Pregnancy Reporting and Biases in Under-5 Mortality in Three HDSS in Western, Eastern, and Southern Africa

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 of 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 HDSS in The Gambia, Kenya, and South Africa. Less than half of births were preregistered as pregnancies in each site. Births with preregistered pregnancies had higher observed mortality overall and especially at early ages. Cox proportional hazards models with inverse probability weights were used to investigate differences in aggregate mortality estimates, 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 preregistered births 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.

Keywords

HDSS, Pregnancy data, Under-5 mortality, Measurement error, Sub-Saharan Africa

Introduction

An estimated 2.8 million African children under age 5 died from mostly preventable causes in 2019 (UN IGME, 2020). 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., 2015). 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. 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 demographic, health, and social dynamics data in areas of sub-Saharan Africa where high-quality data is not available. Although, while 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 or 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 5, and especially within the first few weeks of life.

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 preregistered as pregnancies. We then match mother-level information on pregnancy registration to child survival, and compare U5M estimates for cohorts of births which were preregistered as pregnancies to those that were not. Mortality estimates for cohorts 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 improved ascertainment of early deaths following pregnancy registration, or selection bias from shared risk factors for pregnancy registration and U5M. To answer this question, we employ micro-level analysis to investigate aggregate differences in the observed mortality of preregistered and non-preregistered births. Our approach utilizes Cox regression models with inverse probability weighting to standardize cohorts of births with and without registered pregnancies with regard to the characteristics of the mothers, 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.

Downward Bias in HDSS Estimates of Early Childhood 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 & McGregor, 1981; Cantrelle, 1969). Certain studies collected information on pregnancies to reduce the risk of omissions (Cantrelle, 1974; Cantrelle & Leridon, 1971), though such data was incomplete, and downward bias likely persisted (Garenne, 1981; Pison & 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 (Assefa et al., 2016; Kahn et al., 2012; Kishamawe et al., 2015; Rossier et al., 2012; Sankoh & Byass, 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., 2021). 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., 2021; 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 et al., 1985; Cantrelle & Leridon, 1971; Garenne, 1981; Jasseh, 2003; Pison & Langaney, 1985). However, recent research has 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.

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 prompt the HDSS fieldworker 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 (Alabi et al., 2014; EN-INDEPTH, 2016). 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.

Data and Methods

The analysis was conducted on cohorts of births taking place over 3-4 years in HDSS sites in The Gambia, Kenya, and South Africa. 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. Selected cohorts were defined to include years where such factors were consistent throughout to the greatest extent possible. The quality of the data was also prioritized. This was assessed with a thorough data quality evaluation of the precision and completeness of reported dates of pregnancies, births, and deaths; as well as other socio-demographic variables. It was also necessary to have at least five full years of follow-up time for individuals born into the cohort, and thus births taking place less than five years before the most recent data collection round were not eligible for inclusion.

HDSS sites

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, n.d.). The prevalence of communicable diseases among the population is high. Malaria prevention measures have contributed to declining rates of infant and under-5 mortality and an increasing concentration of deaths at neonatal ages (Jasseh et al., 2015).

The current analysis included 29,447 births taking place in the site between 2011 and 2015. During this time, the HDSS collected data in interview rounds occurring three times per year (Rerimoi et al., 2019). Resident village reporters were also tasked with keeping track of demographic events on an ongoing basis, and their records were used by HDSS fieldworkers 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 preregistered, it was the practice of the HDSS to impute a pregnancy report for the day before birth.

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).

The cohort from Siaya included 28,877 births taking place between 2010 and 2014. Routine data collection rounds occurred three times per year up to 2014, and switched to twice per year thereafter. This transition did not affect the registration of pregnancies or births in the cohort; 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). 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. A parallel continuous village reporter system is also used to collect information on births and deaths, though not pregnancy reports

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).

This analysis included 4,633 births taking place in Somkhele HDSS between 2001 and 2004. 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 proxy respondent who was typically the head of the household, or the next available senior adult household member. The proxy respondent 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.

Reporting of Pregnancies and Births

We identified births that were preregistered as pregnancies within each cohort. This was straightforward for registered pregnancies that had been successfully followed-up with pregnancy outcomes. However, it was not uncommon to observe standalone 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.

Figure 2 shows the date of first observation, whether as a pregnancy or birth, for all individuals included in the cohorts. Preregistered births composed 46% of the total sample in Basse, 38% in Siaya, and 41% in Somkhele. Pregnancy registration was more likely to occur at later gestational ages, close to the date of delivery in all sites. In Siaya and Somkhele HDSS, the median pregnancy was registered approximately 2-3 months prior to birth, and around 80% of pregnancies were registered less than 4 months prior to birth. Approximately 20% of pregnancies were registered more than 6 months before birth in Basse, compared to 3-4% in the other two sites. This reporting pattern is suspicious. Given the lack of access to methods of early pregnancy detection, women would often be uncertain of their pregnancy status at this stage. That so many early pregnancies would be reported by the household proxy respondent is even more unlikely.

Though we discounted all pregnancy registrations which showed obvious signs of imputation, some may have escaped detection in Basse. Alternatively, these pregnancies may only appear to be have been registered very early due to reporting errors in the date of birth.

In the absence of a pregnancy registration, newborns are not under prospective surveillance in the HDSS until they are first enumerated in an HDSS data collection round. In Basse HDSS, approximately 50% of individuals with non-preregistered births were enumerated within four months. Given that data collection rounds were conducted every four months, this suggests that the other half were not enumerated in the first HDSS round following birth. From a data quality perspective, delays in enumeration can be particularly problematic for accurate reporting of dates of birth and death. During this time, non-preregistered births are also at high risk of omission in the case of early death.

For Siaya and Somkhele HDSS, the exact dates of enumerations were not available and were approximated from the first household interview following an individual's birth. As such, the distribution of enumerations in Figure 2 represents the best-case scenario for Siaya and Somkhele, where all births were enumerated as soon as possible. Siaya HDSS was conducting triannual data collection rounds during this time, and 78% of non-preregistered births had a household interview in the following four months. For the remainder, it is possible that there was some irregularity in the timing of household visits, or data integrity issues regarding the harmonization of household visitation records. In Somkhele HDSS, 95% of births were followed by a data collection round in the next six months. This indicates that the site's biannual rounds were comprehensive and consistent. While it is not certain that individuals were enumerated in the first round following their birth, households appear to have been regularly visited every six months except in a few cases.

Mothers with Preregistered and Non-Preregistered Births

Table 1 summarizes the characteristics of mothers with pre-registered and non-preregistered births. In Siaya HDSS, 66% of women with registered pregnancies had a household interview in the 20 weeks prior to delivery. Among those missing pregnancy registration, only 52% had a household interview during the same period. The proportion of women with a household interview in the 20 weeks prior to delivery was also higher among those with preregistered births in Somkhele. Among this group, 61% were present at the time of the household interview, and 56% were self-reporting. For women with non-preregistered births in Somkhele, only 25% were self-reporting at the interview taking place less than 20 weeks prior to delivery.

In Siaya, women with preregistered births were more likely to reside in Asembo and Gem localities, while those lacking pregnancy registration most often lived in Karemo. A household wealth index was calculated separately for each site as a weighted average of variables denoting socioeconomic status. Weights were generated from a Principal Components analysis which 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. In Basse HDSS, women in the top wealth quintile made up the largest share of those with non-preregistered births. In both Siaya and Somkhele, women with preregistered births were more likely to be wealthier.

There were small but significant differences in age for the two groups of mothers in Basse and Siaya HDSS. Those with preregistered births were slightly older in both sites. In Siaya, 83% of women with preregistered births had no education or primary education, compared to 79% of women with non-preregistered births. The ethnicity of the mother was also considered in Basse. Fula and Mandinka women made up a larger share of the group with preregistered births, while Sarahule women were more likely to have non-preregistered births.

