating the Lost Benefits of Antiretroviral Drug Use in South Africa". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 49.4, pp. 410–415. ISSN: 1525-4135. DOI: 10.1097/QAI.0b013e31818a6cd5.

Clark, Sam et al. (1995). "Ten Thousand Tonga: A Longitudinal Anthropological Study from Southern Zambia, 19561991". In: *Population Studies* 49.1, pp. 91–109. ISSN: 0032-4728, 1477-4747. DOI: 10.1080/0032472031000148266.

Clark, Samuel J. (2019). "A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality". In: *Demography* 56.3, pp. 1131–1159. ISSN: 0070-3370, 1533-7790. DOI: 10.1007/s13524-019-00785-3.

— (2021). "Health and Demographic Surveillance Systems and the 2030 Agenda: Sustainable Development Goals". In: *arXiv* 1 (2103:03910).

Clark, Shelley and Dana Hamplová (2013). "Single Motherhood and Child Mortality in Sub-Saharan Africa: A Life Course Perspective". In: *Demography* 50.5, pp. 1521–1549. ISSN: 0070-3370, 1533-7790. DOI: 10.1007/s13524-013-0220-6.

Cleland, J and V Verma (1989). "The World Fertility Survey: An Appraisal of Methodology". In: *Journal of the American Statistical Association* 84.407, pp. 756–767.

Cleland, John (1996). "Demographic Data Collection in Less Developed Countries 1946-1996". In: *Population Studies* 50.3, pp. 433–450. ISSN: 0032-4728, 1477-4747. DOI: 10.1080/0032472031000149556.

Clouse, Kate et al. (2013). "Loss to follow-up before and after delivery among women testing HIV positive during pregnancy in Johannesburg, South Africa". In: *Tropical Medicine & International Health* 18.4, pp. 451–460. ISSN: 1365-3156. DOI: 10.1111/tmi.12072.

Coale, Ansley J and Paul George Demeny (1966). *Regional Model Life Tables and Stable Populations*, Princeton, N.J.: Princeton University Press. ISBN: 978-0-691-09330-7.

Cock, Kevin M De et al. (1994). "Retrospective Study of Maternal HIV-1 and HIV-2 Infections and Child Survival in Abidjan, Côte d'Ivoire". In: *BMJ* 308, pp. 441–443. DOI: 10.1136/bmj.308.6926.441.

Collinson, Mark A. et al. (2014). "Migration and the Epidemiological Transition: Insights from the Agincourt Sub-District of Northeast South Africa". In: *Global Health Action* 7.1, p. 23514. ISSN: 1654-9716, 1654-9880. DOI: 10.3402/gha.v7.23514.

Cook, L. J., L. M. Olson, and J. M. Dean (2001). "Probabilistic Record Linkage: Relationships between File Sizes, Identifiers and Match Weights". In: *Methods of Information in Medicine* 40.3, pp. 196–203. ISSN: 0026-1270.

Cunningham, Solveig A et al. (2019). "Health and Demographic Surveillance Systems Within the Child Health and Mortality Prevention Surveillance Network". In: *Clinical Infectious Diseases* 69 (Supplement 4), S274–S279. ISSN: 1058-4838, 1537-6591. DOI: 10.1093/cid/ciz609.

Curto, José C. (1994). "Sources for the Pre-1900 Population History of Sub-Saharan Africa : The Case of Angola, 1773-1845". In: *Annales de démographie historique* 1994.1, pp. 319–338. ISSN: 0066-2062. DOI: 10.3406/adh.1994.1876.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney (2006). "The Determinants of Mortality". In: *The Journal of Economic Perspectives* 20.3, pp. 97–120. DOI: 10.1257/jep.20.3.97.

Defo, Barthélémy Kuate (2014). "Demographic, Epidemiological, and Health Transitions: Are They Relevant to Population Health Patterns in Africa?" In: *Global Health Action* 7.1, p. 22443. ISSN: 1654-9716, 1654-9880. DOI: 10.3402/gha.v7.22443.

Deribew, Amare et al. (2016). "Under-Five Mortality Rate Variation between the Health and Demographic Surveillance System (HDSS) and Demographic and Health Survey (DHS) Approaches". In: *BMC Public Health* 16.1, p. 1118. ISSN: 1471-2458. DOI: 10.1186/s12889-016-3786-2.

DHS (2019). *DHS-8 Questionnaires: Revision Process and New Content*. ICF International. URL: https://dhsprogram.com/pubs/pdf/DHSM11/DHSM11.pdf (visited on 03/14/2022).

DHS (2022). *Quality Information to Plan, Monitor and Improve Population, Health, and Nutrition Programs*. The Demographic and Health Surveys Program. URL: https://dhsprogram.com/ (visited on 03/14/2022).

Doctor, Henry V., Sangwani Nkhana-Salimu, and Maryam Abdulsalam-Anibilowo (2018). "Health Facility Delivery in Sub-Saharan Africa: Successes, Challenges, and Implications for the 2030 Development Agenda". In: *BMC Public Health* 18.1, p. 765. ISSN: 1471-2458. DOI: 10.1186/s12889-018-5695-z.

Dwyer-Lindgren, Laura et al. (2013). "Error and Bias in Under-5 Mortality Estimates Derived from Birth Histories with Small Sample Sizes". In: *Population Health Metrics* 11.1, p. 13. ISSN: 1478-7954. DOI: 10.1186/1478-7954-11-13.

Eilerts, Hallie et al. (2021a). "Age Patterns of Under-5 Mortality in Sub-Saharan Africa during 19902018: A Comparison of Estimates from Demographic Surveillance with Full Birth Histories and the Historic Record". In: *Demographic Research* 44.18, pp. 415–442. ISSN: 1435-9871. DOI: 10.4054/DemRes.2021.44.18.

Eilerts, Hallie et al. (Dec. 9, 2021b). "Pregnancy Reporting and Biases in Under-5 Mortality in Three HDSS in Western, Eastern, and Southern Africa". In: *Session 136, Data Quality: From Pregnancy to Death*. International Population Conference. International Union for the Scientific Study of Population.

EN-INDEPTH (June 15–17, 2016). "Workshop Report". In: *Every Newborn Action Plan Metrics*. ENAP & INDEPTH Research Protocol Design Workshop. Kampala, Uganda: Every Newborn Action Plan partnering with Ending Preventable Maternal Mortality.

Espeut, Donna and Stan Becker (2015). "The Validity of Birth and Pregnancy Histories in Rural Bangladesh". In: *Journal of Health, Population and Nutrition* 33.1, p. 17. ISSN: 2072-1315. DOI: 10.1186/s41043-015-0027-8.

Etyang, Anthony O. et al. (2014). "Burden of Disease in Adults Admitted to Hospital in a Rural Region of Coastal Kenya: An Analysis of Data from Linked Clinical and Demographic Surveillance Systems". In: *The Lancet Global Health* 2.4, e216–e224. ISSN: 2214109X. DOI: 10.1016/S2214-109X(14)70023-3.

Farahani, Mansour et al. (2014). "Outcomes of the Botswana National HIV/AIDS Treatment Programme from 2002 to 2010: A Longitudinal Analysis". In: *The Lancet Global Health* 2.1, e44–e50. ISSN: 2214109X. DOI: 10.1016/S2214-109X(13)70149-9.

Fellegi, Ivan P. and Alan B. Sunter (1969). "A Theory for Record Linkage". In: *Journal of the American Statistical Association* 64.

Floyd, Sian et al. (2010). "Population-Level Reduction in Adult Mortality after Extension of Free Anti-Retroviral Therapy Provision into Rural Areas in Northern Malawi". In: *PLoS ONE* 5.10. Ed. by Landon Myer, e13499. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0013499.

Ford, Jane B., Christine L. Roberts, and Lee K. Taylor (2006). "Characteristics of Unmatched Maternal and Baby Records in Linked Birth Records and Hospital Discharge Data". In: *Paediatric and Perinatal Epidemiology* 20.4, pp. 329–337. ISSN: 0269-5022, 1365-3016. DOI: 10.1111/j.1365-3016.2006.00715.x.

