Measuring neonatal mortality from survey data in low- and middle-income countries: results and implications of a validation study in Guinea-Bissau

Liu, Li[1]
Fisker, Ane[2]
Chu, Yue[3]
Rodrigues, Amabelia[2]
Helleringer, Stephane[1,*]

[1] Johns Hopkins University, Bloomberg School of Public Health
[2] Bandim Health Project
[3] The Ohio State University, Department of Sociology
* Correspondence to Stephane Helleringer: sheller7@jhu.edu

Background: Reducing neonatal mortality is a key global health priority. Data on neonatal deaths in low- and middle-income countries (LMICs) come primarily from periodic household surveys. There are however no recent validation studies of such data. Our objective was to assess reporting errors in survey data on neonatal mortality, and to evaluate their effects on estimates of neonatal mortality rates (NMR).

Methods and findings: We conducted a validation study of survey data on neonatal mortality in Guinea-Bissau. We used data from a large health and demographic surveillance system (HDSS) that monitors pregnancies and child outcomes prospectively as our reference dataset. Using HDSS records, we selected a stratified random sample of 599 women aged 15-49 years old, and collected their full birth history (FBH). We estimated the sensitivity and specificity of FBH in recording neonatal and post-neonatal deaths by cross-tabulating FBH and HDSS data. We attributed the errors we observed to date displacements, age errors, omissions or misclassifications. We used a mathematical model to draw the implications of reporting errors for the accuracy of NMR estimates. Through logistic regressions, we explored the association of reporting errors with maternal and child characteristics. Lastly, we evaluated the potential impact of reporting errors in FBH data on measures of socioeconomic inequalities in neonatal mortality.

Our survey data collection achieved a 85.2%% participation rate. The sensitivity of FBH data in recording neonatal deaths was 79.1% (95% confidence interval (CI): 72.7-85.5%), and the specificity was 99.3% (95% CI: 98.9-99.6%). Specificity was lower among respondents who had also experienced a stillbirth (86.6%, 95% CI: 74.8-93.4%) or a post-neonatal death (84.0%, 95% CI: 73.2-90.9%), than among respondents who had not experience such events. Omissions of births and misclassifications between stillbirths and neonatal deaths accounted for most errors in survey data on neonatal deaths. Age errors accounted for the majority of errors in the reporting of post-neonatal deaths. In populations with known distributions of stillbirths, neonatal deaths and post-neonatal deaths, these reporting errors would lead to NMR estimates that are too high. Low educational levels of the mother were associated with lower sensitivity of FBH data in recording neonatal deaths (69.6% among mothers without schooling, vs. 87.0% among mothers with secondary schooling or higher). This might lead to under-estimating differentials in neonatal mortality associated with maternal education.

Conclusions: Our validation study indicates that survey data on neonatal mortality might misrepresent progress towards global newborn survival targets, as well as socioeconomic inequalities in neonatal mortality. Methodological research is needed to a) improve survey instruments and b) account for reporting errors in estimates of the NMR.

BACKGROUND

A neonatal death is a death that occurs during the first 28 days of life. In 2017, the UN Inter-agency Group for Child Mortality Estimation (UN-IGME) estimated that 47% of all children who died before age 5 died in this age range (1). Reducing the neonatal mortality rate (NMR) is now a key target of the 3rd Sustainable Development Goal (SDG). That is by 2030, countries should strive to end preventable deaths of newborns, and to reduce neonatal mortality to at least as low as 12 deaths per 1,000 live births (2). To facilitate progress towards this target, accurate data on the levels and trends of neonatal mortality are needed (3-8). This will allow better targeting interventions to reduce neonatal mortality, and could allow monitoring the effects of changes in intervention coverage (9).

In most high-income countries, such data come from civil registration and health information systems that operate continuously. In low- and middle-income countries (LMICs), these systems are often deficient (10-13). A large proportion of births and newborn deaths occur at home, and are not promptly registered by family members. Even when these events occur within health facilities, their recording might be incomplete, delayed or inaccurate (11).

Instead, data on neonatal mortality are collected periodically in most LMICs during household surveys, such as the Demographic and Health Surveys (DHS), and the Multiple Indicator Cluster Surveys (MICS). These surveys frequently conduct full birth histories (FBH): women aged 15-49 years old are asked to report all of their live births, and to state the date of birth of each child, and whether he/she is still alive. For each deceased child, they are also asked to report at what age the child died. From these

data, demographers can directly obtain NMR estimates for recent time periods (e.g., the past 3 or 5 years), by dividing the number of reported neonatal deaths by the number of births reported to have occurred over the same time frame (14). Such estimates form the cornerstone of global studies of the patterns of neonatal mortality (1, 15, 16). They also constitute the standard that new methods to interpret other data sources on child mortality (e.g., census data) are evaluated against (17, 18) or are trained to emulate (19).

Estimates of the NMR obtained from FBH may however be affected by a number of issues. First, they may contain missing data, e.g., on the date of birth of a child, or on his/her age at death. These fields thus need to be imputed, often requiring strong assumptions about the underlying causes of missing data. The frequency of missing data on dates of birth and age at death in FBH has however declined over time in major surveys such as the DHS (20, 21). In most recent such surveys, less than 5% of the reported births have missing information on a component of their date of occurrence such as the month of birth (20). Second, FBH data may display “survivor bias” (22): since they are collected retrospectively, FBH only include data from women who have survived until the time of the survey. If the children of the mothers who died before the survey were also more likely to die than other children (23-26), then NMR estimates will be too low. Simulation studies however suggest that survivor bias might be corrected, even in contexts affected by epidemics of infectious diseases that may be transmitted from mother to child such as HIV and Ebola (27, 28). Finally, FBH data on neonatal mortality may be affected by limited sample size because even in high mortality settings, neonatal deaths remain a rare (statistical) event (22).

In some instances, FBH data collected during a survey may also differ from the true birth

history of a respondent (Figure 1). This may result from four types of reporting or respondent errors. First, *date displacement* occurs when a respondent's answers to the FBH imply a date of birth/death of her child(ren) that differs from the true date(s). Date displacement affects estimates of the NMR when they shift a birth/death into or outside of the reference period for which estimates are sought (e.g., the past 3 or 5 years). Second, *age errors* occur when a respondent misstates the age at which their child died. This is often detected by measuring the extent of heaping in FBH data, i.e., a higher than expected number of events reported to have occurred at specific ages such as 7 days or 12 months (22, 29). Age can be over-stated, for example when a child who died during the neonatal period is reported to have died at an age greater than 28 days. Age can also be under-stated, for example, when a child who died in the post-neonatal period is reported to have died as a newborn. Age errors can affect NMR estimates when deaths are erroneously shifted in and out of the neonatal period.

[FIGURE 1 ABOUT HERE]

Third*, omissions* occur when respondents do not list all of their live births during the FBH. This may happen because of recall issues (e.g., respondents not reporting births that have occurred in a more distant past), because a birth may evoke painful memories (e.g., the death of a child) or because a birth is associated with stigmatized behaviors (e.g., out-of-wedlock childbearing). Omissions might affect NMR estimates if they occur among neonatal deaths. Finally, *misclassifications* occur when respondents erroneously report stillbirths (i.e., a baby born with no signs of life at or after 28 weeks' gestation) as neonatal deaths, or leave a neonatal death out of the FBH, because they considered that the child was not viable.

There are several factors that explain the emergence of these reporting errors in FBH. Demographers have often argued that poor maternal recall of events that occurred several years ago, as well as limited numeracy among some population groups, might explain errors in FBH data. Adjustment procedures have been proposed to account for such recall patterns when analyzing survey data on mortality (30, 31). These methods however make strong assumptions. For example, they consider that respondents accurately report the most recent events, e.g., those that occurred in the past 0-3 years prior to the survey. In validation studies of survey data on mortality among adults, this assumption did not hold (32). They also consider that the proportion of events (i.e. deaths) that are recalled by a respondent declines linearly with the amount of time elapsed between the event and the survey. Studies in cognitive psychology however suggest that this assumption may not be met in practice: more distant events might be easier to recall than more recent ones depending on a complex array of emotional, neural and cognitive factors (33).

Others have emphasized possible social desirability biases in reporting of adverse perinatal events such as neonatal deaths and stillbirths (34). Women might conceal these aspects of their birth histories during a survey interview because they fear judgmental attitudes, gossip or stigma, either from the interviewer, relatives or other community members. Errors may also emerge in FBH because the local terms used by women to describe pregnancy outcomes in LMICs often do not match the western biomedical concepts of stillbirths and neonatal deaths. Haws et al. (34), for example, documented that women in Tanzania used terms (e.g., “immature baby”) that could be applied to both stillbirths and neonatal deaths. Even in health facilities, the fetal heart rate is not systematically monitored in LMICs, so that health workers might mistakenly classify some babies who do not show signs of movement (e.g., due to birth asphyxia)

as stillbirths (35).

Some of these factors may be accentuated by the behaviors of data collectors. Interviewers, for example, may be tempted to skip recording some births reported by the respondent in order to speed up the interview. This would then lead to a number of omissions. They might also report some children as being born earlier than they are in order to skip long series of questions that are only applicable for specific age groups. This might for example occur in the DHS or in the MICS, where time-consuming modules about post-natal care, child health and other topics are only applicable to children born within the past 5 years (20, 36). This would then lead to an increased frequency of date displacement and/or age errors.

Few studies have quantified the extent of these reporting errors in FBH data. Liu et al. (37) used data collected in Malawi to assess how common misclassifications might be in FBH data. Compared to a more intensive verbal autopsy (VA) questionnaire conducted several months after the FBH survey, they found that 21% of the neonatal deaths reported during FBH might in fact have been stillbirths. Based on qualitative interviews, Haws et al (34) also uncovered discrepancies between a woman's reproductive history and her reports of reproductive events during a survey interview. For example, they described several cases of women who reported a stillbirth while indicating that the baby had moved or breathed at the time of delivery, or who reported a neonatal death while describing that the baby's body was cold or had deteriorating skin (thus suggesting it might have been a stillbirth).