Information on HIV status was available for mothers in the South African and Kenyan sites. It 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). 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. There was not a statistically significant difference in the HIV status of mothers with and without preregistered births. In all sites, women in formal union made up a larger share of the group with preregistered births compared to non-preregistered. The difference was especially notable in Somkhele HDSS, where 95% of those with preregistered births were in union, compared to only 77% of those with non-preregistered births.

In each site, pregnancy registration was less common for women who already had another child under the age of 18 months. The mother's migration status was also significantly associated with pregnancy registration in Basse and Siaya. Mothers were considered recent migrants if they had resided outside of the HDSS area for any part of the year preceding their delivery. Internal moves within the HDSS site were not considered migrations. In Basse and Siaya, women who had recently migrated were less likely to have registered pregnancies. There were no statistically significant differences in the sex ratio at birth of the cohorts of preregistered and non-preregistered births in any site.

Pregnancy and Birth Cohort Estimates of U5M

Figure 3 displays the mortality schedules for the children of mothers with preregistered and non-preregistered births in each site. In this figure and hereafter, children whose mothers had pregnancy registrations are referred to as the *pregnancy cohort*. Alternatively, those with reported births that were not preregistered are referred to as the *birth cohort*.

Deaths and exposure time were enumerated for the cohorts by single days of age from birth to 5 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.

During the first week of life, observed mortality of newborns with preregistered births was higher than for those with non-preregistered births in all sites. 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 difference between mortality rates of the pregnancy and birth cohort were not significant after the first week of life. This may be due to the smaller sample size of this site, however there also appear to be some data quality issues with age at death reporting. In the pregnancy cohort, the mortality rates for ages 2 to 5 months were higher than those of the late neonatal period (age two to four weeks). This reversal is unusual, given the rapid decline in mortality that typically characterizes the first weeks and months of life. The abnormal pattern may have resulted from upward rounding of ages at death taking place in the first month. Across all sites, mortality rates for the pregnancy and birth cohorts became more similar at later ages.

The mortality rates for each cohort were converted into the cumulative probabilities of dying under age 5 and shown in Figure 4. This figure also includes a naïve mortality estimate, which was calculated from all births (preregistered and non-preregistered) combined. The pregnancy cohort yielded the highest estimate of U5M in each site. In Basse, the cumulative probability of dying by age 5 for individuals whose birth was not preregistered as a pregnancy was 47 per 1,000 births (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, 0.056 − 0.062). The naïve estimate for all births was in between these two values at 52 per 1,000 (95% CI, 0.050 − 0.054).

In Siaya HDSS, the pregnancy cohort had a cumulative probability of dying by age 5 of close to 10% (0.099, 95% CI, 0.095 − 0.103). Observed mortality in the birth cohort was 92 per 1,000 (95% CI, 0.088 − 0.095), and the naïve estimate was 94 per 1,000 (95% CI, 0.092 − 0.097). In Somkhele HDSS, mortality in the pregnancy cohort was 89 per 1,000 (95% CI, 0.079 − 0.100). This was 5 deaths per 1,000 higher than the naïve estimate, and 10 deaths per 1,000 births higher than the estimate for the birth cohort. However, the difference between the mortality of pregnancy and birth cohorts at age 5 was not as statistically significant as in the other two sites, and had overlapping confidence intervals (95% CI, -0.007 − 0.028).

Regression 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 (Buchanan et al., 2014; Hernán et al., 2000). The inverse probability weights (IPWs) were used to standardize preregistered and non-preregistered births with respect to observable characteristics. Preregistered births were weighted by the inverse probability of pregnancy registration; while non-preregistered births 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 preregistered or non-preregistered) were weighted downwards, while unlikely members of either group 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 whether the birth was preregistered as a pregnancy. As explanatory variables, the models included covariates which were *a priori* hypothesized to affect U5M, or both pregnancy registration and U5M. While not mutually exclusive, covariates pertaining to the first group included the sex of the child, household wealth, mother’s HIV status, and whether the child had a sibling who was less than 18 months old at the time of birth. Covariates for the month of the child’s birth; mother’s age, education level, ethnicity (only Basse), marital status, and household locality within the HDSS (only Siaya) were thought to fall into the latter category. 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 et al., 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 individuals with preregistered and non-preregistered births, 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 mothers whose pregnancy registration status was correctly predicted. For comparison, we fit additional probit models for Siaya and Somkhele HDSS which included covariates for whether a household interview took place in the 20 weeks prior to delivery, and whether the mother was self-reporting at the interview. These additional covariates related to timing and nature of data collected on pregnancy status were considered unlikely to affect U5M (and thus inappropriate to include in the model used to generate IPWs), but potentially explicative of pregnancy registration. We compared the performance metrics for each probit model to assess which better predicted pregnancy registration.

Cox models were fit for each site and neonatal (birth - 28 days), post-neonatal (28 days - 1 year), child (1 year to 5 years), and under-5 (birth - 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. As a sensitivity analysis, we fit additional unweighted and IP weighted models that adjusted for all covariates included in the probit models. The latter of these is similar to a doubly robust estimator; robust to misspecification in the weight model or Cox model, but not both (Funk et al., 2011). The proportional hazards assumption of preregistered and non-preregistered births was verified graphically for unweighted and IP weighted samples.

All statistical analysis was performed using R version 3.6.1.

Results

Table 2 displays the results of the probit models regressing pregnancy registration on measureable characteristics of the mother, child, and household. The table reports the estimated marginal effect of each covariate on the probability of pregnancy registration. In each site, mothers who already had young children were less likely to register pregnancies by 10-17%. 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. The month of birth was a significant covariate in each site, but the marginal effects were largest for Somkhele. This is perhaps due to the site conducting biannual as opposed to triannual data collection rounds, with a longer inter-round interval. Pregnancies coming to term during this time would be especially unlikely to be registered.

While the covariate for age of the mother was significantly associated with pregnancy registration in each site, the marginal effect of this variable was very small. Having higher or an unknown education level increased the probability of pregnancy registration in Basse, and reduced it in Siaya. The variable was not highly 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 was associated with a reduced probability of pregnancy registration, perhaps indicating that such individuals were less likely to have complete information in the HDSS in general. Marital status had the strongest effect on pregnancy registration of all variables in Somkhele. Unpartnered women and women of unknown marital status were 30 and 45% less likely to have registered pregnancies than women in union, respectively. The same relationship was found in the other sites, but with smaller effects. Being a recent migrant significantly decreased the probability of pregnancy registration in Basse and Siaya.

Despite the presence of several significant covariates, the goodness of fit 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 slightly higher at 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%; a naïve classifier which assigns all observations to the larger group of non-preregistered births. There was a difference of 2 percentage points in accuracy and the NIR for the Siaya model (accuracy = 0.64, NIR = 0.62), and 4 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. This indicated low potential for selection bias in pregnancy cohort estimates of mortality from observable covariate imbalances between preregistered and non-preregistered births.

Additional probit models were fit for Siaya and Somkhele HDSS which included more variables related to data collection (see Table A1). The model for Siaya included an indicator for whether a household interview took place less than 20 weeks prior to delivery. This model had slightly higher goodness of fit (pseudo-R^2 = 0.07) and predictive accuracy (accuracy = 0.66). For Somkhele, the household interview covariate also included information for whether the mother self-reported her pregnancy status. This resulted in a larger improvement in predictive performance (pseudo-R^2 = 0.13, accuracy = 0.69). Women who were interviewed directly at a household interview taking place less than 20 weeks prior to delivery were 27% more likely to have a registered pregnancy.

The issue of selectivity in pregnancy registration was investigated further with Cox 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 Figure A1). Results for Cox models regressing child survival on pregnancy registration are displayed in Figure 5.