Frøen, J Frederik et al. (2009). "Making Stillbirths Count, Making Numbers Talk - Issues in Data Collection for Stillbirths". In: *BMC Pregnancy and Childbirth* 9.1, p. 58. ISSN: 1471-2393. DOI: 10.1186/1471-2393-9-58.

Gabrysch, Sabine and Oona MR Campbell (2009). "Still Too Far to Walk: Literature Review of the Determinants of Delivery Service Use". In: *BMC Pregnancy and Childbirth* 9.1, p. 34. ISSN: 1471-2393. DOI: 10.1186/1471-2393-9-34.

Garenne, M and P Cantrelle (1997). "Three Decades of Research on Population and Health: The ORSTOM Experience in Rural Senegal, 1962-1991." In: *Prospective Community Studies in Developing Countries*. Oxford: Clarendon Press, pp. 233–52.

Garenne, Michel (1981). "The Age Pattern of Infant and Child Mortality in Ngayokheme (Rural West Africa)". In: *University of Pennsylvania Scholarly Commons* African Demography Working Paper Series, p. 42.

— (2006). "Health Transitions in Sub-Saharan Africa: Overview of Mortality Trends in Children under 5 Years Old (1950-2000)". In: *Bulletin of the World Health Organization* 84.6, pp. 470–478. ISSN: 00429686. DOI: 10.2471/BLT.05.029231.

Gareta, Dickman et al. (2021). "Cohort Profile Update: Africa Centre Demographic Information System (ACDIS) and Population-Based HIV Survey". In: *International Journal of Epidemiology* 50.1, pp. 33–34. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dyaa264.

Gaylin, Daniel S and Jennifer Kates (1997). "Refocusing the Lens: Epidemiologic Transition Theory, Mortality Differentials, and the AIDS Pandemic". In: *Social Science & Medicine* 44.5, pp. 609–621. ISSN: 02779536. DOI: 10.1016/S0277-9536(96)00212-2.

Global Age Patterns of Under-5 Mortality (2022). URL: https://web.sas.upenn.edu/global-age-patterns-under-five-mortality/ (visited on 03/03/2022).

Glynn, Judith et al. (2000). "Decreased Fertility Among HIV-1 Infected Women Attending Antrenatal Clinics in Three African Cities". In: *Journal of Acquired Immune Deficiency Syndromes*. DOI: 10.1097/00042560-200012010-00008.

Goga, Ameena E et al. (2015). "First Population-Level Effectiveness Evaluation of a National Programme to Prevent HIV Transmission from Mother to Child, South Africa". In: *Journal of Epidemiology and Community Health* 69.3, pp. 240–248. ISSN: 0143-005X, 1470-2738. DOI: 10.1136/jech-2014-204535.

Gordeev, Vladimir Sergeevich et al. (2021). "Paradata Analyses to Inform Population-Based Survey Capture of Pregnancy Outcomes: EN-INDEPTH Study". In: *Population Health Metrics* 19.1, p. 10. ISSN: 1478-7954. DOI: 10.1186/s12963-020-00241-0.

Gould, W.T.S. (1998). "African Mortality and the New 'urban Penalty'". In: *Health & Place* 4.2, pp. 171–181. ISSN: 13538292. DOI: 10.1016/S1353-8292(98)00009-4.

Gourlay, Annabelle et al. (2015). "Factors Associated with Uptake of Services to Prevent Mother-to-Child Transmission of HIV in a Community Cohort in Rural Tanzania". In: *Sexually Transmitted Infections* 91.7, pp. 520–527. ISSN: 1368-4973, 1472-3263. DOI: 10.1136/sextrans-2014-051907.

Guillot, Michel et al. (2012). "Child Mortality Estimation: A Global Overview of Infant and Child Mortality Age Patterns in Light of New Empirical Data". In: *PLoS Medicine* 9.8. Ed. by Peter Byass, e1001299. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001299.

Guillot, Michel et al. (2020). "Modeling Age Patterns of Under-5 Mortality: Results From a Log-Quadratic Model Applied to High-Quality Vital Registration Data". In: *University of Pennsylvania Scholarly Commons*. Population Center Working Papers (PSC/PARC), p. 72.

— (2022). "Modeling Age Patterns of Under-5 Mortality: Results From a Log-Quadratic Model Applied to High-Quality Vital Registration Data". In: *Demography* 59.1, pp. 321–347. ISSN: 0070-3370, 1533-7790. DOI: 10.1215/00703370-9709538.

Gwatkin, Davidson R. (1980). "Indications of Change in Developing Country Mortality Trends: The End of an Era?" In: *Population and Development Review* 6.4, p. 615. ISSN: 00987921. DOI: 10.2307/1972929.

Hallett, Timothy B et al. (2010). "Measuring and Correcting Biased Child Mortality Statistics in Countries with Generalized Epidemics of HIV Infection". In: *Bulletin of the World Health Organization* 88.10, pp. 761–768. ISSN: 0042-9686. DOI: 10.2471/BLT.09.071779.

Hamel, Mary J. et al. (2011). "A Reversal in Reductions of Child Mortality in Western Kenya, 2003-2009". In: *The American Journal of Tropical Medicine and Hygiene* 85.4, pp. 597–605. ISSN: 0002-9637, 1476-1645. DOI: 10.4269/ajtmh.2011.10-0678.

Hammer, Gaël P. et al. (2006). "Risk Factors for Childhood Mortality in Sub-Saharan Africa". In: *Acta Tropica* 98.3, pp. 212–218. ISSN: 0001706X. DOI: 10.1016/j.actatropica.2006.04.005.

Harling, Guy et al. (2018). "Interviewer-Driven Variability in Social Network Reporting: Results from Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community (HAALSI) in South Africa". In: *Field methods* 30.2, pp. 140–154. ISSN: 1525-822X. DOI: 10.1177/1525822X18769498.

Hathi, Payal (Dec. 9, 2021). "The Inclusion of Stillbirths in Demographic Estimates of Child Mortality". In: *Session 169, Creating and Using International or Historical Datasets*. International Population Conference. International Union for the Scientific Study of Population.

Haws, Rachel A. et al. (2010). "These Are Not Good Things for Other People to Know: How Rural Tanzanian Womens Experiences of Pregnancy Loss and Early Neonatal Death May Impact Survey Data Quality". In: *Social Science & Medicine* 71.10, pp. 1764–1772. ISSN: 02779536. DOI: 10.1016/j.socscimed.2010.03.051.

Helleringer, Stéphane et al. (2020). "Biases in Survey Estimates of Neonatal Mortality: Results From a Validation Study in Urban Areas of Guinea-Bissau". In: *Demography*. ISSN: 0070-3370, 1533-7790. DOI: 10.1007/s13524-020-00911-6.

Herbst, Abraham J et al. (2009). "Adult Mortality and Antiretroviral Treatment Rollout in Rural KwaZulu-Natal, South Africa". In: *Bulletin of the World Health Organization* 87.10, pp. 754–762. ISSN: 00429686. DOI: 10.2471/BLT.08.058982.

Hernán, Miguel Ángel, Babette Brumback, and James M. Robins (2000). "Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men:" in: *Epidemiology* 11.5, pp. 561–570. ISSN: 1044-3983. DOI: 10.1097/00001648-200009000-00012.

Hill, Kenneth (1991). "Approaches to the Measurement of Childhood Mortality: A Comparative Review". In: *Population Index* 57.3, p. 368. ISSN: 00324701. DOI: 10.2307/3643873.

Hill, Kenneth and Agbessi Amouzou (2006). "Trends in Child Mortality, 1960 to 2000". In: *Disease and Mortality in Sub-Saharan Africa*. Washington, DC: The International Bank for Reconstruction and Development/The World Bank, p. 16. DOI: 10.1596/978-0-8213-6397-3.

Hill, Kenneth and Yoonjoung Choi (2006). "Neonatal Mortality in the Developing World". In: *Demographic Research* 14, pp. 429–452. ISSN: 1435-9871. DOI: 10.4054/DemRes.2006.14.18.

Hill, Kenneth et al. (2007). "Interim Measures for Meeting Needs for Health Sector Data: Births, Deaths, and Causes of Death". In: *The Lancet* 370.9600, pp. 1726–1735. ISSN: 01406736. DOI: 10.1016/S0140-6736(07)61309-9.