A more comprehensive validation study of FBH data was conducted in the 1990's in Bangladesh (38). It indicated that a number of neonatal and post-neonatal deaths might

not be accurately reported during FBH. In that study, FBH respondents only reported 81.0% of the neonatal deaths they had experienced according to a reference prospective dataset extracted from the Matlab Health and Demographic Surveillance System (39). Unfortunately, while this study investigated the effects of reporting errors on estimates of the total fertility rate, it did not investigate the implications of reporting errors for survey-based estimates of the NMR. Other assessments of the quality of FBH data have been reliability studies, in which estimates of mortality rates obtained from FBH were compared to estimates obtained from other sources or to predictions from statistical models (21, 22). Such studies are thus based on benchmarks that contain significant limitations (e.g., model life tables or census data). They have also focused on infant mortality (29), rather than specifically on neonatal mortality.

The impact of reporting errors on estimates of the NMR obtained from FBH data is difficult to predict. This is so because some errors might typically lead to downward biases (e.g., omissions), whereas other errors might lead to upward biases (e.g., misclassifications of stillbirths as neonatal deaths). In some settings, these errors might offset each other and result in a roughly accurate NMR estimate. In other settings though, they might cumulate and result in severely biased NMR estimates. In this paper, we report the results of a detailed validation study of FBH conducted in an urban setting of Guinea-Bissau in 2016-2017. We then use a simple mathematical model to investigate the net effects of reporting errors in FBH data on estimates of 1) the level of the NMR, and 2) differences in NMR between population groups.

DATA AND METHODS

Study setting: Guinea-Bissau is a low-income country in West Africa, with a population of approximately 1.8 million and an estimated life expectancy of 56 years in 2010-15 (40). The most recent nationally representative FBH survey was conducted in 2014. It documented a NMR of 36, and a post-neonatal mortality rate of 20, per 1,000 live births nationwide (41). The survey also indicated a gradient in neonatal mortality associated with educational level of the mother: there were more than 40 neonatal deaths per 1,000 live births among uneducated mothers, vs. 28 among mothers with secondary schooling or higher.

Reference dataset: We worked within the urban Health and Demographic Surveillance System (HDSS) of the Bandim Health Project (BHP). This is an open cohort that monitors the populations of six neighborhoods of Bissau, the capital city (e.g.,42, 43-45). The areas covered by the BHP are situated approximately 2 km away from the city centre and include more than 103,000 inhabitants. In these areas, the HDSS records pregnancies, births, deaths and migrations since 1978. Every month, fieldworkers visit every household in HDSS areas to register pregnancies and record their outcomes (e.g. stillbirth vs. live birth). Children are then followed every 3 months to record survival, nutritional status and health-seeking behaviors. BHP also registers all births at the National Hospital (Simão Mendes) (45) and at a health center serving parts of the HDSS population (Centro de Saude de Bandim), and these data are linked with the HDSS records. New households (e.g. those established after construction of a new dwelling) are added to the BHP data set on a continuous basis.

We used data from the HDSS as the reference dataset against which we evaluated FBH data. Among HDSSs, the BHP has one of the most rigorous protocols for monitoring

perinatal events and infant survival, because it is based on monthly household visits to detect and monitor pregnancies and their outcomes (46). Despite its high quality, however, the HDSS data do not constitute a gold standard measure of neonatal mortality. This is so in part because some pregnancies might be missed during monthly household visits, and because some pregnant women and their children might migrate outside of the HDSS area during follow-up. In addition, neither the HDSS nor the local hospitals and health centers collect precise data on fetal heart rate after delivery using highly sensitive monitors. As a result, a small number of stillbirths may be misclassified as neonatal deaths in these HDSS data, and vice-versa.

Data collection: We selected a stratified random sample of women aged 15-49 years old from the lists of HDSS residents. We oversampled women in three strata: those who had experienced in the past 5 years either 1) a neonatal death among their live-born children, 2) a post-neonatal death among their live-born children, or 3) a stillbirth. This was necessary to ensure that we had sufficient reported numbers of such events to evaluate the accuracy of FBH. The fourth sampling stratum was constituted of women who did not experience any of these events according to the HDSS over the past 5 years prior to the validation study e.g., all of their children were still alive, or had died after age 1 year. We designed sampling weights to account for differences in the probability of inclusion and participation rates in each of these strata.

Our questionnaire included a subset of the modules of the 2014 Guinea-Bissau MICS questionnaire: a) questions about the respondent's socioeconomic background (e.g., age, educational level, marital status), b) summary questions about her fertility, and c) a standard FBH. The FBH module first asked respondent to list all the live births they have ever had in chronological order. Then, for each reported live birth, it asked the gender of

the child, whether he/she was part of a multiple birth, whether he/she was still alive at the time of the survey, and the date of birth. For children who were still alive, the FBH asked respondents to report their age and residence. For children who had died, it asked respondents to state the age at death in days if they had died within their first month of life, in months if they had died between 1 and 23 months, and in years if they had died at older ages.

Similar to the 2014 MICS, the questionnaire also included d) questions about pregnancies that did not result in a live. We asked respondents if they had ever experienced such a pregnancy termination. If so, we asked them how many such terminations they had experienced over their lifetime, and when was the most recent one. We did not ask respondents to state the type of pregnancy termination they experienced (e.g., stillbirth vs. miscarriage or abortion). Finally, as in the 2014 MICS, we included e) questions about live births of the past two years prior to the survey. These questions focused on antenatal and post-natal care, birth and (if applicable) death registration, and the care of recent illnesses such as cough, diarrhea and/or fevers.

We recruited interviewers who had collected FBH data during the 2014 MICS in Guinea-Bissau. We provided them with refresher training for a week before beginning data collection for the validation study. The interviewers were not given HDSS data prior to conducting FBH interviews. In particular, they did not know whether the women they had to interview had experienced any births, adverse perinatal events (e.g., stillbirths, neonatal deaths) or post-neonatal deaths over the past 5 years. We devised an electronic data collection tool that emulated the tool used in MICS and DHS surveys. Specifically, we used the Qualtrics platform (47) on android tablets. We selected Qualtrics because it allows collecting FBH data within a single screen, by displaying all

reported live births in a large table, as is currently done in DHS and MICS (which collect data on more expensive Windows-based tablets using CSpro). Other android platforms such as Survey Solutions or Open Data Kit (48, 49) do not allow collecting FBH data in this fashion. Instead, they ask interviewers to establish a list of live births, and then they ask each question about these live births on a separate screen. FBH data collected using these platforms might thus not be comparable to FBH data collected by DHS or MICS.

As in MICS and DHS, we incorporated data consistency checks into our Qualtrics program, which alerted interviewers to potential errors or implausible FBH patterns (e.g., intervals between live births that are too short). All data were uploaded every day to a cloud-based server, and checked by a data editor. All study instruments were translated into Portuguese and Creole, the vernacular language most commonly used in Bissau. As was done during the 2014 MICS, questions were displayed in Portuguese, but asked in Creole, except for 2 interviews that required a translator and were conducted in French (among Fulani migrants recently arrived from the neighboring Republic of Guinea).

Data analysis: We first described the constitution of the study sample, including reasons for non-participation and non-inclusion. We investigated the selectivity of the study sample by comparing the HDSS records of the women who participated in the validation study, to the HDSS records who were selected but could not be interviewed. This comparison focused on maternal age and history of neonatal deaths, post-neonatal deaths, and stillbirths. Details are included in appendix A1. We then described the characteristics of study participants. We considered maternal characteristics such as age group (15-24y, 25-34y and 35y and older), educational level (no schooling vs. primary schooling vs. secondary schooling and higher), marital status (currently in an

union vs. never in an union, vs. previously in an union), religion (Catholic vs. Muslim vs. Protestant vs. other) and ethnicity (Pepel vs. Fulani vs other ethnic groups). We also considered characteristics of the "index" pregnancy, i.e., the pregnancy/birth that prompted inclusion of the respondent in one of the sampling strata described above. For example, consider a woman who had 3 pregnancies recorded by the HDSS over the past 5 years: two of these pregnancies resulted in live births who then survived until age 1, whereas the third pregnancy resulted in a live birth that later died after 7 days. This woman was then included in the "neonatal death" stratum, and the pregnancy characteristics we considered in these analyses concern the pregnancy that resulted in a neonatal death (the "index" pregnancy). These characteristics included gender of the child, place of delivery (at home, vs. at hospital, vs. at health center, vs. elsewhere) and whether the delivery took place via cesarean section.

Evaluation of the accuracy of FBH data: We cross-tabulated the FBH data with reference data extracted from HDSS records. We did so separately for neonatal and post-neonatal deaths. We defined the sensitivity of FBH data as the proportion of respondents with neonatal/post-neonatal deaths in the previous 5 years according to the HDSS, who reported experiencing such events in the same timeframe during FBH. Conversely, we defined the specificity of FBH data in recording neonatal/post-neonatal deaths as the proportion of respondents without neonatal/post-neonatal deaths in the previous 5 years according to the HDSS, who did not report such events in the same timeframe during FBH. We investigated whether the specificity of FBH data in recording neonatal deaths varied across the three sampling strata in which women had not experienced a neonatal death according to the HDSS (e.g., women who had experienced a stillbirth or a post-neonatal death over the past 5 years).

In this context, we defined a false negative report as a respondent with a neonatal/post-neonatal death according to HDSS who did not report such a death during FBH. A false negative was attributed to 1) a date error if the respondent reported a neonatal/post-neonatal death before the reference period; 2) an age error if the respondent reported a death in another age group during the reference period; 3) an omission if the respondent did not report any under-5 death or pregnancy termination during the reference period, and 4) a misclassification if the respondent did not report any under-5 death but reported a pregnancy termination during the reference period. Similarly, a false positive (i.e., a respondent without a neonatal/post-neonatal death according to HDSS who nonetheless reported such a death during FBH) was attributed to 1) a date error if the respondent reported a neonatal/post-neonatal death that had occurred prior to the reference period as having occurred during the reference period; 2) an age error if the respondent had experienced a neonatal, post-neonatal or child death according to HDSS but mistakenly reported it as having occurred in another age group during the reference period; and 3) a misclassification if the respondent had experienced a pregnancy termination according to HDSS but reported it as a neonatal death during the reference period of the FBH. All these analyses used sampling weights to account for differences in the probabilities of selection and participation across strata. For each estimate of a proportion described above, we calculated the 95% confidence interval.