In the unweighted baseline model for Basse HDSS, newborns whose births were preregistered had almost double the observed risk of neonatal mortality (hazard ratio [HR] 1.94; 95% CI 1.61 $-$ 2.33) compared to non-preregistered births. After standardizing for covariate imbalances between mothers with preregistered and non-preregistered births using the IPWs, the hazard ratio decreased slightly (HR 1.92; robust 95% CI 1.58 $-$ 2.33). This indicated that the relationship between pregnancy registration and neonatal mortality was subject to some positive confounding in Basse HDSS. Adjustment from IPWs also reduced the hazard ratio for U5M. In the baseline model for U5M, the risk for those with preregistered births was measured as 28% higher than those missing pregnancy registrations (HR 1.28; 95% CI 1.16 − 1.42). This decreased to 26% (HR 1.26; robust 95% CI 1.13 − 1.40) after adjustment from IPWs.

Similar to Basse, observed risk of dying in the neonatal period was significantly higher for newborns with preregistered births in Siaya HDSS. In the baseline model, newborns with preregistered births were 1.71 times more likely to die in neonatal ages (95% CI 1.41 − 2.07) compared to those with non-preregistered births. However, unlike Basse, adjustment from IPWs led to an increased hazard ratio of 1.78 (robust 95% CI 1.45 − 2.17). This suggests that having a registered pregnancy in Siaya was associated with characteristics that were protective of mortality. The elevated hazard for neonates with preregistered births was not reflective of higher innate risk, but rather improved ascertainment of neonatal deaths following pregnancy registration. As such, after controlling for characteristics associated with pregnancy registration and mortality, the estimated hazard ratio for pregnancy registration increased even further.

Total risk for U5M among children with preregistered births was also higher than that of children with non-preregistered births in Siaya. In the baseline model, children with preregistered births had 1.10 times the risk of mortality under 5 (95% CI, 1.02 − 1.20). The hazard ratio increased and became more significant in the model adjusted with IPWs (HR 1.13; robust 95% CI, 1.03 − 1.23). In both Basse and Siaya, the hazard ratio for pregnancy registration was not highly significant in post-neonatal or child age groups.

In Somkhele HDSS, the application of IPWs tended to increase the hazard ratio of pregnancy registration as it did in Siaya. However, Somkhele differed from both sites in having important differences in the observed post-neonatal mortality of preregistered and non-preregistered births. Compared to non-preregistered births, newborns with births that were preregistered as pregnancies had approximately 50% higher risk for post-neonatal mortality in the IP weighted model (HR 1.51; robust 95% CI 1.12 − 2.05). This was an increase from the baseline estimate of 1.32 (95% CI 1.00 − 1.73). In the baseline model for U5M, mortality risk for those with preregistered births was not significantly different from those without a pregnancy registration. After adjustment from IPWs, the observed U5M of preregistered births was 27% higher than non-preregistered births, with moderate significance (robust 95% CI 1.01 − 1.59).

For all sites, the hazard ratios of IP weighted models were similar to those of the other Cox models included in the sensitivity analysis (i.e. unweighted, covariate adjusted models; and double robust models adjusted with both weights and covariates). Full results are provided in Table A2.

Discussion

Using data from three African HDSS, we found that estimates of U5M for children whose births were preregistered as a pregnancy were higher than for non-preregistered births. This difference in observed mortality was primarily attributable to early (neonatal and post-neonatal) 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 post-neonatal 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. Overall, pregnancy registration seemed to result in improved ascertainment of early deaths. For all sites, mortality of preregistered and non-preregistered births was not statistically different between the ages of 1 and 5 years.

The proportion of preregistered births made up a minority of each sample, and was relatively similar across sites (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 pregnancy registration. Indeed, 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 theoretically result in around 75% of births being preregistered. That this level was not achieved in either site indicates considerable underreporting. Furthermore, it suggests that marginally increasing the frequency of interview rounds alone would not be sufficient to achieve exhaustive pregnancy reporting.

In all sites, most pregnancies were registered less than 4 months prior to delivery, and the frequency of pregnancy registration generally increased with gestational age. Pregnancies were more likely to be registered if there was a household interview shortly before delivery, however, factors related to proxy reporting also played a large role. In Somkhele HDSS, women were more likely to have registered pregnancies if they were present during the household visit of the HDSS team, and self-reported their own pregnancy status. Across all three sites, women in union were more likely to have registered pregnancies. Underreporting of pregnancies for women outside of formal union could be related to strong social norms against having a child outside of wedlock. Though 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. While women in rural sub-

Saharan Africa often conceal their pregnancy from the broader community for as long as possible, disclosing one's status to a sexual partner is often 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 underreporting of all pregnancies, but especially those of women who are not in formal union.

Given the incompleteness of pregnancy reporting in the HDSS, it was important to examine whether the higher mortality of pregnancy cohorts was attributable to such mothers being negatively selected in terms of U5M risk factors. Probit models for pregnancy registration had low goodness of fit. Mothers with preregistered and non-preregistered births did not differ substantially in terms of measured characteristics, which were not strongly predictive of pregnancy registration. This finding indicated low potential for a confounding effect between pregnancy registration and child survival. Nevertheless, marginal structural Cox models with inverse probability weights were used to estimate the effect of pregnancy registration on observed mortality, controlling for which women were more or less likely to register pregnancies.

Observed mortality in the pregnancy cohort remained higher after standardizing preregistered and non-preregistered births with respect to measured characteristics of the mother, child, and household. In the case of Siaya and Somkhele HDSS, the hazard ratios increased or became more significant. In Basse, there was a small reduction in the baseline effect of pregnancy registration on neonatal and under-5 mortality after adjustment, though it remained significant. 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.

This study draws on the richness of HDSS data to investigate macro-level differences in the mortality of preregistered and non-preregistered births, 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. There is thus the possibility for residual confounding of the association between pregnancy registration and U5M by omitted or insufficiently controlled for factors. Clinical markers are usually a better predictor of neonatal mortality and were not available for inclusion. However, for these to explain the higher estimates of mortality in pregnancy cohorts, they would also have to be strongly associated with pregnancy registration. There is not a strong *a priori* reason to presume that this would be the case, especially in light of our finding that pregnancy registration was better predicted by the timing of data collection and usage of a proxy respondent than mother-level characteristics.

Additionally, 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 (Helleringer et al., 2020; Liu et al., 2016). 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 under-5 mortality were still higher than birth cohort estimates for all sites (Figure A2). It is thus not likely that misclassification of stillbirths could fully account for the higher observed mortality in the pregnancy cohort group.

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 longstanding 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.

As a first step, we recommend that HDSS protocols and procedures for collecting data on pregnancies be extensively documented. These details are often not included in site descriptions and publicly available data resource profiles, though they provide necessary context for the mortality and fertility estimates arising from such data. In terms of changes to data collection itself, conducting more frequent interview rounds is an appealing prospect, though not always financially feasible. Additionally, while the frequency of data collection rounds is an important determinant of pregnancy reporting completeness, it is not the only one. Data collection procedures and protocols exert an enormous influence on pregnancy registration. Sites should strive to pose questions about pregnancy status directly and avoid proxy reports. In the case that direct reports are collected, sites could achieve further gains by using female interviewers, and training interviewers to collect data in a sensitive manner. Record linkage with routine program data is another promising avenue to enhance HDSS reporting. In areas where coverage of antenatal care services is high, sites could leverage such data to ascertain pregnancy status information for residents. Of course, these efforts would all incur financial and logistical challenges of their own. However, given the many benefits that pregnancy registration confers to monitoring maternal and newborn health in HDSS, we argue that these steps would be a worthwhile investment. Such data will be essential for guiding evidence-based public health interventions and accelerating progress towards mortality reduction.

References

AbouZahr, C., de Savigny, D., Mikkelsen, L., Setel, P. W., Lozano, R., & Lopez, A. D. (2015). Towards universal civil registration and vital statistics systems: The time is now. *The Lancet*, *386*(10001), 1407–1418. https://doi.org/10.1016/S0140-6736(15)60170-2

AHRI. (2021). AHRI | About Us | Working towards the elimination of HIV and TB disease. In Africa Health Research Institute. https://www.ahri.org/about/.