Holian, J. (2000). "Live Birth and Infant Death Record Linkage: Methodological and Policy Issues". In: *Journal of Health & Social Policy* 12.1, pp. 1–10. ISSN: 0897-7186. DOI: 10.1300/J045v12n01_01.

Holian, John, M. Joan Mallick, and Carolyn M. Zaremba (2004). "Maternity and Infant Care, Race and Birth Outcomes". In: *Journal of Health & Social Policy* 18.4, pp. 1–11. ISSN: 0897-7186. DOI: 10.1300/J045v18n04_01.

Hontelez, Jan A. C. et al. (2018). "HIV Treatment Substantially Decreases Hospitalization Rates: Evidence From Rural South Africa". In: *Health Affairs* 37.6, pp. 997–1004. ISSN: 0278-2715, 1544-5208. DOI: 10.1377/hlthaff.2017.0820.

Houle, Brian et al. (2015). "The Impacts of Maternal Mortality and Cause of Death on Childrens Risk of Dying in Rural South Africa: Evidence from a Population Based Surveillance Study (1992-2013)". In: *Reproductive Health* 12.S1, S7. ISSN: 1742-4755. DOI: 10.1186/1742-4755-12-S1-S7.

Houlihan, Catherine F et al. (2011). "Cohort Profile: Hlabisa HIV Treatment and Care Programme". In: *International Journal of Epidemiology* 40.2, pp. 318–326. ISSN: 0300-5771. DOI: 10.1093/ije/dyp402.

Howard, L (2004). "The Psychosocial Outcome of Pregnancy in Women with Psychotic Disorders". In: *Schizophrenia Research* 71.1, pp. 49–60. ISSN: 09209964. DOI: 10.1016/j.schres.2004.01.003.

Hug, Lucia et al. (2019). "National, Regional, and Global Levels and Trends in Neonatal Mortality between 1990 and 2017, with Scenario-Based Projections to 2030: A Systematic Analysis". In: *The Lancet Global Health* 7.6, e710–e720. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(19)30163-9.

Hung, Yuen W. et al. (2020). "Using Routine Health Information Data for Research in Low- and Middle-Income Countries: A Systematic Review". In: *BMC Health Services Research* 20.1, p. 790. ISSN: 1472-6963. DOI: 10.1186/s12913-020-05660-1.

Hunter, Lori M., Wayne Twine, and Aaron Johnson (2011). "Adult Mortality and Natural Resource Use in Rural South Africa: Evidence From the Agincourt Health and Demographic Surveillance Site". In: *Society & Natural Resources* 24.3, pp. 256–275. ISSN: 0894-1920, 1521-0723. DOI: 10.1080/08941920903443327.

IHME (2022). *Reproductive Health Survey (RHS) | GHDx*. Global Health Data Exchange. URL: http://ghdx.healthdata.org/series/reproductive-health-survey-rhs (visited on 03/14/2022).

INDEPTH Network (2002). *Population and Health in Developing Countries*. Volume 1: Population, health, and survival at INDEPTH sites vols. Ottawa, ON, Canada: International Development Research Centre. 1 p. ISBN: 0-88936-948-8.

— (2008a). *Electronic Data Capture*. INDEPTH Resource Kit for Demographic Surveillance Systems. URL: http://www.indepth-network.org/Resource%20Kit/INDEPTH%20DSS%20Resource%20Kit/Electronic_data_capture.htm (visited on 03/03/2022).

— (2008b). *INDEPTH Resource Kit for Demographic Surveillance Systems (Beta Version 0.9)*. URL: http://www.indepth-network.org/Resource%20Kit/INDEPTH%20DSS%20Resource%20Kit/INDEPTH%20DSS%20Resource%20Kit.htm (visited on 06/30/2021).

— (2021). *INDEPTH Network, Better Health Information for Better Health Policy*. URL: http://www.indepth-network.org (visited on 06/16/2021).

Issaka, Abukari I., Kingsley E. Agho, and Andre M. N. Renzaho (2016). "The Impact of Internal Migration on Under-Five Mortality in 27 Sub-Saharan African Countries". In: *PLoS ONE* 11.10. Ed. by Massimo Ciccozzi. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0163179.

Jahn, Andreas et al. (2008). "Population-Level Effect of HIV on Adult Mortality and Early Evidence of Reversal after Introduction of Antiretroviral Therapy in Malawi".

In: *Lancet* 371.9624, pp. 1603–1611. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(08)60693-5.

Jaro, Matthew A. (1989). "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida". In: *Journal of the American Statistical Association* 84.406, pp. 414–420. ISSN: 0162-1459. DOI: 10.1080/01621459.1989.10478785.

Jasseh, Momodou (2003). "Age Patterns of Mortality within Childhood in Sub-Saharan Africa". PhD thesis. London School of Hygiene & Tropical Medicine.

Jasseh, Momodou et al. (2015). "Health & Demographic Surveillance System Profile: Farafenni Health and Demographic Surveillance System in The Gambia". In: *International Journal of Epidemiology* 44.3, pp. 837–847. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dyv049.

Jia, Peng, Osman Sankoh, and Andrew J. Tatem (2015). "Mapping the Environmental and Socioeconomic Coverage of the INDEPTH International Health and Demographic Surveillance System Network". In: *Health & Place* 36, pp. 88–96. ISSN: 13538292. DOI: 10.1016/j.healthplace.2015.09.009.

Johnson, Kiersten et al. (2009). *Fieldwork-Related Factors and Data Quality in the Demographic and Health Surveys Program*. 19. Calverton, Maryland, USA: ICF Macro.

Johnson, Leigh F. et al. (2020). "Steep Declines in Pediatric AIDS Mortality in South Africa, Despite Poor Progress Toward Pediatric Diagnosis and Treatment Targets". In: *Pediatric Infectious Disease Journal* 39.9, pp. 843–848. ISSN: 0891-3668.

Johnson, Peter D., Nobuko Mizoguchi, and Athena Pantazis (Apr. 23, 2020). "Improved Method for Adjusting for Bias Due to HIV Mortality in Estimates of Child Mortality". In: *Population Association of America Annual Meeting*. Washington, DC.

Jones, Gareth et al. (2003). "How Many Child Deaths Can We Prevent This Year?" In: *The Lancet* 362.9377, pp. 65–71. DOI: 10.1016/S0140-6736(03)13811-1.

Joubert, Jané et al. (2014). "Record-Linkage Comparison of Verbal Autopsy and Routine Civil Registration Death Certification in Rural North-East South Africa: 2006-09". In: *International Journal of Epidemiology* 43.6, pp. 1945–1958. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dyu156.

Kabudula, Chodziwadziwa et al. (Sept. 28, 2017a). *LSHTM ALPHA Network / PIRL RecordLinkageSoftware: Initial Public Release of Point-of-contact Interactive Record Linkage (PIRL) Software*. Zenodo. DOI: 10.5281/zenodo.998868. URL: https://zenodo.org/record/998868 (visited on 11/05/2021).

Kabudula, Chodziwadziwa W et al. (2014). "The Promise of Record Linkage for Assessing the Uptake of Health Services in Resource Constrained Settings: A Pilot Study from South Africa". In: *BMC Medical Research Methodology* 14.1, p. 71. ISSN: 1471-2288. DOI: 10.1186/1471-2288-14-71.

Kabudula, Chodziwadziwa Whiteson et al. (2017b). "Using Record Linkage to Assess the Impact of Late Presentation for Care and Treatment on Short-Term Mortality

among HIV-infected Adults in Rural South Africa." In: *International Journal of Population Data Science* 1.1. ISSN: 2399-4908. DOI: 10.23889/ijpds.v1i1.398.

Kadobera, Daniel et al. (2012). "The Effect of Distance to Formal Health Facility on Childhood Mortality in Rural Tanzania, 2005-2007". In: *Global Health Action* 5.1. ISSN: 1654-9716, 1654-9880. DOI: 10.3402/gha.v5i0.19099.

Kadobera, Daniel et al. (2017). "Comparing Performance of Methods Used to Identify Pregnant Women, Pregnancy Outcomes, and Child Mortality in the Iganga-Mayuge Health and Demographic Surveillance Site, Uganda". In: *Global Health Action* 10.1. ISSN: 1654-9716, 1654-9880. DOI: 10.1080/16549716.2017.1356641.

