Evaluating pregnancy reporting and childhood mortality estimates in HDSS through record linkage with ANC clinics

Abstract

Health and Demographic Surveillance Systems (HDSS) are important sources of mortality data in settings where civil registration is incomplete. HDSS are powerful when it comes to monitoring the vital status of established residents, but the recording of vital events for newborns is often unreliable. One remedy has been to record pregnancy status reports that cue fieldworkers to follow-up on their outcome, but these reports are often incomplete. In this contribution, we argue that individual-level record linkages between HDSS and antenatal care (ANC) registers can be used to augment HDSS pregnancy reporting, and ultimately produce more accurate estimates of mortality. We use data from the Kenya Medical Research Institute (KEMRI) HDSS in sub-Saharan Africa that have been individually-linked with ANC records to (i) assess the completeness of pregnancy reporting and (ii) evaluate potential bias in estimates of childhood mortality in the HDSS. We discuss the implications of these results for population-based surveillance of maternal and newborn events.

Background

Ending preventable deaths of newborns and children under five years of age by 2030 is a central aim of the international community, and codified in the United Nations' Sustainable Development Goal 3.2 (1). Accurate measurement of under-five mortality is essential to tracking and accelerating progress towards its reduction, but such information is not always available. Sub-Saharan Africa suffers from the highest rates of under-five mortality in the world and a dearth of empirical health and population data. Health and Demographic Surveillance Systems (HDSS) are designed to fill these data gaps, but HDSS estimates of perinatal and early child mortality are often unrealistically low (2–6). Adverse pregnancy outcomes and births followed by early deaths are more likely to be missed by HDSS when they both take place in-between data collection rounds. In this research, we use record linkage with the routine programme data of antenatal clinics to ascertain bias in estimates of pregnancies, pregnancy outcomes, and childhood mortality in an HDSS.

Most HDSS conduct longitudinal surveillance through closely spaced interview rounds of contiguous populations of 30 to 200 thousand. Interviews are usually held with a household representative, the *proxy respondent*, to inquire about vital events for the entire household since the last HDSS interview round (7). Data are collected on pregnancies, births, deaths, and numerous other health and sociodemographic indicators.

The prospective nature of HDSS data collection is highly effective when it comes to tracking the vital status of individuals that have an existing record, and can be followed-up in subsequent interview rounds. However, information on new and transient residents is often poor. This limited scope applies to newborns who are born, and sometimes die, between HDSS data collection rounds. Surviving children are likely to be identified and registered in the HDSS eventually, but those who die may escape surveillance entirely, creating downward bias in estimates of mortality. Many sub-Saharan African HDSS have reported downward bias in estimates of perinatal and early childhood mortality (2–6).

Key to improving HDSS data on newborns is exhaustive capture of pregnancy status information (8). Pregnancy status reports offer HDSS fieldworkers a cue to follow-up on their outcome in the next data collection round. Unfortunately, pregnancy reporting is often incomplete due to a variety of procedural and sociocultural reasons. Pregnancy status reporting deteriorates with less frequent interview rounds and the use of male interviewers (9). Additionally, women may not disclose their pregnancies to the interviewer or even family members so as to avoid gossip, the shame that can accompany giving birth out of wedlock, or stigma associated with pregnancy loss (10,11). In many cases, the proxy respondent may not be aware of the pregnancy status of women in his or her own household (9).

The almost universal coverage of antenatal care (ANC) in many parts of sub-Saharan Africa provides an opportunity to improve HDSS pregnancy data through individual-level record linkage. The World Health Organization recommends that all pregnant women have at least four ANC assessments, with the first visit occurring as early as possible in the pregnancy (12). ANC coverage has expanded dramatically in many parts of sub-Saharan Africa in the last few decades. According to the 2014 Kenya Demographic and Health Survey (DHS), 96% of women giving birth between 2009 and 2014 received ANC from a skilled provider (13). In such settings, routine programme data on pregnancies can be used to augment HDSS pregnancy reporting.

Against this background, this research leverages record-linked ANC data to evaluate pregnancy reporting and estimates of perinatal and early childhood mortality in the Kenya Medical Research Institute (KEMRI) HDSS. The findings will shed light on the completeness of pregnancy reporting, characteristics associated with pregnancy under-reporting, and the potential bias in childhood mortality estimates. This work also demonstrates the potential of using record linkage with routine programme data to augment demographic data in sub-Saharan Africa, and help "break the link between material and data poverty" (14).