Akuze, J., Blencowe, H., Waiswa, P., Baschieri, A., Gordeev, V. S., Kwesiga, D., … Serbanescu, F. (2020). Randomised comparison of two household survey modules for measuring stillbirths and neonatal deaths in five countries: The Every Newborn-INDEPTH study. *The Lancet Global Health*, *8*(4), e555–e566. https://doi.org/10.1016/S2214-109X(20)30044-9

Alabi, O., Doctor, H. V., Afenyadu, G. Y., & Findley, S. E. (2014). Lessons learned from setting up the Nahuche Health and Demographic Surveillance System in the resourceconstrained context of northern Nigeria. *Global Health Action*, *7*. https://doi.org/10.3402/gha.v7.23368

Assefa, N., Oljira, L., Baraki, N., Demena, M., Zelalem, D., Ashenafi, W., & Dedefo, M. (2016). HDSS Profile: The Kersa Health and Demographic Surveillance System. *International Journal of Epidemiology*, *45*(1), 94–101. https://doi.org/10.1093/ije/dyv284

Billewicz, W. Z., & McGregor, I. A. (1981). The Demography of Two West African (Gambian) Villages, 1951. *Journal of Biosocial Science*, *13*(2), 219–240. https://doi.org/10.1017/S0021932000013390

Blacker, J., Hill, A., & Timaeus, I. (1985). Age patterns of mortality in Africa: An examination of recent evidence. *International Union for the Scientific Study of Population, International Population Conference*.

Buchanan, A. L., Hudgens, M. G., Cole, S. R., Lau, B., & Adimora, A. A. (2014). Worth the Weight: Using Inverse Probability Weighted Cox Models in AIDS Research. *AIDS Research and Human Retroviruses*, *30*(12), 1170–1177. https://doi.org/10.1089/aid.2014.0037

Cantrelle, P. (1969). Etude démographique dans la région du Sine-Saloum (Sénégal) : État civil et observation démographique 1963-1965.

Cantrelle, P. (1974). Is there a standard pattern of tropical mortality? International Union for the Scientific Study of Population (IUSSP), Population in African Development.

Cantrelle, P., & Leridon, H. (1971). Breast feeding, mortality in childhood and fertility in a rural zone of Senegal. *Population Studies*, *25*(3), 505–533. https://doi.org/10.1080/00324728.1971.10405821

Eilerts, H., Romero Prieto, J., Eaton, J. W., & Reniers, G. (2021). 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. *Demographic Research*, *44*, 415–442. https://doi.org/10.4054/DemRes.2021.44.18

EN-INDEPTH. (2016, June). Workshop Report. *Every Newborn Action Plan Metrics*.

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly Robust Estimation of Causal Effects. *American Journal of Epidemiology*, *173*(7), 761–767. https://doi.org/10.1093/aje/kwq439

Garenne, M. (1981). The age pattern of infant and child mortality in Ngayokheme (rural West Africa). *University of Pennsylvania Scholarly Commons*, *African Demography Working Paper Series*, 42.

Gareta, D., Baisley, K., Mngomezulu, T., Smit, T., Khoza, T., Nxumalo, S., … Herbst, K. (2021). Cohort Profile Update: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. *International Journal of Epidemiology*, *50*(1), 33–34. https://doi.org/10.1093/ije/dyaa264

Hamel, M. J., Adazu, K., Obor, D., Sewe, M., Vulule, J., Williamson, J.M., … Laserson, K. F. (2011). A Reversal in Reductions of Child Mortality in Western Kenya, 2003. *The American Journal of Tropical Medicine and Hygiene, 85*(4), 597–605. https://doi.org/10.4269/ajtmh.2011.10-0678

Harling, G., Perkins, J. M., Gómez-Olivé, F. X., Morris, K., Wagner, R. G., Montana, L., … Berkman, L. (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. *Field Methods*, *30*(2), 140–154. https://doi.org/10.1177/1525822X18769498

Haws, R. A., Mashasi, I., Mrisho, M., Schellenberg, J. A., Darmstadt, G. L., & Winch, P. J. (2010). "These are not good things for other people to know": How rural Tanzanian women's experiences of pregnancy loss and early neonatal death may impact survey data quality. *Social Science & Medicine*, *71*(10), 1764–1772. https://doi.org/10.1016/j.socscimed.2010.03.051

Helleringer, S., Liu, L., Chu, Y., Rodrigues, A., & Fisker, A. B. (2020). Biases in Survey Estimates of Neonatal Mortality: Results From a Validation Study in Urban Areas of Guinea-Bissau. *Demography*. https://doi.org/10.1007/s13524-020-00911-6

Hernán, M. Á., Brumback, B., & Robins, J. M. (2000). Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men: *Epidemiology*, *11*(5), 561–570. https://doi.org/10.1097/00001648-200009000-00012

Jasseh, M. (2003). *Age Patterns of Mortality within Childhood in Sub-Saharan Africa* [{{PhD}} Thesis]. London School of Hygiene & Tropical Medicine.

Jasseh, M., Gomez, P., Greenwood, B. M., Howie, S. R., Scott, S., Snell, P.C., … D'Alessandro, U. (2015). Health & Demographic Surveillance System Profile: Farafenni Health and Demographic Surveillance System in The Gambia. *International Journal of Epidemiology*, *44*(3), 837–847. https://doi.org/10.1093/ije/dyv049

Johnson, K., Grant, M., Khan, S., Moore, Z., Armstrong, A., & Sa, Z. (2009). *Fieldwork-Related Factors and Data Quality in the Demographic and Health Surveys Program* (No. 19). ICF Macro.

Kadobera, D., Waiswa, P., Peterson, S., Blencowe, H., Lawn, J., Kerber, K., & Tumwesigye, N. M. (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. *Global Health Action*, *10*(1), 1356641. https://doi.org/10.1080/16549716.2017.1356641

Kahn, K., Collinson, M. A., Gomez-Olive, F. X., Mokoena, O., Twine, R., Mee, P., … Tollman, S. M. (2012). Profile: Agincourt Health and Socio-demographic Surveillance System. *International Journal of Epidemiology*, *41*(4), 988–1001. https://doi.org/10.1093/ije/dys115

Kishamawe, C., Isingo, R., Mtenga, B., Zaba, B., Todd, J., Clark, B., … Urassa, M. (2015). Health & Demographic Surveillance System Profile: The Magu Health and Demographic Surveillance System (Magu HDSS). *International Journal of Epidemiology*, *44*(6), 1851–1861. https://doi.org/10.1093/ije/dyv188

Kwesiga, D., Tawiah, C., Imam, M. A., Tesega, A. K., Nareeba, T., Enuameh, Y. A. K., … The Every Newborn-INDEPTH Study Collaborative Group. (2021). Barriers and enablers to reporting pregnancy and adverse pregnancy outcomes in population-based surveys: EN-INDEPTH study. *Population Health Metrics*, *19*(1), 15. https://doi.org/10.1186/s12963-020-00228-x

Kwon, C., Naser, A. M., Eilerts, H., Reniers, G., & Cunningham, S. A. (2021). *Pregnancy Surveillance Methods within Health and Demographic Surveillance Systems* (No. 5:144). Gates Open Research. https://doi.org/10.12688/gatesopenres.13332.1

Liu, L., Kalter, H. D., Chu, Y., Kazmi, N., Koffi, A. K., Amouzou, A., … Black, R. E. (2016). Understanding Misclassification between Neonatal Deaths and Stillbirths: Empirical Evidence from Malawi. *PLOS ONE*, *11*(12), e0168743. https://doi.org/10.1371/journal.pone.0168743

Mcfadden, D. (1977). *Quantitative methods for analyzing travel behaviour of individuals: Some recent developments* (Discussion {{Paper}} No. 474; p. 35). Cowles Foundation for Research in Economics.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (P. Zarembka (ed.), pp. 105–142). Academic Press.