Kahn, Kathleen et al. (2007). "Mortality Trends in a New South Africa: Hard to Make a Fresh Start1". In: *Scandinavian Journal of Public Health* 35.69, pp. 26–34. ISSN: 1403-4948, 1651-1905. DOI: 10.1080/14034950701355668.

Kahn, Kathleen et al. (2012). "Profile: Agincourt Health and Socio-demographic Surveillance System". In: *International Journal of Epidemiology* 41.4, pp. 988–1001. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dys115.

Kaplan, Richard et al. (2008). "Loss to Follow-up and Mortality amongst Pregnant Women Referred to a Community Clinic for Antiretroviral Treatment". In: *AIDS (London, England)* 22.13, pp. 1679–1681. ISSN: 0269-9370. DOI: 10.1097/QAD.0b013/e32830ebcee.

Kenya National Bureau of Statistics (2015). *Kenya Demographic and Health Survey 2014*. URL: https://www.dhsprogram.com/pubs/pdf/FR308/FR308.pdf.

Khan, Shane and Attila Hancioglu (2019). "Multiple Indicator Cluster Surveys: Delivering Robust Data on Children and Women across the Globe". In: *Studies in Family Planning* 50.3, pp. 279–286. ISSN: 0039-3665, 1728-4465. DOI: 10.1111/sifp.12103.

Kiguli, Juliet et al. (2015). "Weeping in Silence: Community Experiences of Stillbirths in Rural Eastern Uganda". In: *Global Health Action* 8.1. ISSN: 1654-9716, 1654-9880. DOI: 10.3402/gha.v8.24011.

Kishamawe, Coleman et al. (2015). "Health & Demographic Surveillance System Profile: The Magu Health and Demographic Surveillance System (Magu HDSS)". In: *International Journal of Epidemiology* 44.6, pp. 1851–1861. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dyv188.

Kitagawa, Evelyn M. (1955). "Components of a Difference Between Two Rates*". In: *Journal of the American Statistical Association* 50.272, pp. 1168–1194. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1955.10501299.

Kramer, M. S. (2002). "Analysis of Perinatal Mortality and Its Components: Time for a Change?" In: *American Journal of Epidemiology* 156.6, pp. 493–497. ISSN: 00029262. DOI: 10.1093/aje/kwf077.

Kupka, Roland et al. (2009). "Predictors of Stillbirth among HIV-infected Tanzanian Women". In: *Acta Obstetricia et Gynecologica Scandinavica* 88.5, pp. 584–592. ISSN: 1600-0412. DOI: 10.1080/00016340902835901.

Kwesiga, Doris et al. (2021). "Barriers and Enablers to Reporting Pregnancy and Adverse Pregnancy Outcomes in Population-Based Surveys: EN-INDEPTH Study". In: *Population Health Metrics* 19.1, p. 15. ISSN: 1478-7954. DOI: 10.1186/s12963-020-00228-x.

Kwon, Christie et al. (2021). "Pregnancy Surveillance Methods within Health and Demographic Surveillance Systems". In: *Gates Open Research* 5.144. DOI: 10.12688/gatesopenres.13332.1.

Lawn, Joy E, Simon Cousens, and Jelka Zupan (2005). "4 Million Neonatal Deaths: When? Where? Why?" In: *The Lancet* 365.9462, pp. 891–900. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(05)71048-5.

Lawn, Joy E. et al. (2008). "Four Million Neonatal Deaths: Counting and Attribution of Cause of Death". In: *Paediatric and Perinatal Epidemiology* 22.5, pp. 410–416. ISSN: 02695022, 13653016. DOI: 10.1111/j.1365-3016.2008.00960.x.

Lawn, Joy E et al. (2016). "Stillbirths: Rates, Risk Factors, and Acceleration towards 2030". In: *The Lancet* 387.10018, pp. 587–603. ISSN: 01406736. DOI: 10.1016/S0140-6736(15)00837-5.

Ley, Benedikt et al. (2019). "Analysis of Erroneous Data Entries in Paper Based and Electronic Data Collection". In: *BMC Research Notes* 12.1, p. 537. ISSN: 1756-0500. DOI: 10.1186/s13104-019-4574-8.

Liu, Li, Kenneth Hill, and Shefali Oza (2016). "Levels and Causes of Mortality under Age Five Years". In: *Reproductive, Maternal, Newborn, and Child Health: Disease Control Priorities*. 3rd ed. Vol. 2.

Liu, Li et al. (2016). "Understanding Misclassification between Neonatal Deaths and Stillbirths: Empirical Evidence from Malawi". In: *PLoS ONE* 11.12. Ed. by Jacobus P. van Wouwe, e0168743. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0168743.

Lockwood, Matthew (1990). "Demographic Evidence for Central Africa - Demography from Scanty Evidence: Central Africa in the Colonial Era". In: *The Journal of African History* 34.1, pp. 161–162. ISSN: 0021-8537, 1469-5138. DOI: 10.1017/S0021853700033156.

Maher, D and J Sekajugo (2010). "Health Transition in Africa: Practical Policy Proposals for Primary Care". In: *Bulletin of the World Health Organization* 88.12, pp. 943–948. ISSN: 00429686. DOI: 10.2471/BLT.10.077891.

Maher, D. et al. (2010). "Translating Global Health Research Aims into Action: The Example of the ALPHA Network*". In: *Tropical Medicine & International Health* 15.3, pp. 321–328. ISSN: 1365-3156. DOI: 10.1111/j.1365-3156.2009.02456.x.

Mahy, Mary (2003). *Measuring Child Mortality in AIDS-affected Countries*. New York: United Nations Population Division, p. 9.

Mahy, Mary et al. (2017). "Improving Estimates of Children Living with HIV from the Spectrum AIDS Impact Model". In: *AIDS* 31 (Supplement 1), S13–S22. ISSN: 0269-9370. DOI: 10.1097/QAD.0000000000001306.

Marsh, Andrew D et al. (2020). "Effective Coverage Measurement in Maternal, Newborn, Child, and Adolescent Health and Nutrition: Progress, Future Prospects, and Implications for Quality Health Systems". In: *The Lancet Global Health* 8.5, e730–e736. ISSN: 2214109X. DOI: 10.1016/S2214-109X(20)30104-2.

Marston, M., B. Zaba, and J. W. Eaton (2017). "The Relationship between HIV and Fertility in the Era of Antiretroviral Therapy in Sub-Saharan Africa: Evidence from 49 Demographic and Health Surveys". In: *Tropical Medicine & International Health* 22.12, pp. 1542–1550. ISSN: 1365-3156. DOI: 10.1111/tmi.12983.

Marston, M. et al. (2011). "Net Survival of Perinatally and Postnatally HIV-infected Children: A Pooled Analysis of Individual Data from Sub-Saharan Africa". In: *International Journal of Epidemiology* 40.2, pp. 385–396. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dyq255.

Marston, Milly et al. (2012). "The Impact of Antiretroviral Therapy on Adult Mortality in Rural Tanzania". In: *Tropical Medicine & International Health* 17.8, e58–e65. ISSN: 1365-3156. DOI: 10.1111/j.1365-3156.2011.02924.x.

Marston, Milly et al. (2017). "The Effects of HIV on Fertility by Infection Duration: Evidence from African Population Cohorts before Antiretroviral Treatment Availability". In: *AIDS* 31 (Supplement 1), S69–S76. ISSN: 0269-9370. DOI: 10.1097/QAD.0000000000001305.

Masquelier, Bruno, Georges Reniers, and Gilles Pison (2014). "Divergences in Trends in Child and Adult Mortality in Sub-Saharan Africa: Survey Evidence on the Survival of Children and Siblings". In: *Population Studies* 68.2, pp. 161–177. ISSN: 0032-4728, 1477-4747. DOI: 10.1080/00324728.2013.856458.

Mathers, Colin D et al. (2005). "Counting the Dead and What They Died from: An Assessment of the Global Status of Cause of Death Data". In: *Bulletin of the World Health Organization* 83.3, pp. 171–177. ISSN: 0042-9686.

McFadden, Daniel (1974). "Conditional Logit Analysis of Qualitative Choice Behavior". In: *Frontiers in Econometrics*. P. Zarembka. New York: Academic Press, pp. 105–142.