Effects of FBH reporting errors on NMR estimates: We devised a mathematical model to investigate the effects of reporting errors in FBH data on the accuracy of NMR estimates. Briefly, this model links the true NMR in a population to a survey estimate of the NMR, via a set of conditional probabilities describing reporting errors in FBH (e.g., sensitivity, specificity). A complete description of the model is given in appendix A2. Following a common approach in assessments of perinatal mortality data (22), we

applied this model to data series from England and Wales, where live births, stillbirths, neonatal deaths, infant and under-5 deaths have been reported annually since the 1920's (appendix A3). For each 5-year time period, we used our model and our validation results to calculate the estimate of the NMR that would have been obtained if a FBH survey had been conducted in this population. Then, we compared this counterfactual estimate to the estimate that was recorded through vital statistics in England and Wales.

Our estimates of the accuracy of FBH data (e.g., sensitivity) are affected by sampling errors. To account for this uncertainty in our assessment of the effects of reporting errors on NMR estimates, we calculated 95% confidence intervals. For each population-period, we generated 1,000 survey estimates of the NMR through random draws from the distributions of each reporting parameter (e.g., sensitivity), in conjunction with model equations. We then calculated the 2.5^{th}, 50^{th} and 97.5^{th} percentiles of the distribution of these estimates, and compared them to the NMR estimates available from vital statistics (England & Wales)

Robustness tests: Because our reference data (HDSS) in this validation study do not constitute a gold standard measure of neonatal mortality, we performed several tests to evaluate the robustness of our findings to possible errors in HDSS data. We conducted a series of re-analyses of the validation study data in which we 1) randomly reclassified events recorded by the HDSS datasets (e.g., some stillbirths were recoded as neonatal deaths, and vice-versa), and 2) introduced additional neonatal deaths in the HDSS dataset to account for the potential incompleteness of HDSS registration of pregnancies. We developed 4 conservative scenarios, reflecting the likely magnitude of possible errors in HDSS data on neonatal mortality. The scenarios are described in Appendix A4.

We replicated each scenario 1,000 times, and re-calculated our study outcomes (e.g., sensitivity of FBH data) in each of these replicates. We then evaluated whether our assessment of the direction of bias in NMR estimates derived from survey data would be altered in each scenario. Throughout these robustness tests, we assumed that the recording of dates and ages is accurate in HDSS due to the intensive follow-up by HDSS fieldworkers. We thus do not introduce age errors or date displacements in HDSS data.

Differential reporting between population groups: Finally, we investigated differences in FBH accuracy between population groups. We focused on non-reported neonatal deaths in FBH (i.e., false negatives) because there were too few cases of other errors (e.g., false positives). Among respondents who experienced a neonatal death according to HDSS, we created a binary variable indicating whether they reported such a death during the FBH survey (i.e. a true positive). We then used logistic regressions to assess the association between maternal and child characteristics on one hand, and the likelihood of reporting a neonatal death during the survey on the other hand. For these analyses we also considered the sampling weights described above.

Using the mathematical model in appendix A2, we investigated whether differences in the sensitivity of FBH data between population groups might lead to bias in survey estimates of the differences in NMR between those groups. We focused on differentials in NMR associated with maternal education because the 2014 MICS suggested a large NMR gradient by maternal education (41). We analyzed two hypothetical scenarios, which represent two possible patterns of inequalities in perinatal and under-5 survival. In the first scenario (“constant inequality”), the risks of stillbirth and under-5 mortality rates (e.g., neonatal and post-neonatal),are elevated by a similar factor among mothers without schooling, relative to mothers with secondary school or higher. In the second

scenario (“age-varying inequality”), the risks of stillbirth and neonatal deaths are elevated by a similar factor among mothers without schooling, but there are no differences in the risk of post-neonatal mortality and in the risk of dying at ages 1-4 years old between the two educational groups. This latter scenario reflects contexts in which post-neonatal mortality is reduced to low levels throughout the population, due for example to a strong and equitable vaccination program (50, 51). The scenarios are described more fully in appendix A5. In each scenario, we let the true risk ratio of neonatal mortality associated with lack of maternal education vary between 1 and 2.5. We then plot the relationship between the true and estimated risk ratios. All data analyses were carried out using STATA 14.

RESULTS

We selected 599 records of women aged 15-49 years old among HDSS rosters (figure 2). After reviewing HDSS and linked hospital records, we excluded 13 women who had died prior to the study, 3 who were duplicate registrations, and 1 who was older than 49 years old. Thus, we contacted 582 women to offer study participation. Among those, 7 had died prior to our visit, and 80 had permanently migrated outside of the HDSS area since the last HDSS visit. They were thus ineligible for study participation. Among the 495 women who were eligible for enrollment, 66 were absent at all 3 study visits, and 3 were unknown. In addition, in two instances, we interviewed someone who was not the target HDSS resident. This was detected initially by large discrepancies between the age reported by the respondent (>20 years) and her birth record in the HDSS. It was then confirmed by household inquiries made by study supervisors. In two other instances the respondent was registered by the HDSS only after the birth of (one of) her child(ren). As a result, prospective data on neonatal mortality are not available for these two

respondents and they were excluded. Our analytical sample thus included 422 women, yielding an overall participation rate of 85.3% (422/495).

[FIGURE 2 ABOUT HERE]

The women included in this analytical sample were not older than those who were not included (appendix A1). In each sampling stratum, the women included in our analytical sample also did not differ from other women on key aspects of their pregnancy histories (Appendix A1). One exception to this pattern is related to the experience of post-neonatal deaths: women included in the analytical sample appeared less likely to have experienced such a death than women who were not included (p=0.006). In total, 178 respondents had experienced at least one neonatal death over the past 5 years according to the HDSS, 86 had experienced a post-neonatal death and 89 had experienced at least one stillbirth (figure 2).

Approximately one in four respondents were aged 15-24 years old, whereas one in five were 35 years or older (table 1). The educational level was low, with 17.0% of respondents never having attended school. Two thirds of the respondents were in a union at the time of the survey, and 4.7% were divorced, widowed or had separated. Catholics and Muslims were the two main religious groups (46.6% and 32.5%, respectively), whereas Fula and Pepel constituted the two largest ethnic groups (20.1% and 27.3%, respectively). The majority of (index) children (live born or stillborn) were male; reflecting increased neonatal and post-neonatal mortality among boys in Bandim (52-54). Approximately, one in five index deliveries occurred at home, whereas 44.0% of those deliveries occurred in the national hospital Simao Mendes. Finally, only a small minority of the index deliveries (8.3%) occurred via cesarean section.

[TABLE 1 ABOUT HERE]

The sensitivity of FBH in recording neonatal deaths was moderate (table 2, 79.1%, 95% CI = 72.0% to 84.8%), but the specificity was high (99.3%, 95% CI = 98.9% to 99.6%). The specificity varied however depending on the pregnancy history recorded by the HDSS. Among respondents with at least one stillbirth in the past 5 years, specificity was 86.6% (95% CI = 74.8 to 93.4%), whereas it was 84.0% (95% CI, 73.2% to 90.9%) among respondents with a post-neonatal death in the past 5 years, and 100.0% (i.e., no false positive reports) among other respondents. The specificity of FBH in recording post-neonatal deaths was 99.9% (95% CI = 99.8 to 100.0%), whereas its sensitivity was 69.1% (95% CI = 57.7% to 78.6%).

[TABLE 2 ABOUT HERE]

False negative reports of neonatal deaths were due in large part to misclassifications of neonatal deaths as stillbirths (43.4%, figure 3). However, omissions also accounted for close to a third of false negatives (30.4%). Date displacements and age errors accounted for smaller proportions of false negatives (18.0% and 8.2%, respectively). For post-neonatal deaths, the majority of false negatives stemmed from age under-statement and omissions.

[FIGURE 3 ABOUT HERE]

Among false negative reports of neonatal deaths, the sources of error differed between deaths that had occurred in the early neonatal period (0-6 days) and those that occurred

in the late neonatal period (7-27 days) according to the HDSS (Appendix A6). Among early neonatal deaths that were not reported during FBH, close to two thirds were attributed to misclassifications. All of those deaths were neonatal deaths that had occurred on the day of birth according to the HDSS. On the other hand, there were no misclassifications among the deaths that occurred in the late neonatal period according to the HDSS. Instead, errors in that period originated in roughly equal proportion from date errors, age over-estimates and omissions.

False positive reports of neonatal deaths in FBH data were primarily due to misclassifications of stillbirths as neonatal deaths (62.7%, figure 3). Other sources of false positive reports of neonatal deaths included age under-statement (31.4%) and date displacement (5.9%). All the false positive reports of neonatal deaths attributable to misclassifications were reported to have occurred within the first two days after birth according to the respondent. There were few false positive reports of post-neonatal deaths, but all of them were due to age over-statement.

When inserted into our mathematical model, these patterns of reporting errors appeared to imply that FBH data over-estimates the NMR. When applied to data from England & Wales between 1927 and 2017 (figure 4), our model predicted a survey estimate of the NMR that would be 4.7% to 19.8% higher than the NMR reported from vital statistics. This upward bias appeared to be highest for time periods when the “true” NMR was either high (approximately 30 per 1,000), or low (< 5 neonatal deaths per 1,000) in England and Wales. The confidence intervals associated with our estimates of bias in survey estimates of the NMR were however wide. Except for a cluster of estimates when neonatal mortality was around 30 deaths per 1,000 live births in England and Wales, the

confidence intervals routinely included the possibility that a FBH survey would result in a NMR that was either unbiased or too low (i.e., bias <1).

[FIGURE 4 ABOUT HERE]

In robustness tests (figure 5), we explored how our findings might be affected by errors in reference HDSS data. We found that our assessment of the *direction* of bias in NMR estimates obtained from survey data would not be altered by the introduction of such errors. Even in a very conservative scenario (scenario D) where we considered that misclassifications were common in HDSS data, and that the HDSS might have failed to record up to 1 in 5 pregnancies, FBH data would on average still lead to over-estimates of the NMR in most settings similar to those observed in England and Wales.