Mikkelsen, L., Phillips, D. E., AbouZahr, C., Setel, P. W., de Savigny, D., Lozano, R., & Lopez, A. D. (2015). A global assessment of civil registration and vital statistics systems: Monitoring data quality and progress. *The Lancet*, *386*(10001), 1395–1406. https://doi.org/10.1016/S0140-6736(15)60171-4

Nareeba, T., Dzabeng, F., Alam, N., Biks, G. A., Thysen, S. M., Akuze, J., … The Every Newborn-INDEPTH Study Collaborative Group. (2021). Neonatal and child mortality data in retrospective population-based surveys compared with prospective demographic surveillance: EN-INDEPTH study. *Population Health Metrics*, *19*(1), 7. https://doi.org/10.1186/s12963-020-00232-1

Ngom, P. (2001). Demographic surveillance and health equity in sub-Saharan Africa. *Health Policy and Planning*, *16*(4), 337–344. https://doi.org/10.1093/heapol/16.4.337

Odhiambo, F. O., Laserson, K. F., Sewe, M., Hamel, M. J., Feikin, D. R., Adazu, K., … Vulule, J. M. (2012). Profile: The KEMRI/CDC Health and Demographic Surveillance System–Western Kenya. *International Journal of Epidemiology*, *41*(4), 977–987. https://doi.org/10.1093/ije/dys108

Pison, G., & Langaney, A. (1985). The Level and Age Pattern of Mortality in Bandafassi (Eastern Senegal): Results From a Small-Scale and Intensive Multi-Round Survey. *Population Studies*, *39*(3), 387–405. https://doi.org/10.1080/0032472031000141586

Rerimoi, A. J., Jasseh, M., Agbla, S. C., Reniers, G., Roca, A., & Timæus, I. M. (2019). Under-five mortality in The Gambia: Comparison of the results of the first demographic and health survey with those from existing inquiries. *PLOS ONE*, *14*(7), e0219919. https://doi.org/10.1371/journal.pone.0219919

Robins, J. M., Hernán, M. Á., & Brumback, B. (2000). Marginal Structural Models and Causal Inference in Epidemiology: *Epidemiology*, *11*(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

Rossier, C., Soura, A., Baya, B., Compaore, G., Dabire, B., Dos Santos, … Zourkaleini, Y. (2012). Profile: The Ouagadougou Health and Demographic Surveillance System. *International Journal of Epidemiology*, *41*(3), 658–666. https://doi.org/10.1093/ije/dys090

Sankoh, O., & Byass, P. (2012). The INDEPTH Network: Filling vital gaps in global epidemiology. *International Journal of Epidemiology*, *41*(3), 579–588. https://doi.org/10.1093/ije/dys081

Tanser, F., Hosegood, V., Barnighausen, T., Herbst, K., Nyirenda, M., Muhwava, W., … Newell, M.-L. (2008). Cohort Profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. *International Journal of Epidemiology*, *37*(5), 956–962. https://doi.org/10.1093/ije/dym211

UKRI. (n.d.). Basse Health and Demographic Surveillance System. In *UK Research and Innovation*. https://gtr.ukri.org/projects?ref=MC_UP_1003%2F1.

UN IGME. (2020). Levels & Trends in Child Mortality: Report 2020, Estimates developed by the United Nations Inter-agency Group for Child Mortality Estimation. United Nations Inter-agency Group for Child Mortality Estimation, United Nations Children's Fund.

UNAIDS. (2020). *Gambia Country Factsheet*. https://www.unaids.org/en/regionscountries/countries/gambia.

Utazi, C. E., Sahu, S. K., Atkinson, P. M., Tejedor-Garavito, N., Lloyd, C. T., & Tatem, A. J. (2018). Geographic coverage of demographic surveillance systems for characterising the drivers of childhood mortality in sub-Saharan Africa. *BMJ Global Health*, *3*(2), e000611. https://doi.org/10.1136/bmjgh-2017-000611

Verhulst, A., Romero Prieto, J., Alam, N., Katz, J., Lankoande, B., Liu, L., … Guillot, M. (2021). Identifying New Age Patterns of Under-Five Mortality Using the Under-Five Mortality Database. Methodological Innovations in Mortality, 6.

Waiswa, P., Moyer, C., Kwesiga, D., Arthur, S., Sankoh, O., Welaga, P., … Kerber, K. (2019). Status of birth and pregnancy outcome capture in Health Demographic Surveillance Sites in 13 countries. *International Journal of Public Health*, *64*(6), 909–920. https://doi.org/10.1007/s00038-019-01241-0

Xu, S., Ross, C., Raebel, M. A., Shetterly, S., Blanchette, C., & Smith, D. (2010). Use of Stabilized Inverse Propensity Scores as Weights to Directly Estimate Relative Risk and Its Confidence Intervals. *Value in Health*, *13*(2), 273–277. https://doi.org/10.1111/j.1524-4733.2009.00671.x

Tables

Table 1: Registered pregnancies in each HDSS site, by observable characteristics of the mother, child, and household.

		Basse HDSS [2011, 2015)			Siaya HDSS [2010, 2014)			Somkhele HDSS [2001, 2004)		
Household Interview <20wks Prior to Delivery					0.66	0.52	<0.01	0.89	0.76	<0.01
Mother is Present	No							0.28	0.47	
	Yes							0.61	0.29	<0.01
	Not Applicable							0.11	0.24	
Mother is Self-Respondent	No							0.33	0.51	
	Yes							0.56	0.25	<0.01
	Not Applicable							0.11	0.24	
Household Locality	Asembo				0.34	0.25				
	Gem				0.35	0.36	<0.01			
	Karemo				0.31	0.39				
Household Wealth Quintile	1	0.16	0.18		0.15	0.14		0.20	0.19	
	2	0.16	0.16		0.16	0.15		0.21	0.18	
	3	0.10	0.10	<0.01	0.15	0.15	<0.01	0.19	0.19	<0.01
	4	0.17	0.14		0.17	0.16		0.18	0.18	
	5	0.14	0.13		0.12	0.15		0.18	0.21	
	Unknown	0.26	0.29		0.24	0.25		0.04	0.05	
Mother Age		27.3	26.1	<0.01	25.9	25.1	<0.01	25.3	25.5	0.42
Mother Education Level	None/Primary	0.01	0.02		0.83	0.79		0.32	0.29	
	Secondary/Reli	0.30	0.31	<0.01	0.16	0.20	<0.01	0.67	0.67	0.20
	Unknown	0.68	0.67		0.00	0.01		0.02	0.03	
Mother Ethnicity	Fula	0.34	0.31							
	Mandinka	0.23	0.20	<0.01						
	Sarahule	0.41	0.46							
	Other	0.02	0.02							
Mother HIV Status	Negative				0.27	0.25		0.47	0.41	
	Positive				0.04	0.03	0.15	0.01	0.01	0.84
	Unknown				0.69	0.72		0.51	0.57	
Mother Martial Status	In union	0.52	0.38		0.79	0.64		0.95	0.77	
	Not in union[a]			<0.01	0.15	0.25	<0.01	0.05	0.16	<0.01
	Unknown	0.48	0.62		0.06	0.11		0.00	0.07	
Mother Has Child <18 Months Old		0.02	0.03	<0.01	0.04	0.05	<0.01	0.02	0.04	<0.01
Mother is Recent Migrant		0.06	0.16	<0.01	0.14	0.28	<0.01	0.26	0.25	0.90
Sex Ratio at Birth		1.00	1.04	0.19	1.00	1.01	0.69	0.99	1.05	0.30
N		0.46	0.54		0.38	0.62		0.41	0.59	

Notes: Categorical variables are presented as weighted proportions and continuous variables as means. All measures are categorical except Mother Age (ranges from 10 to 49) and Sex Ratio at Birth. Two-sample t tests were performed for continuous outcomes, and chi-square tests were performed for categorical outcomes. Tests were performed on complete case data, with "Unknown" values excluded.

[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.

Table 2: Results from probit regression analysis. Table shows marginal effects.