— (1977). *Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments*. Discussion Paper 474. Cowles Foundation for Research in Economics, pp. 1–35.

McGregor, I. A., W. Z. Billewicz, and A. M. Thomson (1961). "Growth and Mortality in Children in an African Village". In: *BMJ* 2.5268, pp. 1661–1666. ISSN: 0959-8138, 1468-5833. DOI: 10.1136/bmj.2.5268.1661.

McGregor, I. A. et al. (1979). *Mortality in a Rural West African Village (Keneba) with Special Reference to Deaths Occurring in the First Five Years of Life*. Medical Research Council Laboratories, The Gambia.

McLean, Estelle et al. (2017). "Implementing Electronic Data Capture at a Well-Established Health and Demographic Surveillance Site in Rural Northern Malawi". In: *Global*

Health Action 10.1, p. 1367162. ISSN: 1654-9716, 1654-9880. DOI: 10.1080/16549716.2017.1367162.

Mehler, Katrin et al. (2016). "Survival Among Infants Born at 22 or 23 Weeks Gestation Following Active Prenatal and Postnatal Care". In: *JAMA Pediatrics* 170.7, p. 671. DOI: 10.1001/jamapediatrics.2016.0207.

Menashe-Oren, Ashira and Philippe Bocquier (2021). "Urbanization Is No Longer Driven by Migration in Low- and Middle-Income Countries (19852015)". In: *Population and Development Review* 47.3, pp. 639–663. ISSN: 1728-4457. DOI: 10.1111/padr.12407.

MICS (2022). *About MICS*. UNICEF MICS. URL: https://mics.unicef.org/about (visited on 03/14/2022).

Mikkelsen, Lene et al. (2015). "A Global Assessment of Civil Registration and Vital Statistics Systems: Monitoring Data Quality and Progress". In: *The Lancet* 386.10001, pp. 1395–1406. ISSN: 01406736. DOI: 10.1016/S0140-6736(15)60171-4.

Modiba, L and A Gw Nolte (2007). "The Experiences of Mothers Who Lost a Baby during Pregnancy". In: *Health SA Gesondheid* 12.2. ISSN: 2071-9736, 1025-9848. DOI: 10.4102/hsag.v12i2.245.

Mohangoo, Ashna D. et al. (2013). "International Comparisons of Fetal and Neonatal Mortality Rates in High-Income Countries: Should Exclusion Thresholds Be Based on Birth Weight or Gestational Age?" In: *PLoS ONE* 8.5. Ed. by Linda Wright, e64869. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0064869.

Molyneux, Malcolm and Elizabeth Molyneux (2016). "Reaching Millennium Development Goal 4". In: *The Lancet Global Health* 4.3, e146–e147. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(16)00009-7.

MRC Centre for Global Infectious Disease Analysis (2020). *Mrc-Ide/Demogsurv*. URL: https://github.com/mrc-ide/demogsurv (visited on 06/10/2020).

Murray, C. J. L. et al. (2003). "Modified Logit Life Table System: Principles, Empirical Validation, and Application". In: *Population Studies* 57.2, pp. 165–182. ISSN: 0032-4728. DOI: 10.1080/0032472032000097083.

Mutale, Wilbroad et al. (2013). "Improving Health Information Systems for Decision Making across Five Sub-Saharan African Countries: Implementation Strategies from the African Health Initiative". In: *BMC Health Services Research* 13.S2, S9. ISSN: 1472-6963. DOI: 10.1186/1472-6963-13-S2-S9.

Mutemaringa, Themba (2011). "Impact of HIV on Estimates of Child Mortality Derived Using the Summary Birth History (CEB/CS) Method". MPhil thesis. Cape Town: Centre for Actuarial Research: University of Cape Town.

Nakiyingi, Jessica S et al. (2003). "Child Survival in Relation to Mothers HIV Infection and Survival: Evidence from a Ugandan Cohort Study". In: *AIDS* 17.12. DOI: 10.1097/00002030-200308150-00012.

Nareeba, Tryphena et al. (2021). "Neonatal and Child Mortality Data in Retrospective Population-Based Surveys Compared with Prospective Demographic Surveillance: EN-INDEPTH Study". In: *Population Health Metrics* 19.1, p. 7. ISSN: 1478-7954. DOI: 10.1186/s12963-020-00232-1.

Ngom, P. (2001). "Demographic Surveillance and Health Equity in Sub-Saharan Africa". In: *Health Policy and Planning* 16.4, pp. 337–344. ISSN: 14602237. DOI: 10.1093/heapol/16.4.337.

Ng'weshemi, Japheth et al. (2003). "HIV Impact on Mother and Child Mortality in Rural Tanzania:" in: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 33.3, pp. 393–404. ISSN: 1525-4135. DOI: 10.1097/00126334-200307010-00015.

Nicoll, A. et al. (1994). "The Impact of HIV-1 Infection on Mortality in Children under 5 Years of Age in Sub-Saharan Africa: A Demographic and Epidemiologic Analysis". In: *AIDS (London, England)* 8.7, pp. 995–1005. ISSN: 0269-9370. DOI: 10.1097/00002030-199407000-00019.

Notestein, Frank W. (1945). "Population- the Long View". In: *Food for the World*. Chicago: University of Chicago Press, pp. 36–57.

Notkola, Veijo (2017). "The Causes of Changes in Fertility in Northern Namibia: Ovamboland, 1927-2010, and Kavango Region, 1935-1979". In: *Finnish Yearbook of Population Research* 51, pp. 23–40. DOI: 10.23979/fypr.60262.

Odhiambo, F. O. et al. (2012). "Profile: The KEMRI/CDC Health and Demographic Surveillance System–Western Kenya". In: *International Journal of Epidemiology* 41.4, pp. 977–987. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dys108.

Odimegwu, Clifford, Vesper H. Chisumpa, and Oluwaseyi Dolapo Somefun (2018). "Adult Mortality in Sub-Saharan Africa Using 2001-2009 Census Data: Does Estimation Method Matter?" In: *Genus* 74.1. ISSN: 2035-5556. DOI: 10.1186/s41118-017-0025-3.

Oduro, A. R. et al. (2012). "Profile of the Navrongo Health and Demographic Surveillance System". In: *International Journal of Epidemiology* 41.4, pp. 968–976. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dys111.

Omran, Abdel R (1971). "The Epidemiologic Transition: A Theory of the Epidemiology of Population Change". In: *Milbank Memorial Fund Quarterly* 49.4, pp. 509–538.

Oomman, Nandini et al. (2013). "Modernising Vital Registration Systems: Why Now?" In: *The Lancet* 381.9875, pp. 1336–1337. ISSN: 01406736. DOI: 10.1016/S0140-6736(13)60847-8.

Ouma, Peter O et al. (2010). "Antenatal and Delivery Care in Rural Western Kenya: The Effect of Training Health Care Workers to Provide "Focused Antenatal Care"". In: *Reproductive Health* 7.1. ISSN: 1742-4755. DOI: 10.1186/1742-4755-7-1.

Oza, Shefali, Simon N Cousens, and Joy E Lawn (2014). "Estimation of Daily Risk of Neonatal Death, Including the Day of Birth, in 186 Countries in 2013: A Vital-Registration and Modelling-Based Study". In: *The Lancet Global Health* 2.11, pp. 635–644. ISSN: 2214109X. DOI: 10.1016/S2214-109X(14)70309-2.

Oza, Shefali et al. (2015). "Neonatal Cause-of-Death Estimates for the Early and Late Neonatal Periods for 194 Countries: 2000-2013". In: *Bulletin of the World Health Organization* 93.1, pp. 19–28. ISSN: 0042-9686. DOI: 10.2471/BLT.14.139790.

Phillips, David E et al. (2015). "Are Well Functioning Civil Registration and Vital Statistics Systems Associated with Better Health Outcomes?" In: *The Lancet* 386.10001, pp. 1386–1394. ISSN: 01406736. DOI: 10.1016/S0140-6736(15)60172-6.