[FIGURE 5 ABOUT HERE]

We found limited differences in the likelihood of accurately reporting a neonatal death (sensitivity) across maternal and child characteristics (table 3). However, the sensitivity of FBH data was associated with the educational level of respondents in our sample: among women with secondary schooling or higher, it was 87.0% vs. 69.6% among women with no schooling. In adjusted logistic regressions, the odds of accurately reporting neonatal deaths were 3.58 times higher among women with secondary education or higher than those with no schooling (95 % CI = 1.05 to 12.2).

[TABLE 3 ABOUT HERE]

In figure 6, we investigated the impact of such differentials in the sensitivity of FBH data on our ability to detect differences in NMR between educational groups using survey data. We found that these reporting patterns would lead to under-estimates of the extent of educational inequalities in NMR in the two scenarios we considered (figure 6). In these scenarios, when there were no educational differences in NMR (rate ratio = 1), FBH data erroneously suggested that the children of women without schooling experienced lower neonatal mortality than the children of women with secondary school education or higher. When there were large educational differences in NMR, FBH data yielded estimates of rate ratios that were too low (e.g., 1.85 vs. 2.5 in the scenario of "age-varying inequality").

[FIGURE 6 ABOUT HERE]

DISCUSSION

Using a high-quality reference dataset from the Bandim Health Project in Guinea-Bissau, we found that FBH data collected during surveys had high specificity (99.3%) but moderate sensitivity (79.1%) in recording neonatal deaths. These estimates of the accuracy of FBH data on neonatal deaths are comparable to those obtained in a prior validation study conducted in Matlab, Bangladesh in the 1990's (38), where a sensitivity of 82.0% was obtained. However, our estimates of the sensitivity of FBH data in recording post-neonatal deaths in Guinea-Bissau were much lower than those obtained in Bangladesh (69.1% in Bandim vs 87.0% in Matlab).

Our validation study adds to prior work on the accuracy of FBH data in several important ways. First, we provided estimates of the specificity of FBH in recording neonatal and post-neonatal deaths. Such estimates were not available from the Matlab study in Bangladesh, even though they are crucial in evaluating the potential consequences of reporting errors on the accuracy of mortality estimates. Indeed, since neonatal (and post-neonatal) deaths are rare events, large biases can result from even small imperfections in the specificity of survey data on these events. Second, we investigated the sources of errors in FBH data. We showed that FBH data on neonatal mortality were vulnerable to all 4 potential sources of reporting errors: date displacement, age errors, omissions and misclassifications. In particular, we found that stillbirths were often misclassified as neonatal deaths, whereas a number of post-neonatal deaths were transferred to the neonatal period due to age errors. This information might play a key role in guiding data improvement strategies and the development of new statistical models of the NMR that account for reporting errors in FBH data.

We found that the reporting errors observed in our validation study might lead to an upward bias in survey estimates of the NMR. Using a simple mathematical model applied to historical data from England and Wales on the distribution of perinatal and other under-5 deaths, we found that FBH data might lead to over-estimates of the NMR ranging from approximately 5 to 20%. These data on England and Wales spanned a broad array of perinatal and under-5 survival contexts, characterized by neonatal mortality rates ranging from 2.5 to >30 per 1,000 live births. The highest levels of bias in survey estimates of the NMR were however apparent at the highest levels of the NMR observed in England and Wales, i.e., approximately 30 neonatal deaths per 1,000 live births. Due to the lower sensitivity of FBH data on neonatal deaths among respondents

with low educational levels, we also found that FBH data might underestimate the extent of educational differentials in NMR.

Our study has several limitations. First, the size of our validation sample was too limited to estimate various reporting parameters precisely. For example, our estimates of the extent of age transfers between neonatal and post-neonatal groups were based on a few cases, thus resulting in high levels of uncertainty. This subsequently affected our assessment of the impact of reporting errors on the accuracy of NMR estimates (figure 4), which resulted in wide confidence intervals that often included the possibility that the survey estimate was unbiased. Because of the limited sample size, we also could not investigate more detailed patterns of reporting errors. We thus could not assess the covariates associated with false positive reports of neonatal and post-neonatal deaths, nor were we able to determine if the sources of errors varied according to key characteristics of the respondent (e.g., educational level).

Second, we used the FBH instrument that was implemented in the 2014 Guinea-Bissau MICS, but that instrument only inquired about pregnancy terminations, rather than asking respondents to also report the type of pregnancy termination they experienced (e.g., stillbirths or miscarriages). Due to these limitations of the survey instrument, we could not investigate the quality of survey measures of the stillbirth rate. Our assessment of the sources of errors in FBH data may also have been slightly biased. Indeed, when a neonatal death was not reported (false negative), we attributed this error to misclassifications between stillbirths and neonatal deaths as long as the respondent reported at least one pregnancy termination during the reference period. But if that pregnancy termination was not a stillbirth (e.g, if it was a miscarriage or an abortion), then the error might instead have been attributed to an omission instead.

Third, our reference HDSS dataset might have been affected by a number of errors in the recording of neonatal deaths. On the one hand, it might have included a number of misclassifications between stillbirths and neonatal deaths. Such errors might even be present in data from hospital settings, because in Bissau as in most other LMICs, health workers do not use monitors to precisely determine whether the delivery resulted in a detectable fetal heartbeat. Additionally, the HDSS might also have failed to record a number of pregnancies that have occurred in the target population. As a result, some participants in our study might have experienced a neonatal death among their children, even though the HDSS failed to record it. In robustness tests however, we obtained similar results after randomly reclassifying a varying fraction of all stillbirths, and of the neonatal deaths that occurred within a few days of delivery. Our robustness tests also indicated that even high levels of incompleteness in pregnancy records included in HDSS datasets would not alter our conclusions about the direction of bias in NMR estimates.

Fourth, our results may not be representative of other settings where FBH data are collected in LMICs. Our study population was urban, and thus our results likely do not extend to rural areas. Our study also did not include a number of potential participants who had either migrated permanently or were temporarily absent from the Bandim HDSS area. Finally, we found that the women who were included our analytical sample experienced fewer post-neonatal deaths among their children than women who could not be included. These patterns of selectivity might have affected our estimates of the sensitivity and specificity of FBH data in recording neonatal and post-neonatal deaths.

Fifth, our study did not measure the full extent of reporting errors among live children and among children who died at ages one year and above. We were only able to document some age errors that transferred those older children to the post-neonatal period, and we did not find any age errors that led to transfers from 1-4 years old to the neonatal period. We were not able to measure the prevalence of omissions and date displacements among live children, nor were we able to measure the extent of age errors that would transfer deaths among 1-4 year olds to the 5-9 year old age group. In our modeling of study data, we thus made assumptions about these reporting patterns, (see appendix A2). This is problematic because such errors in FBH data would lead to an upward bias in estimates of the NMR, by excluding a number of births from its denominator while leaving its numerator largely unaffected. In our analyses, we investigated how much these reporting parameters might impact the survey estimates of the NMR. We found that misreporting of child deaths might only have a limited, because such deaths have become increasingly rare over time. On the other hand, omissions and displacements of live children aged 0-4 years old might have a strong effect on biases in these estimates.

To parameterize our analyses, we used estimates of the prevalence of omissions and date displacements among live children and deaths between ages 1-4 years old computed by Pullum and Becker (2014). However, in doing so, we used their median estimates of the extent of these errors, whereas in some surveys they found that age transfers in particular could be much more prevalent. If errors in the reporting of FBH data about live children and about deaths among 1-4 year olds are more common than we assumed in this study (appendix A2), then the upward bias in survey estimates of the NMR might be larger than we estimated.

Sixth, HDSS participants in Bandim are frequently interviewed and followed-up about their reproductive health and the survival of their children. They may thus report their FBH more accurately during a survey interview than other women who are not in such frequent contact with data collectors. As a result, the extent and magnitude of reporting errors in FBH data might be larger in settings without similar follow-up. Additional validation studies of FBH data should be conducted in diverse LMIC settings with less intensive follow-up, for example through record linkages with antenatal care clinics.

Finally, we did not measure potential errors in the reporting of the clustering of neonatal or post-neonatal deaths among mothers (i.e., women experiencing multiple such events over a short period of time). This was not possible because of the limited extent of this clustering in areas covered by the Bandim HDSS (table A1). Indeed, in our sample, only 7 participants (<2%) had experienced more than 1 neonatal death among their children over the past 5 years according to the HDSS. Among those, only one did not report any neonatal death, whereas the other 6 participants reported all the neonatal deaths that were also recorded by the HDSS.

Despite these limitations, our study has important implications. On the one hand, it suggests that in order to improve the accuracy of FBH data collected during surveys, a combination of strategies might be needed. To avoid omissions of neonatal deaths, FBH questionnaires might be supplemented by probing questions or prompts asking respondents about events that might be more likely to be left out of birth histories (e.g., a death having occurred shortly after birth). To reduce the prevalence of misclassifications between stillbirths and neonatal deaths, FBH questionnaires might be supplemented by a series of questions about the vital signs displayed by newborns, e.g., what their skin condition looked like, and whether they moved, cried or breathed. Such questions are

often included in VA questionnaires, and might allow reclassifying a number of live births as stillbirths if the reported newborn did not display signs of life (37). Since misclassifications only appear to affect newborns that have died in the early neonatal period (appendix A6), these questions might be limited to that sub-group in order to save time and resources during surveys. In addition, survey questionnaires might be expanded to include a full pregnancy history, during which respondents are asked not only to list all of their live births as in FBH, but also the pregnancies that did not result in a live birth. This approach might help collect more accurate data on infant survival (38), by 1) limiting omissions of neonatal deaths, and 2) affording additional opportunities to probe for the proper classification of events as stillbirths or neonatal deaths. Large-scale trials comparing FBH data to data from full pregnancy histories are currently under way (18). Finally, to reduce age errors and date displacements, survey questionnaires might incorporate event history calendars, which have helped improve the quality of adult mortality data (55, 56).