		Basse [2011, 2015)	Siaya [2010, 2014)	Somkhele [2001, 2004)
Child Sex (ref.=Female)	Male	-0.01	-0.00	-0.02
Has Child <18 Months Old (ref. = No)	Yes	-0.17***	-0.10***	-0.16***
Household Locality (ref.=Karemo)	Asembo		0.15***	
	Gem		0.01	
Household Wealth Quintile (ref.=1)	2	0.01	-0.00	0.02
	3	0.01	-0.02†	-0.01
	4	0.03**	-0.01	-0.01
	5	0.01	-0.04***	-0.01
	Unknown	-0.01	-0.03**	-0.01
Month of Birth (ref. = 1)	2	-0.01	-0.03*	0.01
	3	0.02†	0.06***	0.10**
	4	0.02	0.09***	0.14***
	5	0.03†	0.04**	0.11***
	6	-0.02	0.02	0.09**
	7	-0.00	0.03*	0.11***
	8	0.00	0.08***	0.10**
	9	0.04**	0.05***	0.19***
	10	0.07***	0.03*	0.12***
	11	0.05***	0.05***	0.07†
	12	0.07***	0.08***	0.04
Mother Age		0.00***	-0.00***	-0.01***
Mother Education (ref.=None/Primary)	Secondary/Religious	0.07**	-0.04***	-0.03†
	Unknown	0.09***	-0.15***	-0.02
Mother Ethnicity (ref.=Sarahule)	Fula	0.05***		
	Mandinka	0.06***		
	Other	0.00		
Mother HIV Status (ref.=Negative)	Positive		0.02	-0.07
	Unknown		-0.03***	-0.05***
Mother Marital Status (ref.=In union)	Not in union		-0.17***	-0.30***
	Unknown	-0.09***	-0.19***	-0.45***
Mother Recently Migrated (ref.=No)	Yes	-0.22***	-0.17***	0.03*
Observations		29447	28877	4633
No-Information Rate		0.54	0.62	0.59
Accuracy		0.59	0.64	0.63
Pseudo R-squared		0.04	0.06	0.08

Notes: †p<0.10; *p<0.05; **p<0.01; ***p<0.001

Figures

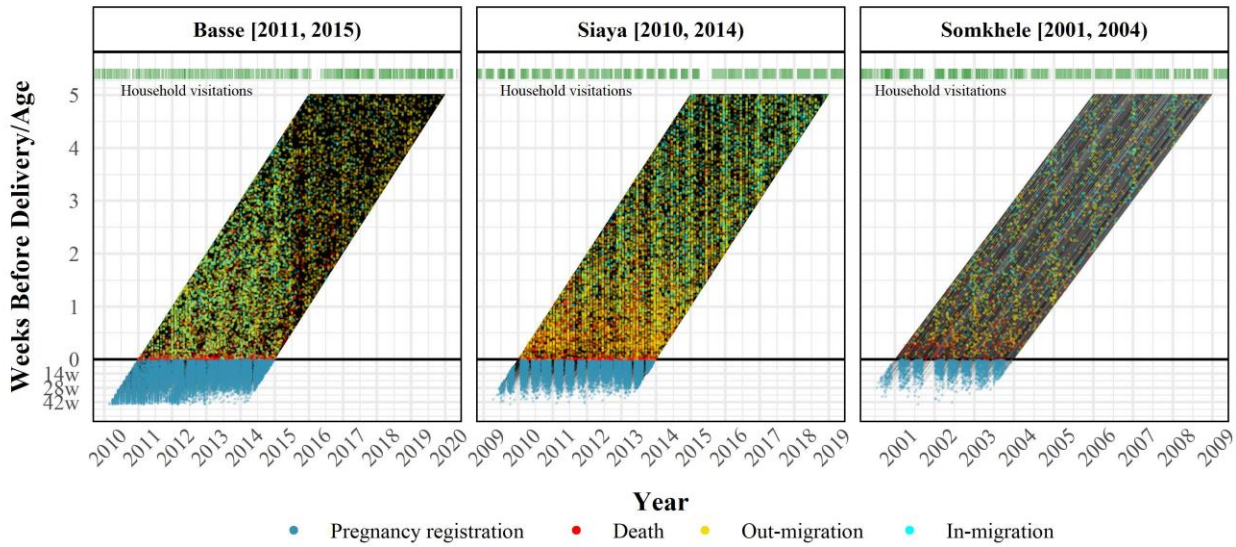

Figure 1: Lexis diagrams depicting the cohorts included in the analysis from each HDSS site. The dates of data collection through regular household visitations are plotted in bars across the top of each panel, with color intensity denoting more or less data collected during the same day of fieldwork. The dates of data entry in Basse were used to infer household visitations, which were not included in the dataset.

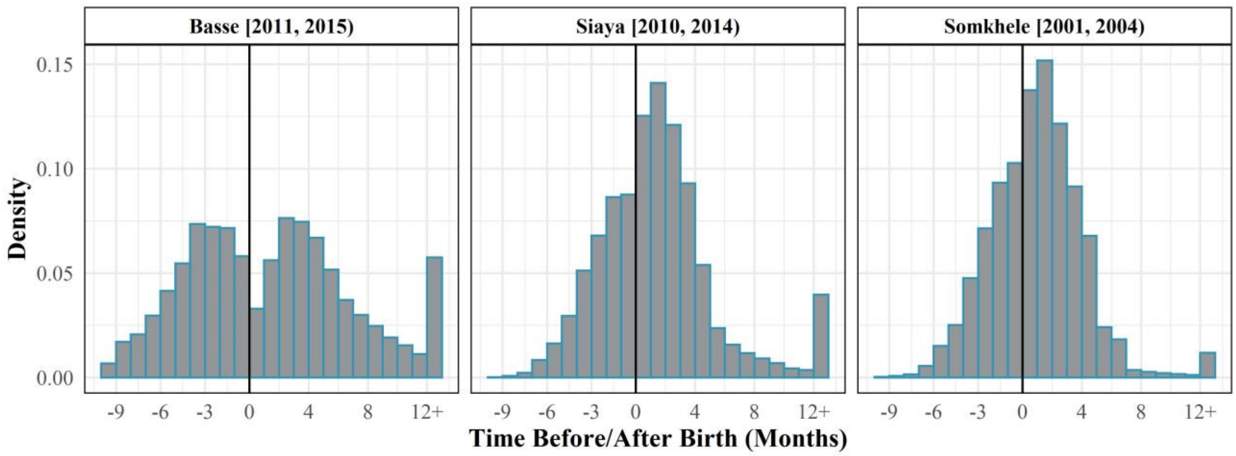

Figure 2: Relative distribution of when individuals were first reported (either as a pregnancy or birth) in each site. Preregistered births are plotted as negative values, showing when the pregnancy registration took place prior to the birth. For non-preregistered births, the duration between birth and enumeration is shown. In Siaya and Somkhele HDSS, the date of enumeration was approximated from the first household visitation following birth.