Pison, Gilles and Andre Langaney (1985). "The Level and Age Pattern of Mortality in Bandafassi (Eastern Senegal): Results From a Small-Scale and Intensive Multi-Round Survey". In: *Population Studies* 39.3, pp. 387–405. ISSN: 0032-4728, 1477-4747. DOI: 10.1080/0032472031000141586.

Pozzi, Lucia and Diego Ramiro Fariñas (2016). "Infant and Child Mortality in the Past". In: *Annales de démographie historique* 129.1, pp. 55–75. ISSN: 0066-2062. DOI: 10.3917/adh.129.0055.

Preston, Samuel H (1975). "The Changing Relation between Mortality and Level of Economic Development". In: *Population Studies* 29. ISSN: 1464-3685, 0300-5771. DOI: 10.1093/ije/dym075.

Preston, Samuel H. (1980). "Causes and Consequences of Mortality Declines in Less Developed Countries during the Twentieth Century". In: *Population and Economic Change in Developing Countries*. Ed. by Richard Ainley Easterlin. Chicago: University of Chicago Press, pp. 289–360. ISBN: 0-226-18027-1.

Pullum, Thomas and Stan Becker (2014). *DHS Methodological Reports 11: Evidence Omission and Displacement in DHS Birth Histories*. Rockville, Maryland, USA: ICF International.

Pullum, Thomas and Sarah Staveteig (2017). *DHS Methodological Reports 19: An Assessment of the Quality and Consistency of Age and Date Reporting in DHS Surveys, 2000-2015*. Rockville, Maryland, USA: ICF.

Pullum, Thomas W. (2006). *DHS Methodological Reports 5: An Assessment of Age and Date Reporting in the DHS Surveys, 1985-2003*. Calverton, Maryland, USA: Macro International Inc.

Pullum, Thomas W. et al. (Aug. 28, 2013). "An Assessment of DHS Estimates of Fertility and Under-Five Mortality". In: *Session 132: Data Quality in Demographic Surveys*. 2013 International Population Conference of the International Union for the Scientific Study of Population (IUSSP), p. 34.

Quattrochi, John et al. (2019). "Measuring and Correcting Bias in Indirect Estimates of Under-5 Mortality in Populations Affected by HIV/AIDS: A Simulation Study". In: *BMC Public Health* 19.1. ISSN: 1471-2458. DOI: 10.1186/s12889-019-7780-3.

Reher, David S. (2004). "The Demographic Transition Revisited as a Global Process". In: *Population, Space and Place* 10.1, pp. 19–41. ISSN: 1544-8452. DOI: 10.1002/psp.313.

Reniers, Georges et al. (2014). "Mortality Trends in the Era of Antiretroviral Therapy: Evidence from the Network for Analysing Longitudinal Population Based HIV/AIDS Data on Africa (ALPHA)". In: *AIDS* 28 (Supplement 4), S533–S542. ISSN: 0269-9370. DOI: 10.1097/QAD.0000000000000496.

Reniers, Georges et al. (2016). "Data Resource Profile: Network for Analysing Longitudinal Population-based HIV/AIDS Data on Africa (ALPHA Network)". In: *International Journal of Epidemiology* 45.1, pp. 83–93. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dyv343.

Rentsch, Christopher T. et al. (2017a). "Point-of-Contact Interactive Record Linkage (PIRL): A Software Tool to Prospectively Link Demographic Surveillance and Health Facility Data". In: *Gates Open Research* 1, p. 8. ISSN: 2572-4754. DOI: 10.12688/gatesopenres.12751.1.

Rentsch, Christopher T. et al. (2017b). "Point-of-Contact Interactive Record Linkage (PIRL) between Demographic Surveillance and Health Facility Data in Rural Tanzania". In: *International Journal of Population Data Science* 2.1. ISSN: 2399-4908. DOI: 10.23889/ijpds.v2i1.408.

Rentsch, Christopher T. et al. (2018). "Linkage to Care and Antiretroviral Therapy Initiation by Testing Modality among Individuals Newly Diagnosed with HIV in Tanzania, 2014-2017". In: *Tropical Medicine & International Health* 23.12, pp. 1384–1393. ISSN: 13602276. DOI: 10.1111/tmi.13153.

Rerimoi, Anne J. (2019). "Development and Assessment of Systems for Population-Based Estimation of Neonatal and Perinatal Mortality in The Gambia". PhD thesis.

Rerimoi, Anne J. et al. (2019). "Under-Five Mortality in The Gambia: Comparison of the Results of the First Demographic and Health Survey with Those from Existing Inquiries". In: *PLoS ONE* 14.7. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0219919.

Riordan, Daniel Vincent et al. (2012). "Family Size and Perinatal Circumstances, as Mental Health Risk Factors in a Scottish Birth Cohort". In: *Social Psychiatry and Psychiatric Epidemiology* 47.6, pp. 975–983. ISSN: 1433-9285. DOI: 10.1007/s00127-011-0405-5.

Risher, Kathryn A et al. (2021). "Age Patterns of HIV Incidence in Eastern and Southern Africa: A Modelling Analysis of Observational Population-Based Cohort Studies". In: *The Lancet HIV* 8.7, pp. 429–439. ISSN: 23523018. DOI: 10.1016/S2352-3018(21)00069-2.

Robins, James M., Miguel Ángel Hernán, and Babette Brumback (2000). "Marginal Structural Models and Causal Inference in Epidemiology". In: *Epidemiology* 11.5, pp. 550–560. ISSN: 1044-3983. DOI: 10.1097/00001648-200009000-00011.

Ross, Amanda et al. (2004). "HIV-1 Disease Progression and Fertility: The Incidence of Recognized Pregnancy and Pregnancy Outcome in Uganda". In: *AIDS* 18.5, pp. 799–804. ISSN: 0269-9370. DOI: 10.1097/01.aids.0000111419.91384.50.

Rossier, C. et al. (2012). "Profile: The Ouagadougou Health and Demographic Surveillance System". In: *International Journal of Epidemiology* 41.3, pp. 658–666. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dys090.

Rusnock, Andrea (2009). *Vital Counts: Quantifying Infant Mortality in England and France, 1750-1800*. ISBN: 9780521101233.

Rutstein, Shea O (2000). "Factors Associated with Trends in Infant and Child Mortality in Developing Countries during the 1990s". In: *Bulletin of the World Health Organization* 78.10, pp. 1256–1270. ISSN: 0042-9686.

Rutstein, Shea O. and Guillermo Rojas (2006). *Guide to DHS Statistics*. Calverton, Maryland, USA: Demographic and Health Surveys, ORC Macro.

Sankoh, Osman (2015). "CHESS: An Innovative Concept for a New Generation of Population Surveillance". In: *The Lancet Global Health* 3.12, e742. ISSN: 2214109X. DOI: 10.1016/S2214-109X(15)00180-1.

— (2017). "Why Population-Based Data Are Crucial to Achieving the Sustainable Development Goals". In: *International Journal of Epidemiology* 46.1, pp. 4–7. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dyx010.

Sankoh, Osman and Peter Byass (2012). "The INDEPTH Network: Filling Vital Gaps in Global Epidemiology". In: *International Journal of Epidemiology* 41.3, pp. 579–588. ISSN: 1464-3685, 0300-5771. DOI: 10.1093/ije/dys081.

Sartorius, Benn et al. (2011). "Survived Infancy but Still Vulnerable: Spatial-Temporal Trends and Risk Factors for Child Mortality in the Agincourt Rural Sub-District, South Africa, 1992-2007". In: *Geospatial health* 5.2, p. 285. ISSN: 1970-7096, 1827-1987. DOI: 10.4081/gh.2011.181.

Schoenfeld, David (1982). "Partial Residuals for the Proportional Hazards Regression Model". In: *Biometrika* 69.1, pp. 239–241. ISSN: 0006-3444. DOI: 10.1093/biomet/69.1.239.

Schoumaker, Bruno (Mar. 31–Apr. 2, 2011). "Omissions of Births in DHS Birth Histories in Sub-Saharan Africa: Measurement and Determinants". In: Annual Meeting, Population Association of America. Washington, DC.

— (2014). *DHS Methodological Reports 12: Quality and Consistency of DHS Fertility Estimates, 1990 to 2012*. DHS Methodological Reports. ICF International.