On the other hand, our study also indicates that FBH should be used cautiously in the development and evaluation of new methods to estimate neonatal mortality. Indeed, FBH have recently served as a benchmark against which new data collection platforms or statistical models to measure NMR are evaluated (17, 19). While they represent the current best practice in mortality measurement in LMICs with limited vital records, FBH data also misrepresent the true extent of neonatal mortality in potentially complex ways. As a result, biases in NMR estimates obtained from FBH data might be transferred to these new methods. When possible, evaluations of new measurement tools should account for the possibility of bias in FBH data.

Finally, our study highlights that survey data collected through FBH might misrepresent progress towards the achievement of global health objectives related to newborn health, particularly for the most disadvantaged population groups. In order for FBH data to adequately play their expected role in the stewardship of global health programs targeting newborn and child health, they require statistical adjustments that properly account for the added bias and uncertainty linked to reporting errors. The development of such adjustments will likely require additional validation studies in other LMICs, to better understand whether the patterns of sensitivity and specificity of FBH data might differ in other settings, relative to what we observed in Guinea-Bissau. It will also require the development of statistical models that adequately reflect these properties of FBH data in their estimates of the NMR and associated assessments of uncertainty. This should constitute a priority for institutes and organizations currently engaged in assessing progress towards the achievement of the neonatal mortality target of the 3rd sustainable development goal.

REFERENCES:

1. IGME. 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; 2018.
2. UN. Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development. New York: United Nations; 2018.
3. Lawn JE, Blencowe H, Kinney MV, Bianchi F, Graham WJ. Evidence to inform the future for maternal and newborn health. Best Pract Res Clin Obstet Gynaecol. 2016;36:169-83.
4. Gulmezoglu AM, Lawrie TA, Hezelgrave N, Oladapo OT, Souza JP, Gielen M, et al. Interventions to Reduce Maternal and Newborn Morbidity and Mortality. In: Black RE, Laxminarayan R, Temmerman M, Walker N, editors. Reproductive, Maternal, Newborn, and Child Health: Disease Control Priorities, Third Edition (Volume 2). Washington (DC)2016.
5. Victora CG, Requejo JH, Barros AJ, Berman P, Bhutta Z, Boerma T, et al. Countdown to 2015: a decade of tracking progress for maternal, newborn, and child survival. Lancet. 2016;387(10032):2049-59.
6. Yoshida S, Martines J, Lawn JE, Wall S, Souza JP, Rudan I, et al. Setting research priorities to improve global newborn health and prevent stillbirths by 2025. J Glob Health. 2016;6(1):010508.
7. Moxon SG, Ruysen H, Kerber KJ, Amouzou A, Fournier S, Grove J, et al. Count every newborn; a measurement improvement roadmap for coverage data. BMC Pregnancy Childbirth. 2015;15 Suppl 2:S8.
8. Lawn JE, Blencowe H, Oza S, You D, Lee AC, Waiswa P, et al. Every Newborn: progress, priorities, and potential beyond survival. Lancet. 2014;384(9938):189-205.
9. Davey G, Deribe K. Precision public health: mapping child mortality in Africa. Lancet. 2017;390(10108):2126-8.
10. Phillips DE, AbouZahr C, Lopez AD, Mikkelsen L, de Savigny D, Lozano R, et al. Are well functioning civil registration and vital statistics systems associated with better health outcomes? Lancet. 2015;386(10001):1386-94.
11. Mikkelsen L, Phillips DE, AbouZahr C, Setel PW, de Savigny D, Lozano R, et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. Lancet. 2015;386(10001):1395-406.
12. AbouZahr C, de Savigny D, Mikkelsen L, Setel PW, Lozano R, Lopez AD. Towards universal civil registration and vital statistics systems: the time is now. Lancet. 2015;386(10001):1407-18.
13. Fisker AB, Rodrigues A, Helleringer S. Differences in barriers to birth and death registration in Guinea-Bissau: implications for monitoring national and global health objectives. Trop Med Int Health. 2018.
14. Moultrie T, Dorrington R, Hill A, Hill K, Timæus I, Zaba B. Tools for demographic estimation: International Union for the Scientific Study of Population; 2013.
15. Wang H, Liddell CA, Coates MM, Mooney MD, Levitz CE, Schumacher AE, et al. 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. Lancet. 2014;384(9947):957-79.
16. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet. 2015;385(9966):430-40.

17. Helleringer S, Arhinful D, Abuaku B, Humes M, Wilson E, Marsh A, et al. Using community-based reporting of vital events to monitor child mortality: Lessons from rural Ghana. PLoS One. 2018;13(1):e0192034.
18. Baschieri A, Gordeev V, Akuze J, Kwesiga D, Blencowe H, Cousens S, et al. Every Newborn-INDEPTH" (EN-INDEPTH) study protocol for a randomised comparison of household survey modules for measuring stillbirths and neonatal deaths in five Health and Demographic Surveillance site. Journal of Global Health. 2019;Fortchoming.
19. Burstein R, Wang H, Reiner RC, Jr., Hay SI. Development and validation of a new method for indirect estimation of neonatal, infant, and child mortality trends using summary birth histories. PLoS Med. 2018;15(10):e1002687.
20. Pullum T, Staveteig S. An assessment of the quality and consistency of age and date reporting in DHS Surveys, 2000-2015. Rockville, Maryland, USA: ICF; 2017.
21. Pullum T, Becker S. Evidence of Omission and Displacement in DHS Birth Histories. Rockland, Md: ICF International; 2014.
22. Hill K, Choi Y. Neonatal mortality in the developing world. Demogr Res. 2006;14:429-52.
23. Finlay JE, Moucheraud C, Goshev S, Levira F, Mrema S, Canning D, et al. The effects of maternal mortality on infant and child survival in rural Tanzania: a cohort study. Maternal and child health journal. 2015;19(11):2393-402.
24. Armstrong Schellenberg JR, Nathan R, Abdulla S, Mukasa O, Marchant TJ, Tanner M, et al. Risk factors for child mortality in rural Tanzania. Tropical Medicine & International Health. 2002;7(6):506-11.
25. Zaba B, Whitworth J, Marston M, Nakiyingi J, Ruberantwari A, Urassa M, et al. HIV and mortality of mothers and children: evidence from cohort studies in Uganda, Tanzania, and Malawi. Epidemiology. 2005:275-80.
26. Ng'weshemi J, Urassa M, Isingo R, Mwaluko G, Ngalula J, Boerma T, et al. HIV impact on mother and child mortality in rural Tanzania. Journal of acquired immune deficiency syndromes (1999). 2003;33(3):393-404.
27. Hallett TB, Gregson S, Kurwa F, Garnett GP, Dube S, Chawira G, et al. Measuring and correcting biased child mortality statistics in countries with generalized epidemics of HIV infection. Bull World Health Organ. 2010;88(10):761-8.
28. Walker N, Hill K, Zhao F. Child mortality estimation: methods used to adjust for bias due to AIDS in estimating trends in under-five mortality. PLoS Med. 2012;9(8):e1001298.
29. Guillot M, Gerland P, Pelletier F, Saabneh A. Child mortality estimation: a global overview of infant and child mortality age patterns in light of new empirical data. PLoS Med. 2012;9(8):e1001299.
30. Masquelier B, Reniers G, Pison G. Divergences in trends in child and adult mortality in sub-Saharan Africa: survey evidence on the survival of children and siblings. Popul Stud (Camb). 2014;68(2):161-77.
31. Obermeyer Z, Rajaratnam JK, Park CH, Gakidou E, Hogan MC, Lopez AD, et al. Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974-2006. PLoS Med. 2010;7(4):e1000260.
32. Helleringer S, Pison G, Kante AM, Duthe G, Andro A. Reporting errors in siblings' survival histories and their impact on adult mortality estimates: results from a record linkage study in Senegal. Demography. 2014;51(2):387-411.
33. Wixted JT. The psychology and neuroscience of forgetting. Annu Rev Psychol. 2004;55:235-69.
34. Haws RA, Mashasi I, Mrisho M, Schellenberg JA, Darmstadt GL, Winch PJ. "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. Soc Sci Med. 2010;71(10):1764-72.
35. Spector JM, Daga S. Preventing those so-called stillbirths. Bull World Health Organ. 2008;86(4):315-6.
36. Pullum TW. An assessment of age and date reporting in the DHS Surveys 1985-2003. 2006.
37. Liu L, Kalter HD, Chu Y, Kazmi N, Koffi AK, Amouzou A, et al. Understanding Misclassification between Neonatal Deaths and Stillbirths: Empirical Evidence from Malawi. PLoS One. 2016;11(12):e0168743.
38. Espeut D, Becker S. The validity of birth and pregnancy histories in rural Bangladesh. J Health Popul Nutr. 2015;33:17.
39. Alam N, Ali T, Razzaque A, Rahman M, Zahirul Haq M, Saha SK, et al. Health and Demographic Surveillance System (HDSS) in Matlab, Bangladesh. Int J Epidemiol. 2017;46(3):809-16.
40. Nations U. United Nations Department of Economic and Social Affairs/Population Division (2017): World Population Prospects: The 2017 Revision. New York, USA, Division P; 2017.
41. INE. Inquérito aos Indicadores Múltiplos (MICS5) 2014, Relatório Final. Bissau, Guine-Bissau.: Ministério da Economia e Finanças, Direcção Geral do Plano Instituto Nacional de Estatística (INE). 2014.
42. Ursing J, Rombo L, Rodrigues A, Aaby P, Kofoed PE. Malaria transmission in Bissau, Guinea-Bissau between 1995 and 2012: malaria resurgence did not negatively affect mortality. PLoS One. 2014;9(7):e101167.
43. Roth AE, Benn CS, Ravn H, Rodrigues A, Lisse IM, Yazdanbakhsh M, et al. Effect of revaccination with BCG in early childhood on mortality: randomised trial in Guinea-Bissau. BMJ. 2010;340:c671.
44. Rodrigues A, Schellenberg JA, Kofoed PE, Aaby P, Greenwood B. Changing pattern of malaria in Bissau, Guinea Bissau. Trop Med Int Health. 2008;13(3):410-7.
45. Bjerregaard-Andersen M, Lund N, Pinstrup Joergensen AS, Starup Jepsen F, Werner Unger H, Mane M, et al. Stillbirths in urban Guinea-Bissau: A hospital- and community-based study. PLoS One. 2018;13(5):e0197680.
46. Akuze J, Baschieri A, Kerber K, Gordeev V, Waiswa P, Blencowe H, et al. Do different HDSS surveillance systems result in different quality of pregnancy outcome data? 2017.
47. Snow J, Mann M. Qualtrics survey software: handbook for research professionals. Qualtrics Labs, Inc. 2013.
48. Hartung C. Open Data Kit: Technologies for Mobile Data Collection and Deployment Experiences in Developing Regions: University of Washington; 2012.
49. Borriello G, editor Open Data Kit: creating an open source community for mobile data collection. Proceedings of the 3rd ACM international workshop on MobiArch; 2011: ACM.
50. Portnoy A, Jit M, Helleringer S, Verguet S. Impact of measles supplementary immunization activities on reaching children missed by routine programs. Vaccine. 2018;36(1):170-8.
51. Verguet S, Jassat W, Bertram MY, Tollman SM, Murray CJ, Jamison DT, et al. Impact of supplemental immunisation activity (SIA) campaigns on health systems: findings from South Africa. J Epidemiol Community Health. 2013;67(11):947-52.
52. Fisker AB, Biering-Sorensen S, Lund N, Djana Q, Rodrigues A, Martins CL, et al. Contrasting female-male mortality ratios after routine vaccinations with pentavalent vaccine versus measles and yellow fever vaccine. A cohort study from urban Guinea-Bissau. Vaccine. 2016;34(38):4551-7.