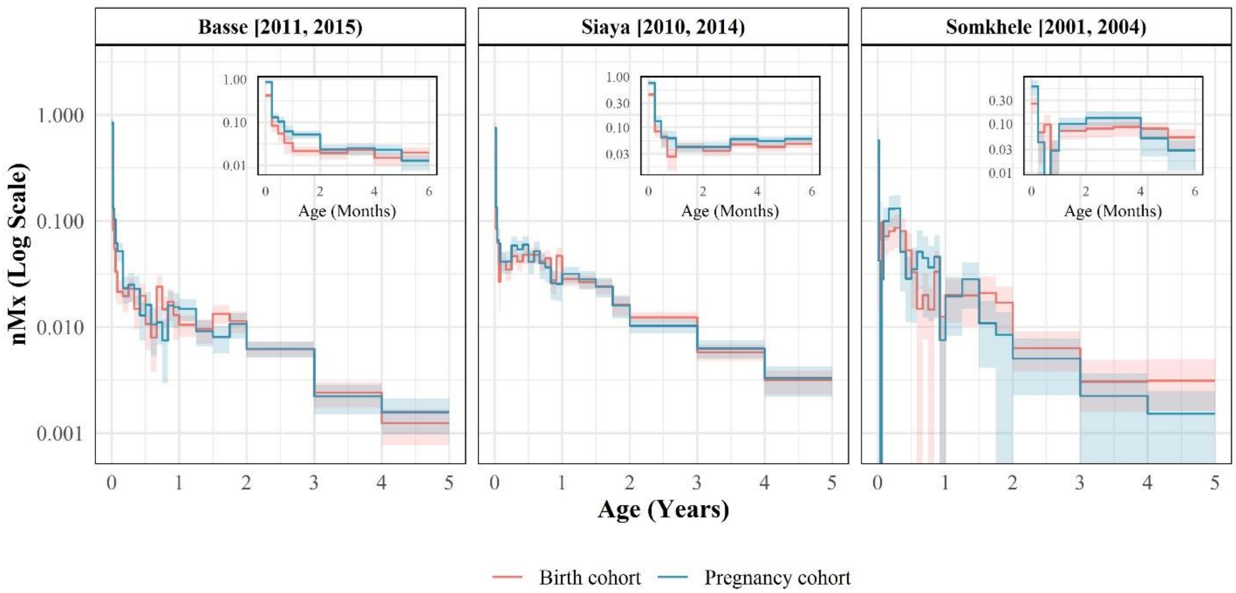

Figure 3: Age-specific mortality rates under age 5 years - birth cohort vs. pregnancy cohort. Inset panels show mortality under 6 months. Bootstrap 95% confidence intervals were calculated for 5,000 resamples with replacement.

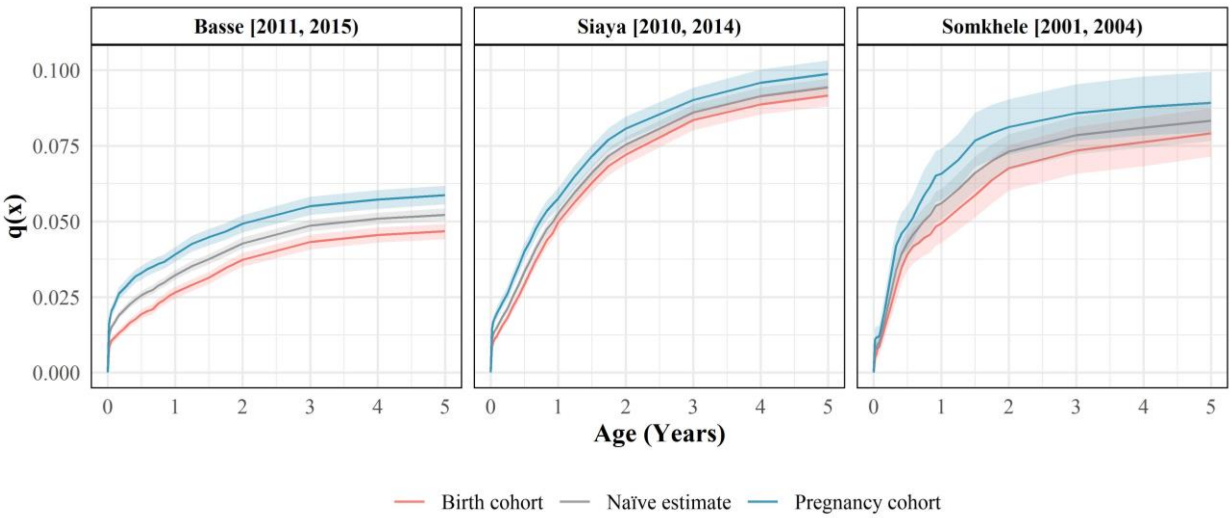

Figure 4: Cumulative probabilities of dying by age 5 years - birth cohort, pregnancy cohort, and naïve estimate. Bootstrap 95% confidence intervals were calculated for 5,000 resamples with replacement.

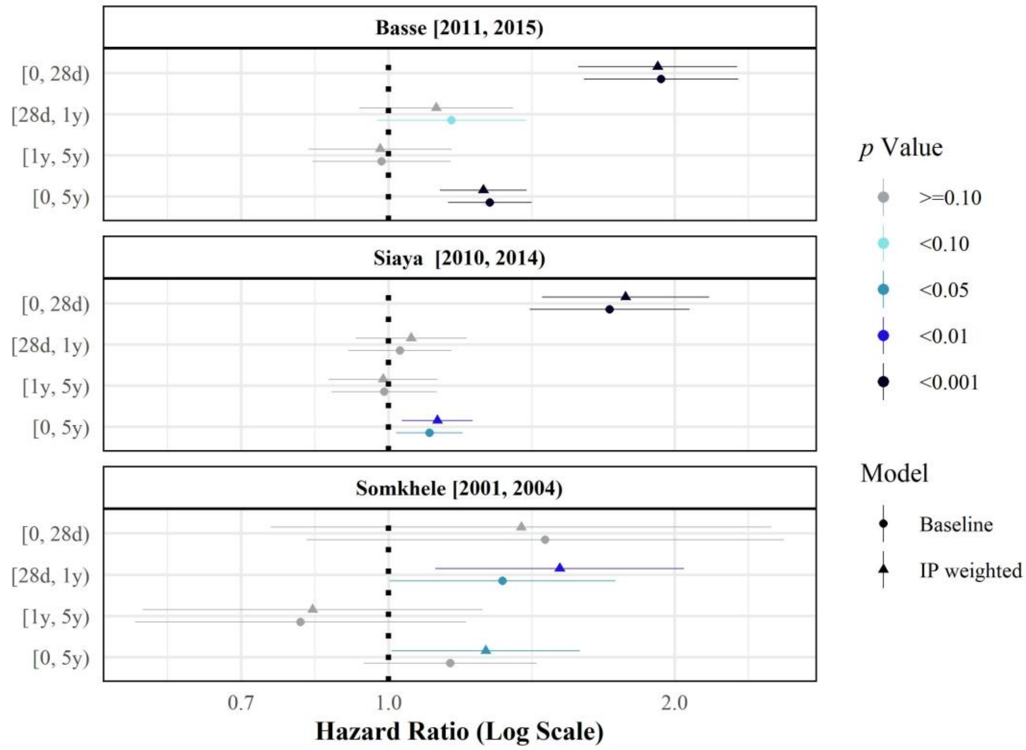

Figure 5: Results from Cox regression models for mortality under age 5. Estimated hazard ratios for births with a prior pregnancy registration are displayed as dots with horizontal lines for 95% confidence intervals. Inverse probability weighted (IP weighted) estimates have robust 95% confidence intervals.

Appendix

Table A1: Results from probit regression analysis incorporating additional covariates related to the timing of household interview, presence of the mother, and identity of the respondent. Table shows marginal effects.

		Siaya [2010, 2014)	Somkhele [2001, 2004)
Child Sex (ref.=Female)	Male	-0.00	-0.01
Has Child <18 Months Old (ref. = No)	Yes	-0.10***	-0.17***
Household Interview <20wks Prior to Delivery (ref.=No)	Yes	0.14***	
	Yes, Self-Respondent		0.27***
	Yes, Proxy Respondent		0.00
Household Locality (ref.=Karemo)	Asembo	0.16***	
	Gem	0.02**	
Household Wealth Quintile (ref.=1)	2	-0.00	0.02
	3	-0.01	-0.01
	4	-0.00	-0.01
	5	-0.05***	-0.01
	Unknown	-0.04***	-0.02
Month of Birth (ref. = 1)	2	-0.03†	0.03
	3	0.06***	0.09**
	4	0.08***	0.13***
	5	0.03*	0.09**
	6	0.01	0.06†
	7	0.02†	0.09**
	8	0.07***	0.10**
	9	0.04***	0.16***
	10	0.03*	0.10**
	11	0.04***	0.05
	12	0.07***	0.05
Mother Age		-0.00***	-0.01***
Mother Education (ref.=None/Primary)	Secondary/Religious	-0.04***	-0.02
	Unknown	-0.14***	-0.02
Mother HIV Status (ref.=Negative)	Positive	0.02	-0.06
	Unknown	-0.04***	-0.04**
Mother Marital Status (ref.=In union)	Not in union	-0.17***	-0.26***
	Unknown	-0.19***	-0.43***
Mother Recently Migrated (ref.=No)	Yes	-0.17***	0.03*
Observations		28877	4633
No-Information Rate		0.62	0.59
Accuracy		0.66	0.69
Pseudo R-squared		0.07	0.13

Notes: †p<0.10; *p<0.05; **p<0.01; ***p<0.001

Table A2: Results from Cox regression analysis. Hazard ratios and 95% confidence intervals for the effect of pregnancy registration on mortality under age 5.