Silva, Romesh (2012). "Child Mortality Estimation: Consistency of Under-Five Mortality Rate Estimates Using Full Birth Histories and Summary Birth Histories". In: *PLoS Medicine* 9.8, e1001296. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001296.

Slogrove, Amy L et al. (2020). "Estimates of the Global Population of Children Who Are HIV-exposed and Uninfected, 200018: A Modelling Study". In: *The Lancet Global Health* 8.1, pp. 67–75. ISSN: 2214109X. DOI: 10.1016/S2214-109X(19)30448-6.

Snow, Robert W. et al. (2012). "The Changing Limits and Incidence of Malaria in Africa". In: *Advances in Parasitology* 78, pp. 169–262. DOI: 10.1016/B978-0-12-394303-3.00010-4.

Snow, Robert W. et al. (2017). "The Prevalence of Plasmodium Falciparum in Sub-Saharan Africa since 1900". In: *Nature* 550.7677, pp. 515–518. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature24059.

Spira, Rosemary et al. (1999). "Natural History of Human Immunodefiency Virus Type 1 Infection in Children: A Five-Year Prospective Study in Rwanda". In: *Pediatrics* 104.5. ISSN: 0031-4005, 1098-4275. DOI: 10.1542/peds.104.5.e56.

Stanton, Cynthia et al. (2006). "Stillbirth Rates: Delivering Estimates in 190 Countries". In: *The Lancet* 367.9521, pp. 1487–1494. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(06)68586-3.

Stoll, Barbara J. et al. (2015). "Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993-2012". In: *JAMA* 314.10, p. 1039. ISSN: 0098-7484. DOI: 10.1001/jama.2015.10244.

Stover, John, Tim Brown, and Milly Marston (2012). "Updates to the Spectrum / Estimation and Projection Package (EPP) Model to Estimate HIV Trends for Adults and Children". In: *Sexually Transmitted Infections* 88 (Suppl 2), pp. i11–i16. ISSN: 1368-4973, 1472-3263. DOI: 10.1136/sextrans-2012-050640.

Streatfield, P. Kim et al. (2014a). "Adult Non-Communicable Disease Mortality in Africa and Asia: Evidence from INDEPTH Health and Demographic Surveillance System Sites". In: *Global Health Action* 7.1, p. 25365. ISSN: 1654-9716, 1654-9880. DOI: 10.3402/gha.v7.25365.

Streatfield, P. Kim et al. (2014b). "HIV/AIDS-related Mortality in Africa and Asia: Evidence from INDEPTH Health and Demographic Surveillance System Sites". In: *Global Health Action* 7.1, p. 25370. ISSN: 1654-9716, 1654-9880. DOI: 10.3402/gha.v7.25370.

Tanser, F. et al. (2008). "Cohort Profile: Africa Centre Demographic Information System (ACDIS) and Population-Based HIV Survey". In: *International Journal of Epidemiology* 37.5, pp. 956–962. ISSN: 0300-5771, 1464-3685. DOI: 10.1093/ije/dym211.

Tatem, Andrew J., Robert W. Snow, and Simon I. Hay (2006). "Mapping the Environmental Coverage of the INDEPTH Demographic Surveillance System Network in Rural Africa". In: *Tropical Medicine and International Health* 11.8, pp. 1318–1326. ISSN: 1360-2276, 1365-3156. DOI: 10.1111/j.1365-3156.2006.01681.x.

Teitelbaum, Michael S. (1975). "Relevance of Demographic Transition Theory for Developing Countries: The Theory Offers Only Partial Explanation of European Trends

and Ambiguous Advice for Developing Countries." In: *Science* 188.4187, pp. 420–425. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1124392.

Thornton, Arland (2001). "The Developmental Paradigm, Reading History Sideways, and Family Change". In: *Demography* 38.4, pp. 449–465. DOI: 10.1353/dem.2001.0039.

Thysen, Sanne M. et al. (2021). "Electronic Data Collection in a Multi-Site Population-Based Survey: EN-INDEPTH Study". In: *Population Health Metrics* 19.S1, p. 9. ISSN: 1478-7954. DOI: 10.1186/s12963-020-00226-z.

Timaeus, I. M. (1998). "Impact of the HIV Epidemic on Mortality in Sub-Saharan Africa: Evidence from National Surveys and Censuses". In: *AIDS (London, England)* 12 Suppl 1, S15–27. ISSN: 0269-9370.

Todd, Jim et al. (2007). "Time from HIV Seroconversion to Death: A Collaborative Analysis of Eight Studies in Six Low and Middle-Income Countries before Highly Active Antiretroviral Therapy". In: *AIDS* 21 (Suppl 6), S55–S63. ISSN: 0269-9370. DOI: 10.1097/01.aids.0000299411.75269.e8.

Tollman, Stephen M. and William M. Pick (2002). "Roots, Shoots, but Too Little Fruit: Assessing the Contribution of COPC in South Africa". In: *American Journal of Public Health* 92.11, pp. 1725–1728. ISSN: 0090-0036, 1541-0048. DOI: 10.2105/AJPH.92.11.1725.

Tollman, Stephen M et al. (2008). "Implications of Mortality Transition for Primary Health Care in Rural South Africa: A Population-Based Surveillance Study". In: *The Lancet* 372.9642, pp. 893–901. ISSN: 01406736. DOI: 10.1016/S0140-6736(08)61399-9.

Tollman, Stephen M. et al. (2021). "Count Every Newborn: EN-INDEPTH Study to Improve Pregnancy Outcome Measurement in Population-Based Surveys". In: *Population Health Metrics* 19.S1, p. 5. ISSN: 1478-7954. DOI: 10.1186/s12963-020-00243-y.

UKRI (2022). *Basse Health and Demographic Surveillance System Overview*. UK Research and Innovation (UKRI) Gateway to Research Portal. URL: https://gtr.ukri.org/projects?ref=MC_UP_1003%2F1#/tabOverview (visited on 03/09/2022).

UN IGME (2013). *Levels & Trends in Child Mortality: Report 2013, Estimates Developed by the United Nations Inter-agency Group for Child Mortality Estimation*. New York: United Nations Children's Fund.

— (2015). *Levels & Trends in Child Mortality: Report 2015, Estimates Developed by the United Nations Inter-agency Group for Child Mortality Estimation*. New York: United Nations Children's Fund.

— (2018). *Levels & Trends in Child Mortality: Report 2018, Estimates Developed by the United Nations Inter-agency Group for Child Mortality Estimation*. New York: United Nations Children's Fund.

UN IGME (2020a). *A Neglected Tragedy: The Global Burden of Stillbirths*. United Nations Children's Fund.

— (2020b). *Levels & Trends in Child Mortality: Report 2020, Estimates Developed by the United Nations Inter-agency Group for Child Mortality Estimation*. New York: United Nations Children's Fund.

— (2021). *Levels & Trends in Child Mortality: Report 2021, Estimates Developed by the United Nations Inter-agency Group for Child Mortality Estimation*. New York: United Nations Children's Fund.

UNAIDS (2020). *Gambia Country Factsheet*. URL: https://www.unaids.org/en/regionscountries/countries/gambia (visited on 08/03/2021).

— (2021). *UNAIDS Data 2021*. UNAIDS 2021 Reference. Joint United Nations Programme on HIV/AIDS. URL: https://www.unaids.org/sites/default/files/media_asset/JC3032_AIDS_Data_book_2021_En.pdf (visited on 12/17/2021).

United Nations (1955). *Age and Sex Patterns of Mortality: Model Life-Tables for under-Developed Countries*. Population Studies 22. United Nations, Dept. of Social Affairs, Population Branch.

— (1982). *Model Life Tables for Developing Countries*. Population Studies / Department of International Economic and Social Affairs no. 77. New York: United Nations. 351 pp.

— (1986). *Determinants of Mortality Change and Differentials in Developing Countries: The Five Country Case Study Project*. Population Studies 94. New York: United Nations, Department of International Economic and Social Affairs.

— (2000). *United Nations Millennium Declaration*. Resolution adopted by the General Assembly.

— (2017). "Transforming Our World: The 2030 Agenda for Sustainable Development". In: *A New Era in Global Health*. Ed. by William Rosa. New York, NY: Springer Publishing Company. ISBN: 978-0-8261-9011-6.