53. Fisker AB, Hornshoj L, Rodrigues A, Balde I, Fernandes M, Benn CS, et al. Effects of the introduction of new vaccines in Guinea-Bissau on vaccine coverage, vaccine timeliness, and child survival: an observational study. Lancet Glob Health. 2014;2(8):e478-87.
54. Benn CS, Diness BR, Roth A, Nante E, Fisker AB, Lisse IM, et al. Effect of 50,000 IU vitamin A given with BCG vaccine on mortality in infants in Guinea-Bissau: randomised placebo controlled trial. BMJ. 2008;336(7658):1416-20.
55. Helleringer S, Pison G, Masquelier B, Kante AM, Douillot L, Ndiaye CT, et al. Improving survey data on pregnancy-related deaths in low-and middle-income countries: a validation study in Senegal. Trop Med Int Health. 2015;20(11):1415-23.
56. Helleringer S, Pison G, Masquelier B, Kante AM, Douillot L, Duthe G, et al. Improving the quality of adult mortality data collected in demographic surveys: validation study of a new siblings' survival questionnaire in Niakhar, Senegal. PLoS Med. 2014;11(5):e1001652.

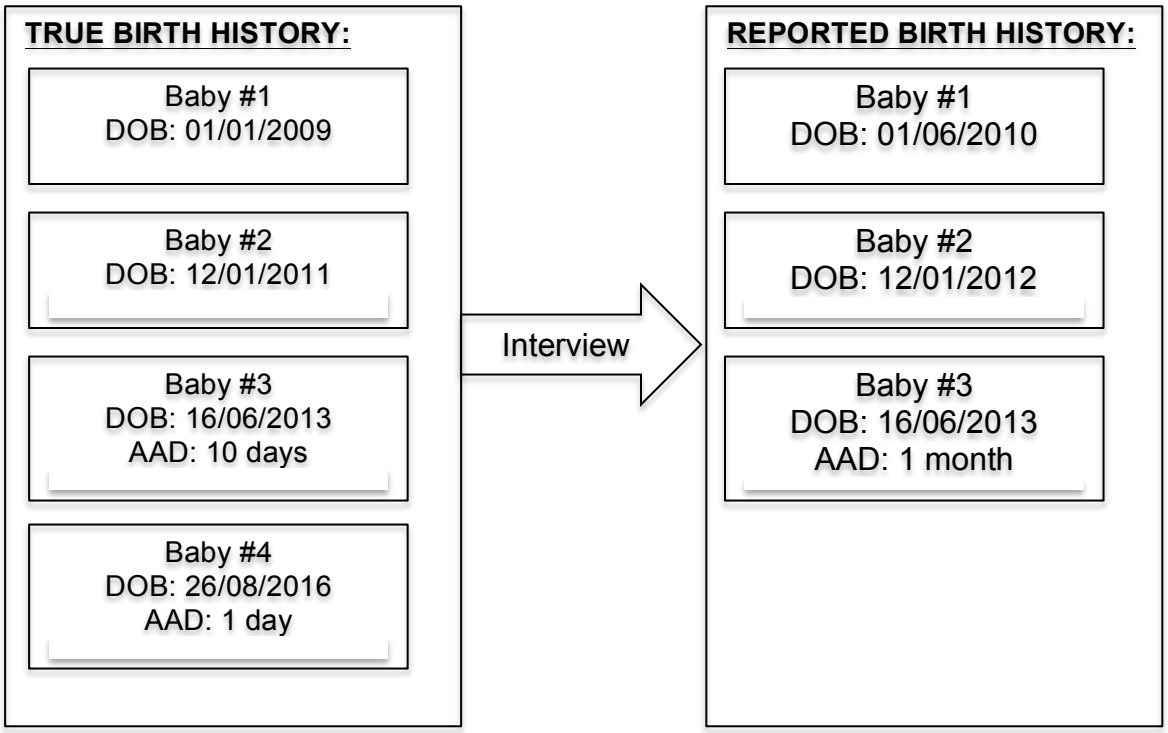
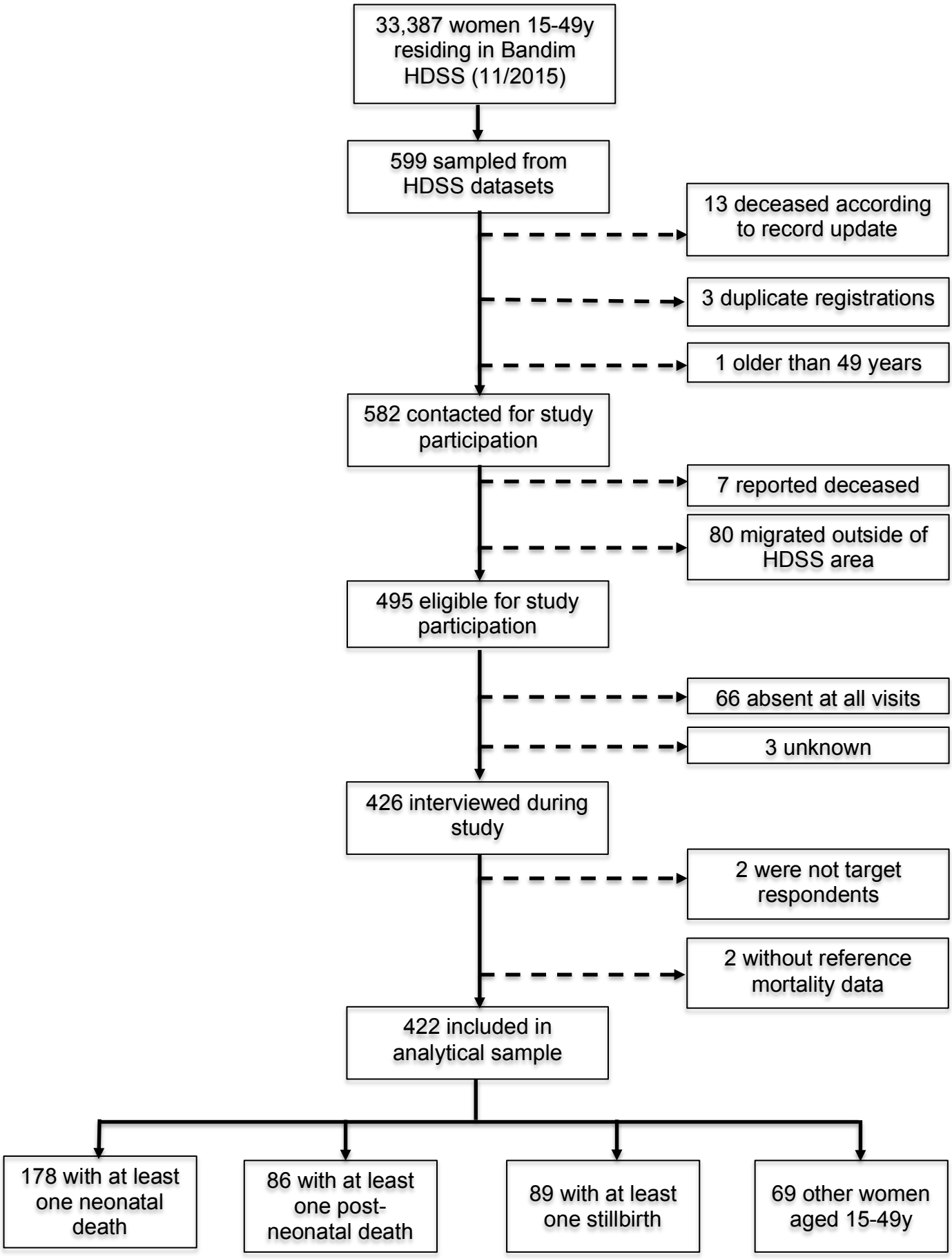


Figure 1: an example of reporting errors in full birth history data

Notes: DOB = Date of birth, AAD = Age at death. In this example, baby #1 and baby #2 were affected by date errors. Baby #3 was affected by age over-statement, whereas baby #4 was an omission.

Figure 2: flow chart of study enrollment



Maternal and child characteristics	N(%)
Maternal variables	
Age group	
15-24y	106 (25.2)
25-34y	226 (53.5)
≥ 35y	90 (21.3)
Schooling	
None	72 (17.0)
Primary	137 (32.5)
Secondary or higher	213 (50.5)
Marital status	
Currently in a union	284 (67.2)
Never in union	118 (28.0)
Previously in union	20 (4.8)
Religion	
Catholic	198 (47.0)
Muslim	135 (32.0)
Protestant	49 (11.5)
Other[a]	40 (9.5)
Ethnicity	
Fula	86 (20.3)
Pepel	116 (27.5)
Other[b]	220 (52.2)
Child variables[c]	
Gender	
Male	261 (61.9)
Female	161 (38.1)
Place of birth	
At home	92 (21.7)
At national hospital	187 (44.5)
At health center	85 (20.3)
Other place	57 (13.5)
Cesarean section	
No	391 (92.6)
Yes	31 (7.4)

Table 1: reported characteristics of participants

Notes: the figures in the table are weighted to account for differences in sampling probabilities and non-response across the sampling strata. Figures in parentheses are column percentages.