		Registered Pregnancies		Non-preregistered Births					
HDSS	Age Group	Deaths	Person-years	Deaths	Person-years	Unweighted HR (95% CI)	Unweighted HR (95% CI) Adjusted for covariates	IP weighted HR (robust 95% CI)	IP weighted HR (robust 95% CI) Adjusted for covariates
Basse	[0,28d)	288	1012.09	178	1214.42	1.94*** (1.61, 2.33)	1.90*** (1.57, 2.30)	1.92*** (1.58, 2.33)	1.92*** (1.58, 2.33)
	[28d, 1y)	233	11813.77	241	14231.39	1.16† (0.97, 1.39)	1.12 (0.93, 1.34)	1.12 (0.93, 1.35)	1.13 (0.93, 1.36)
	[1y, 5y)	251	47847.24	302	56268.24	0.98 (0.83, 1.16)	0.97 (0.82, 1.15)	0.98 (0.82, 1.17)	0.98 (0.82, 1.16)
	[0,5y)	772	60673.10	721	71714.03	1.28*** (1.16, 1.42)	1.25*** (1.13, 1.39)	1.26*** (1.13, 1.40)	1.26*** (1.13, 1.40)
Siaya	[0,28d)	210	827.49	201	1355.79	1.71*** (1.41, 2.07)	1.76*** (1.44, 2.15)	1.78*** (1.45, 2.17)	1.77*** (1.45, 2.17)
	[28d, 1y)	404	9288.69	638	15074.35	1.03 (0.91, 1.17)	1.01 (0.89, 1.15)	1.06 (0.92, 1.21)	1.05 (0.92, 1.21)
	[1y, 5y)	393	33711.11	611	50957.39	0.99 (0.87, 1.12)	1.00 (0.88, 1.14)	0.99 (0.87, 1.13)	0.98 (0.86, 1.12)
	[0,5y)	1007	43827.27	1450	67387.50	1.10* (1.02, 1.20)	1.11* (1.02, 1.21)	1.13** (1.03, 1.23)	1.12** (1.03, 1.22)
Somkhele	[0,28d)	23	142.72	23	209.00	1.46 (0.82, 2.61)	1.36 (0.74, 2.48)	1.38 (0.75, 2.53)	1.40 (0.76, 2.58)
	[28d, 1y)	97	1563.59	109	2323.72	1.32* (1.00, 1.73)	1.48** (1.11, 1.98)	1.51** (1.12, 2.05)	1.52** (1.13, 2.04)
	[1y, 5y)	37	5527.79	68	8157.08	0.81 (0.54, 1.21)	0.91 (0.59, 1.38)	0.83 (0.55, 1.26)	0.83 (0.55, 1.26)
	[0,5y)	157	7234.10	200	10689.79	1.16 (0.94, 1.43)	1.28* (1.03, 1.59)	1.27* (1.01, 1.59)	1.28* (1.02, 1.60)

Notes: †p<0.10; *p<0.05; **p<0.01; ***p<0.001

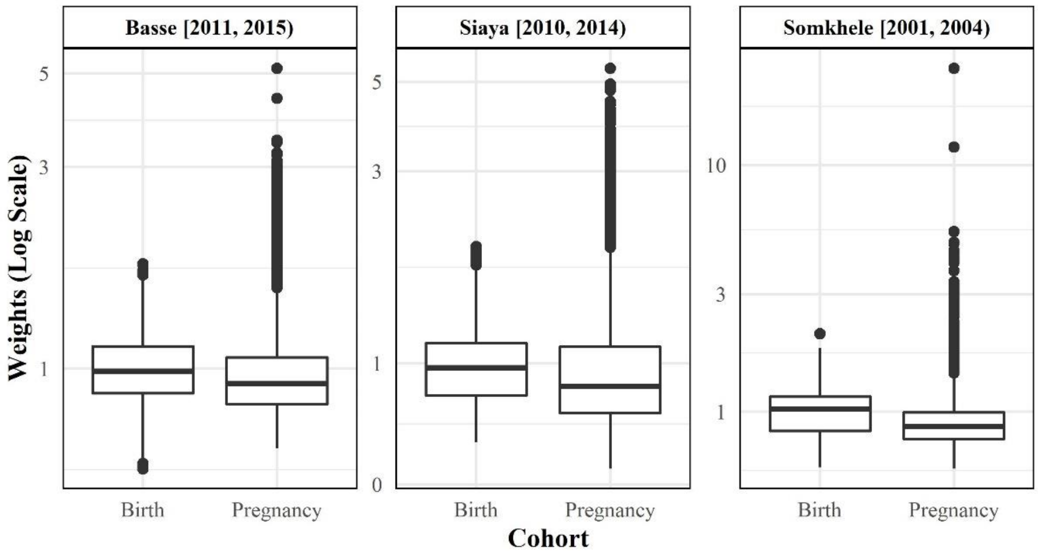

Figure A1: Distribution of inverse probability of pregnancy registration weights for birth and pregnancy cohorts in each HDSS site.

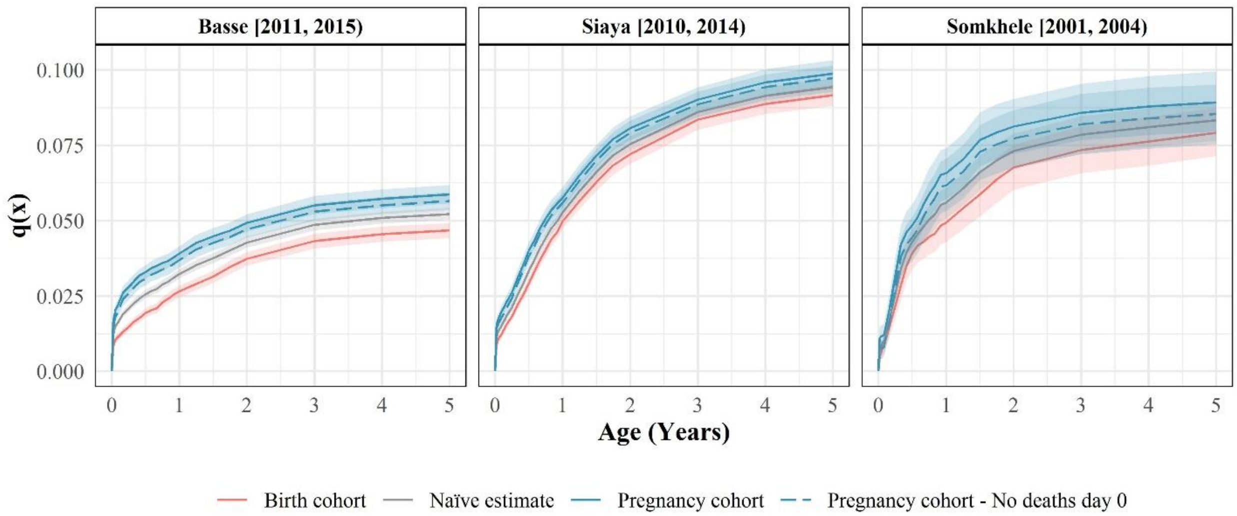

Figure A2: Sensitivity analysis: all deaths taking place during the first day of life were removed from the pregnancy cohort, and cumulative probabilities of dying under 5 were recalculated. Bootstrap 95% confidence intervals were calculated for 1,000 resamples with replacement.