Utazi, C Edson et al. (2018). "Geographic Coverage of Demographic Surveillance Systems for Characterising the Drivers of Childhood Mortality in Sub-Saharan Africa". In: *BMJ Global Health* 3.2, e000611. ISSN: 2059-7908. DOI: 10.1136/bmjgh-2017-000611.

Vandormael, Alain et al. (2014). "Use of Antiretroviral Therapy in Households and Risk of HIV Acquisition in Rural KwaZulu-Natal, South Africa, 200412: A Prospective Cohort Study". In: *The Lancet Global Health* 2.4, e209–e215. ISSN: 2214109X. DOI: 10.1016/S2214-109X(14)70018-X.

Vandormael, Alain et al. (2019). "Declines in HIV Incidence among Men and Women in a South African Population-Based Cohort". In: *Nature Communications* 10.1, pp. 1–10. ISSN: 2041-1723. DOI: 10.1038/s41467-019-13473-y.

Verhulst, Andrea et al. (May 5, 2021). "Identifying New Age Patterns of Under-Five Mortality Using the Under-Five Mortality Database". In: *Methodological Innovations in Mortality*. Population Association of America Annual Meeting 2021.

Waiswa, Peter et al. (2019). "Status of Birth and Pregnancy Outcome Capture in Health Demographic Surveillance Sites in 13 Countries". In: *International Journal of Public Health* 64.6, pp. 909–920. ISSN: 1661-8556, 1661-8564. DOI: 10.1007/s00038-019-01241-0.

Walker, Neff, Kenneth Hill, and Fengmin Zhao (2012). "Child Mortality Estimation: Methods Used to Adjust for Bias Due to AIDS in Estimating Trends in Under-Five Mortality". In: *PLoS Medicine* 9.8. Ed. by Peter Byass, e1001298. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001298.

Walters, Sarah (2016). "Counting Souls: Towards an Historical Demography of Africa". In: *Demographic Research* 34, pp. 63–108. ISSN: 1435-9871. DOI: 10.4054/DemRes.2016.34.3.

— (2021). "African Population History: Contributions of Moral Demography". In: *The Journal of African History* 62.2, pp. 183–200. ISSN: 0021-8537, 1469-5138. DOI: 10.1017/S002185372100044X.

Wang, Bingxia et al. (2011). "Loss to Follow-Up in a Community Clinic in South Africa Roles of Gender, Pregnancy and CD4 Count". In: *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde* 101.4, pp. 253–257. ISSN: 0256-9574.

Wang, Haidong et al. (2014). "Global, Regional, and National Levels of Neonatal, Infant, and under-5 Mortality during 1990-2013: A Systematic Analysis for the Global Burden of Disease Study 2013". In: *The Lancet* 384.9947, pp. 957–979. ISSN: 01406736. DOI: 10.1016/S0140-6736(14)60497-9.

Wang, Haidong et al. (2016). "Global, Regional, National, and Selected Subnational Levels of Stillbirths, Neonatal, Infant, and under-5 Mortality, 1980-2015: A Systematic Analysis for the Global Burden of Disease Study 2015". In: *The Lancet* 388.10053, pp. 1725–1774. ISSN: 01406736. DOI: 10.1016/S0140-6736(16)31575-6.

Wang, Haidong et al. (2017). "Global, Regional, and National under-5 Mortality, Adult Mortality, Age-Specific Mortality, and Life Expectancy, 1970-2016: A Systematic Analysis for the Global Burden of Disease Study 2016." In: *The Lancet* 390.10100, pp. 1084–1150. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(17)31833-0.

Ward, Patrick and Basia Zaba (2008). "The Effect of HIV on the Estimation of Child Mortality Using the Children Surviving/Children Ever Born Technique". In: *Southern African Journal of Demography* 11.1, pp. 39–73.

Warren, Ashley E et al. (2013). "Global Health Initiative Investments and Health Systems Strengthening: A Content Analysis of Global Fund Investments". In: *Globalization and Health* 9.1, p. 30. ISSN: 1744-8603. DOI: 10.1186/1744-8603-9-30.

Watson, Oliver J., Rich FitzJohn, and Jeffrey W. Eaton (2019). "Rdhs: An R Package to Interact with The Demographic and Health Surveys (DHS) Program Datasets". In: *Wellcome Open Research* 4, p. 103. ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.15311.1.

Welz, Tanya et al. (2007). "Continued Very High Prevalence of HIV Infection in Rural KwaZulu-Natal, South Africa: A Population-Based Longitudinal Study". In: *AIDS* 21.11, pp. 1467–1472. ISSN: 0269-9370. DOI: 10.1097/QAD.0b013e3280ef6af2.

White, Lisa J. et al. (Mar. 7, 2012). "Estimation of Gestational Age from Fundal Height: A Solution for Resource-Poor Settings". In: *Journal of the Royal Society Interface* 9.68, pp. 503–510. ISSN: 1742-5689. DOI: 10.1098/rsif.2011.0376.

Wilmoth, John et al. (2012). "A Flexible Two-Dimensional Mortality Model for Use in Indirect Estimation". In: *Population Studies* 66.1, pp. 1–28. ISSN: 1477-4747. DOI: 10.1080/00324728.2011.611411.

Woldesenbet, S.A. et al. (2021). *The 2019 National Antenatal Sentinel HIV Survey*. South Africa: National Department of Health.

World Bank (2022). *Completeness of Birth Registration (%) - Sub-Saharan Africa | Data*. URL: https://data.worldbank.org/indicator/SP.REG.BRTH.ZS?end=2017&locations=ZG&start=2010&view=chart (visited on 01/28/2022).

World Health Organization (2013). *International Statistical Classification of Diseases and Related Health Problems*. 2010th ed. Vol. 2.

— (2014). *Every Newborn: An Action Plan to End Preventable Deaths*. Geneva: World Health Organization. ISBN: 978-92-4-150744-8.

— (2016). *WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience*. Geneva: World Health Organization. ISBN: 978-92-4-154991-2.

Wringe, Alison et al. (2012). "Antiretroviral Therapy Uptake and Coverage in Four HIV Community Cohort Studies in Sub-Saharan Africa". In: *Tropical Medicine & International Health* 17.8. ISSN: 1360-2276, 1365-3156. DOI: 10.1111/j.1365-3156.2011.02925.x.

Xu, Stanley et al. (2010). "Use of Stabilized Inverse Propensity Scores as Weights to Directly Estimate Relative Risk and Its Confidence Intervals". In: *Value in Health* 13.2, pp. 273–277. ISSN: 10983015. DOI: 10.1111/j.1524-4733.2009.00671.x.

Ye, Yazoume et al. (2012). "Health and Demographic Surveillance Systems: A Step towards Full Civil Registration and Vital Statistics System in Sub-Sahara Africa?" In: *BMC Public Health* 12.1, p. 741. ISSN: 1471-2458. DOI: 10.1186/1471-2458-12-741.

You, Danzhen et al. (2015). "Global, Regional, and National Levels and Trends in under-5 Mortality between 1990 and 2015, with Scenario-Based Projections to 2030: A Systematic Analysis by the UN Inter-agency Group for Child Mortality Estimation". In: *The Lancet* 386.10010, pp. 2275–2286. ISSN: 01406736. DOI: 10.1016/S0140-6736(15)00120-8.

Zaba, B. and S. Gregson (1998). "Measuring the Impact of HIV on Fertility in Africa". In: *AIDS (London, England)* 12 Suppl 1, S41–50. ISSN: 0269-9370.

Zaba, Basia, Milly Marston, and Sian Floyd (2003). *The Effect of HIV on Child Mortality Trends in Sub-Saharan Africa*. New York: United Nations Population Division, pp. 1–18.

Zaba, Basia et al. (2003). *HIV and Child Mortality: Evidence from Surveillance Studies in Uganda, Tanzania and Malawi*. University of North Carolina at Chapel Hill: Carolina Population Center, pp. 1–19.

Zaba, Basia et al. (2005). "HIV and Mortality of Mothers and Children: Evidence From Cohort Studies in Uganda, Tanzania, and Malawi". In: *Epidemiology* 16.3, pp. 275–280. ISSN: 1044-3983. DOI: 10.1097/01.ede.0000155507.47884.43.