[a] Other religions include primarily respondents practicing traditional beliefs and religions.

[b] Fulas and Pepels are the two largest ethnic groups. Other ethnic groups include primarily Manjacos, Mandingas and Balantas.

[c] These variables refer to the child/delivery, which prompted inclusion of the respondent in her sampling stratum. For example, if a respondent had 3 live births over the past 5 years including one that resulted in a neonatal death, and two that are still alive, according to the HDSS, then these variables refer to the neonatal death.

	Neonatal deaths		Post-neonatal deaths	
	Sensitivity	**Specificity**	**Sensitivity**	**Specificity**
Overall	79.1% (72.0% to 84.8%)	99.3% (98.9% to 99.6%)	69.1% (57.7% to 78.6%)	99.9% (99.8 to 100.0%)
Sub-groups				
Respondent w/ stillbirth	--	86.6% (74.8% to 93.4%)	--	--
Respondent w/ post-neonatal death	--	84.0% (73.2% to 90.9%)	--	--
Other respondents	--	100.0% (--)	--	--

Table 2: sensitivity and specificity of FBH data in recording neonatal and post-neonatal deaths

Notes: specificity is the proportion of respondents who did not experience a neonatal death during the reference period according to the HDSS, who also did not report such a death during the FBH survey; sensitivity is the proportion of respondents who experienced a neonatal death according to the HDSS during the reference period who also reported such a death during the survey. All figures are weighted to account for sampling design and non-response.

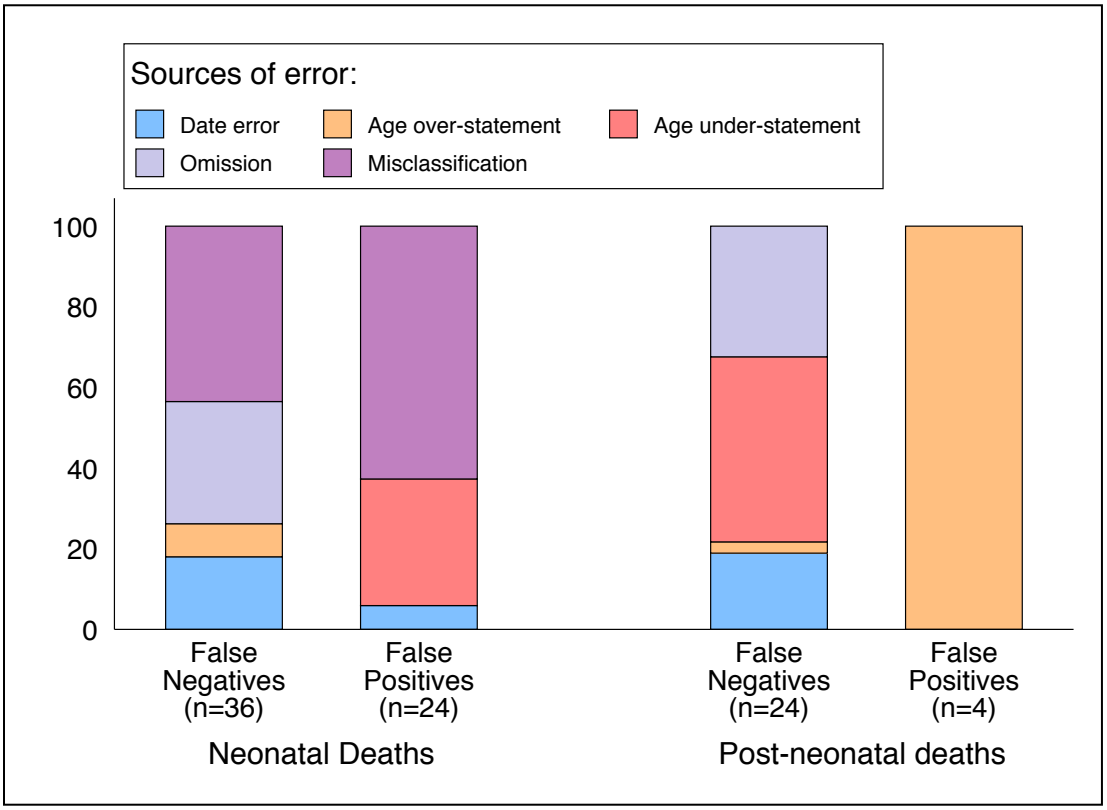


Figure 3: sources of error in FBH data

Notes: Date displacement refers to events reported to have occurred earlier/later than recorded by the HDSS, leading to erroneous exclusion/inclusion from the reference period. Age under-statement refer to errors resulting from the fact that a respondent reported her child's age at death too low compared to the age at death recorded in the HDSS dataset. Age over-statement refers to the opposite situation. Misclassification refers to events that were reported in a manner that does not correspond to HDSS records. For example, a respondent with a neonatal death according to HDSS who did not report such a death during the survey (false negative) might have reported a stillbirth instead. Omission refers to respondents with a neonatal death according to the HDSS who did not report such a death during the survey, and who also did not report a stillbirth nor a post-neonatal death.

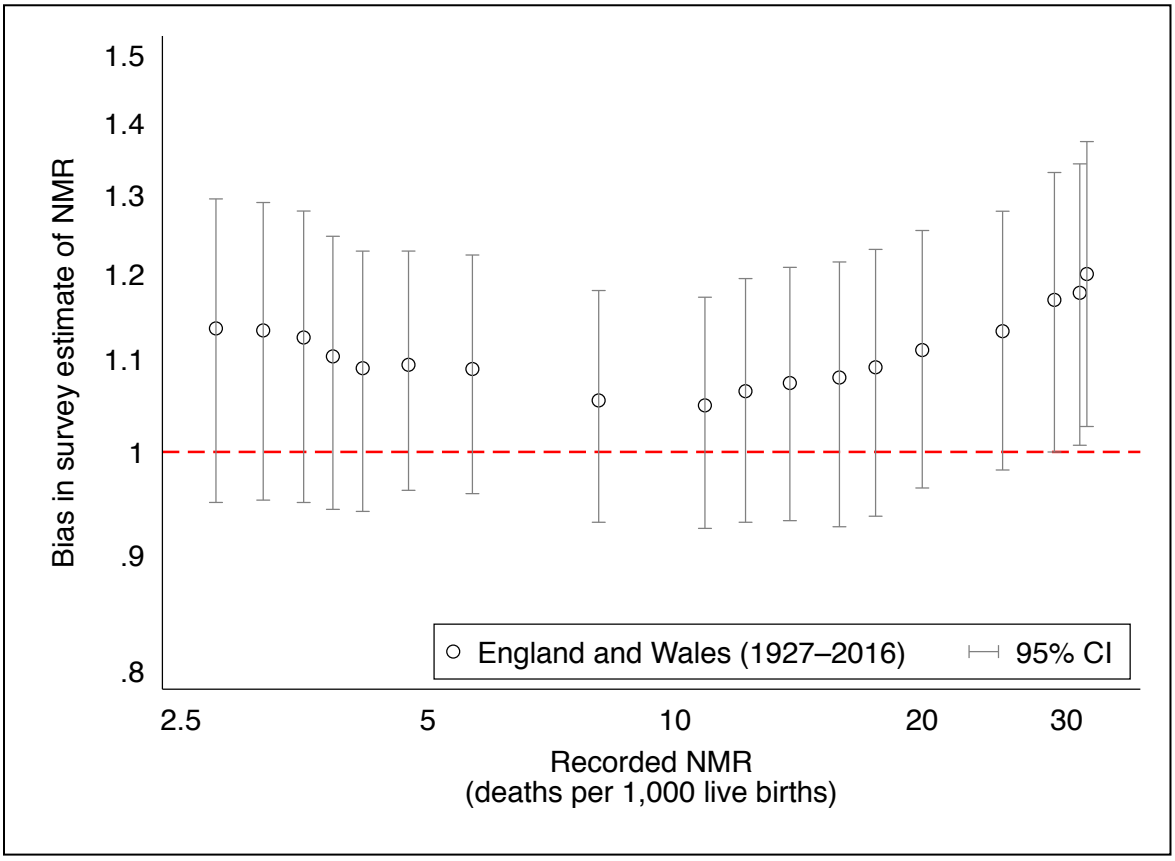


Figure 4: Effects of reporting errors on survey estimates of the neonatal mortality rate

Notes: We used the mathematical model described in appendix A2, along with data from England & Wales (described in appendix A3); to calculate the survey estimates of the NMR that would have been obtained in a population if a FBH survey had been conducted with the reporting parameters observed in Bandim. The bias is obtained by dividing the calculated survey estimate of the NMR by the true level of the NMR in the population of interest. A bias above 1 indicates that FBH data would over-estimate the level of NMR in, whereas a bias under 1 indicates that survey data would under-estimate the level of the NMR. The 95% confidence intervals represented in this figure are obtained from 1,000 draws from the distributions of the reporting parameters. Annual data from England and Wales were aggregated into consecutive 5-year periods (e.g., 1927-1931, 1932-1936) to ensure that our counterfactual survey estimates did not include overlapping sets of events.

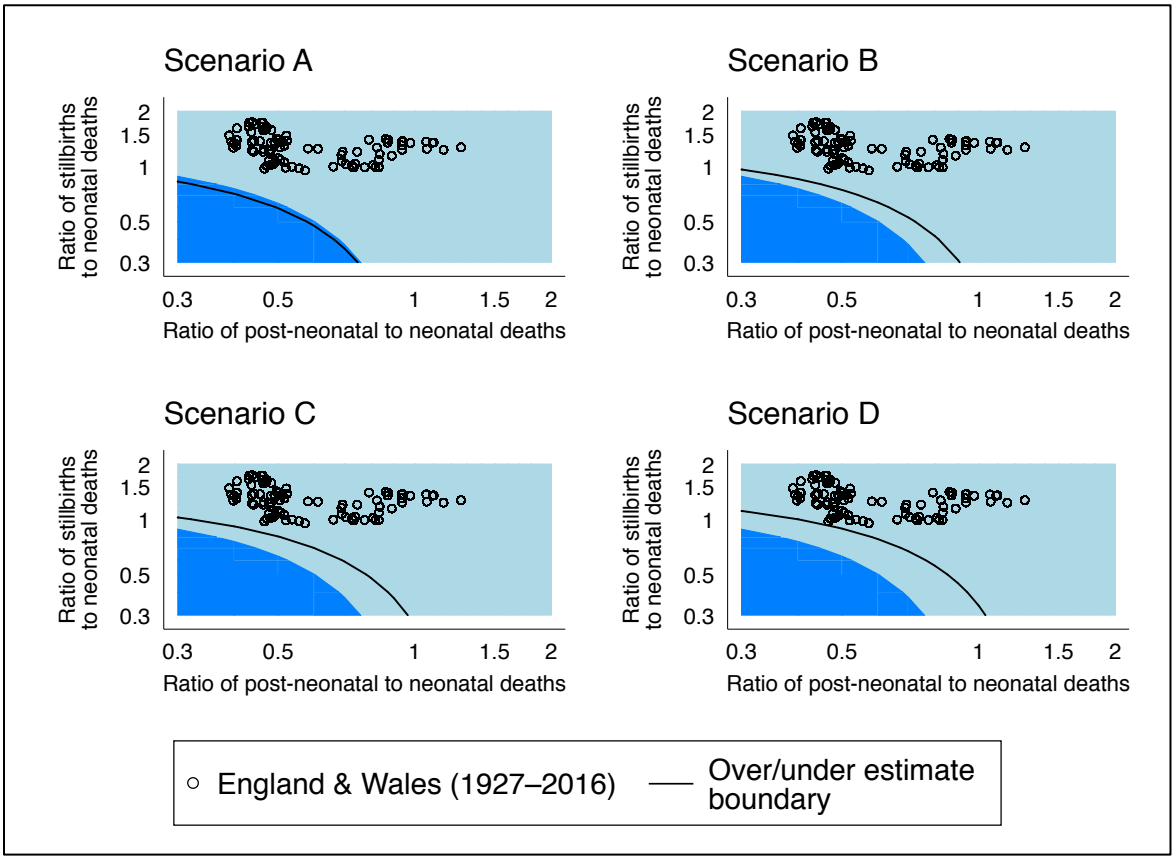


Figure 5: Robustness tests

Notes: in these figures, we recalculated study outcomes (e.g., sensitivity, specificity) after randomly re-classifying and possibly adding a certain number of events in the HDSS datasets according to scenarios defined in appendix A4. In these graphs, the surfaces are colored according to an analysis that takes the HDSS datasets at face value, similar to figure 4. The darker blue areas then represent the combinations of parameters for which a FBH survey would yield an under-estimate of the NMR, whereas the light blue areas represent the combinations of parameters for which a FBH survey would yield an over-estimate of the NMR. At the boundary between the darker blue and light blue areas, the FBH survey yields an unbiased estimate of the NMR. The black contour line on each graph represents the location where the boundary between over and under-estimates would be placed if the HDSS dataset contained the errors assumed in each scenario. For example, in scenario A, the boundary would be shifted downwards slightly, whereas in scenario D, the boundary would be shifted upwards, so that a greater number of parameter combinations would result in under-estimates of the NMR. We replicated each scenario 1,000 times, and in these graphs, the boundary drawn in black represents the 50th percentile of the distribution of these replicates.

	Sensitivity (%)	P-value	uOR of true positive (95% CI)	aOR of true positive (95% CI)
Maternal variables				
Age group		0.900		
15-24y	81.2%		Ref	Ref
25-34y	78.0%		0.82 (0.33 to 2.01)	0.87 (0.33 to 2.32)
≥ 35y	78.3%		0.83 (0.26 to 2.61)	1.12 (0.22 to 5.64)
Schooling		0.068		
None	69.6%		Ref	Ref
Primary	75.9%		1.38 (0.53 to 3.61)	1.95 (0.63 to 6.01)
Secondary or higher	87.0%		2.92 (1.20 to 4.36)	3.58 (1.05 to 12.2)
Religion		0.278		
Catholic	*84.5%*		Ref	Ref
Muslim	*71.8%*		0.47 (0.18 to 1.19)	*0.90 (0.25 to 3.24)*
Protestant	*81.0%*		0.78 (0.22 to 2.75)	*1.06 (0.31 to 3.64)*
Other	*86.1%*		1.14 (0.25 to 5.13)	*3.47 (0.48 to 25.4)*
Marital status		0.271		
Currently in an union	76.3%		Ref	Ref
Never in union	88.2%		2.31 (0.78 to 6.87)	2.27 (0.77 to 6.71)
Previously in union	73.3%		0.85 (0.19 to 3.75)	0.45 (0.07 to 2.94)
Ethnicity		0.527		
Fula	72.2%		Ref	Ref
Pepel	81.9%		1.74 (0.59 to 5.13)	0.69 (0.15 to 3.28)
Other[b]	80.6%		1.60 (0.62 to 4.12)	0.88 (0.26 to 3.02)
Child variables[c]				
Gender		0.426		
Male	81.4%		Ref	Ref
Female	76.1%		0.73 (0.33 to 1.59)	0.55 (0.23 to 1.34)
Place of birth		0.759		
At home	81.6%		Ref	Ref
At hospital	75.7%		0.70 (0.24 to 2.08)	0.43 (0.13 to 1.50)
At health center	81.6%		1.01 (0.28 to 3.59)	0.60 (0.15 to 2.33)
Other place	86.5%		1.45 (0.23 to 9.00)	1.51 (0.13 to 17.9)
Cesarean section		0.549		
No	79.8%		Ref	Ref
Yes	74.1%		0.72 (0.25 to 2.10)	0.74 (0.21 to 2.61)

Table 3: Correlates of the sensitivity of FBH on neonatal deaths (n =174)

Notes: all figures were calculated after accounting for sampling design and non-response by using sampling weights. uOR = unadjusted odds ratios; aOR = adjusted odds ratios. aORs were calculated from a logistic regression that included all covariates appearing in this table.

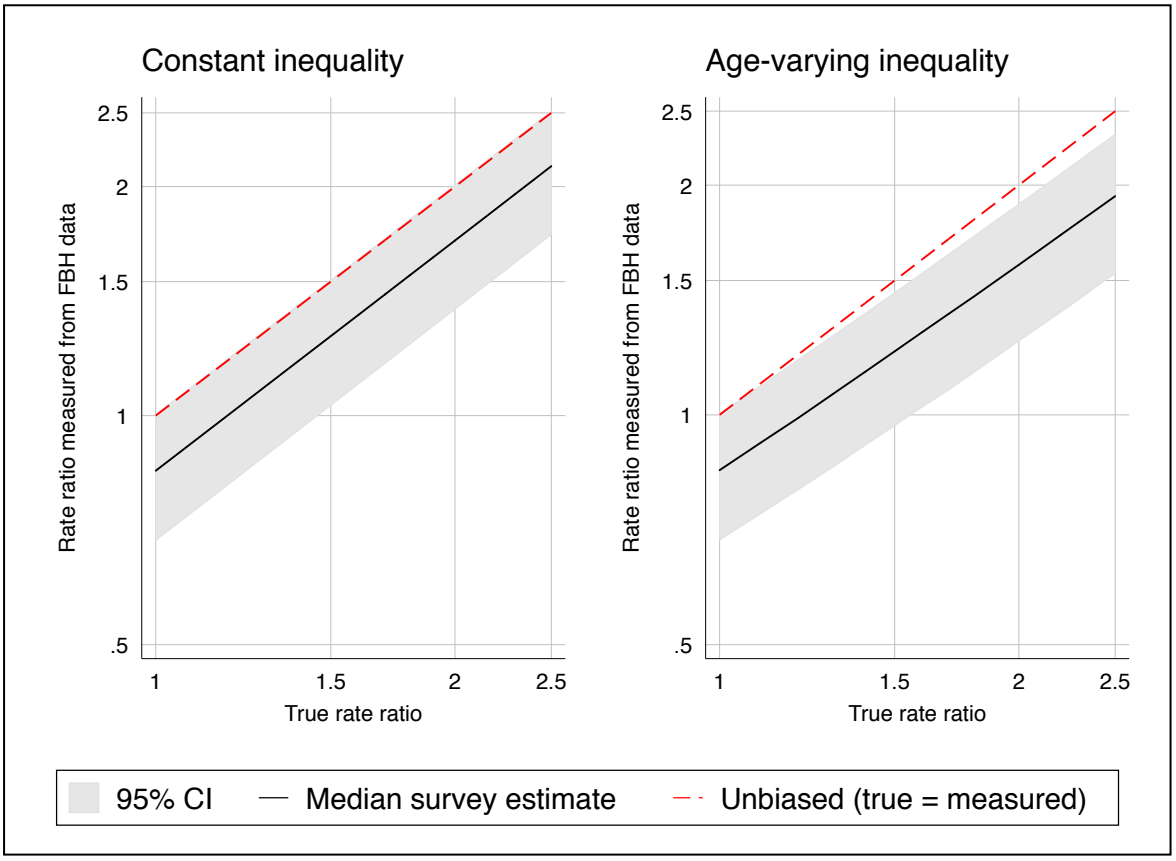


Figure 6: implications of reporting errors for measurement of educational differentials in NMR

Notes: In each panel, the solid red line represents equality between true and estimated rate ratios. Below that line, the survey under-estimates the extent of differences in NMR between educational groups. Above that line, the opposite is true. The left panel represents a hypothetical scenario of "constant inequality", in which the risks of stillbirths, neonatal deaths and post-neonatal deaths are increased by a similar factor among the least educated group. The right panel represents a hypothetical scenario of "age-varying inequality", in which the risks of stillbirths and neonatal deaths are increased by a similar factor among the least educated group, but there are no differences in post-neonatal mortality between the least and most educated groups. The confidence intervals represented in this figure are obtained from 1,000 draws from the distributions of the reporting parameters.