Methods

Study setting

The KEMRI HDSS is located to the northeast of Lake Victoria in Siaya County of the Nyanza Province. It covers an area of approximately 700 square kilometres with a population of 220,000 residents in 54,869 households (15). The HDSS was initially set up to collect information on malaria morbidity, mortality, and interventions as part of an insecticide-treated bed net trial. The HDSS has since expanded its mandate, and collects information on a variety of diseases and interventions, in addition to longitudinal demographic and health data (15). In the HDSS area, HIV prevalence was approximately 18% in 2014, and the total fertility rate was above five children per woman (16). ANC coverage in the area is high with approximately 94% of women accessing care at least once during pregnancy (16). Data is collected on births and deaths in household interview rounds that take place every 6 months, and through a parallel continuous village reporter system (15). Information on pregnancy status is only collected in household interviews, and not captured by village reporters.

Data

Since February 2018, the HDSS has been conducting individual-level record linkage with 14 ANC clinics in the Gem District of Siaya County. In an approach termed "point-of-contact interactive record linkage" (PIRL), patients seeking ANC services are linked to their HDSS record during a brief interview with data clerks stationed in clinic waiting rooms (17). After obtaining informed consent, data clerks enter the patient's identifying information into the PIRL software. This information includes up to three names for the individual, date of birth, village, sub-location, location, region, and up to three names of a household member. A probabilistic search algorithm returns potential matches from the HDSS database, and the patient and data clerk consult to select the true match.

As of October 31, 2019, 2,277 ANC patients had been linked to the KEMRI HDSS through PIRL. For each patient's pregnancy, data from the ANC register was digitized for gestational age at time of visit, last menstrual period (LMP), estimated date of delivery (EDD), gravidity, parity, HIV and treatment status, and dates of clinic visits. The LMP and EDD were used to define the estimated start and end of each pregnancy episode that was linked to an individual in the HDSS.

In order to evaluate pregnancy reporting completeness in the HDSS, it was stipulated that an individual's pregnancy episode must have ended at least 6 months prior to the latest available HDSS data. This allowed time for a pregnancy outcome to be reported in the HDSS following the completion of the pregnancy episode. The most recent HDSS data included in the analysis was collected on October 31, 2019. As such, pregnancies were eligible for the inclusion in the analysis if the EDD took place prior to April 30, 2019. There were 1,266 pregnancies which met this criterion.

Socio-demographic and clinical characteristics of the 1,266 individuals with pregnancies included in the analysis are presented in Table 1. The majority of individuals were aged between 20 and 30. Information on level of education was missing for approximately half of the individuals. Among the other half, most had primary-level education. Approximately 19% of individuals were HIV positive, and 76% were married. This was the first pregnancy for 22% of the individuals, and 47% made their first visit to the ANC clinic in the second trimester of their pregnancy. 13% of individuals were relatively new residents in the HDSS, having only resided in the HDSS area since 2018.

Additionally, it is important to take note of the variable "Months since EDD," which indicates the number of months that have elapsed between the individual's EDD and the cut-off date for inclusion in the study. It will be important to control for this variable when evaluating chracteristics that are associated with pregnancy outcome reporting, as pregnancies which occurred further in the past will have had more time to be reported in the HDSS. For instance, individuals with an EDD that was 0-3 months prior to April 30th, 2019 only had 6-9 months to report the pregnancy outcome in the HDSS prior to October 31, 2019. Whereas individuals with an EDD that was 9-15 months prior to April 30th, 2019 will have had approximately 15-21 months to report the pregnancy outcome in the HDSS.

Table 1: Socio-demographic and clinical characteristics of ANC patients that were linked to the HDSS. "." indicates a missing value.

Variable	Value	N	%
Age	9-14	11	0.87
	15-19	216	17.06
	20-24	377	29.78
	25-29	309	24.41
	30-34	210	16.59
	35-39	119	9.40
	40+	24	1.90
Education level	None	4	0.32
	Primary	392	30.96
	Secondary	201	15.88
	Tertiary	22	1.74
	.	647	51.11
Gestational age	1st trimester	131	10.35
	2nd trimester	595	47.00
	3rd trimester	509	40.21
	.	31	2.45
Gravidity	1	277	21.88
	2	241	19.04
	3	238	18.80
	4	220	17.38
	5+	290	22.91
HIV status	Negative	983	77.65
	Positive	245	19.35
	.	38	3.00
Marital status	Married	960	75.83
	Sep./Div./Wid.	38	3.00
	Single	266	21.01
	.	2	0.16
Months since EDD	(0,3]	268	21.17
	(3,6]	375	29.62
	(6,9]	362	28.59
	(9,15]	261	20.62
Parity	0	277	21.88
	1	242	19.12
	2	238	18.80
	3	219	17.30
	4	163	12.88
	5+	126	9.95
	.	1	0.08
Residency start	before 2000	146	11.53
	2000-2004	163	12.88
	2005-2009	170	13.43
	2010-2014	291	22.99
	2015-2017	332	26.22
	2018-present	164	12.95
Total		1266	100

Analysis

Pregnancy episodes that were inferred from the ANC data were crosstabulated with HDSS records to evaluate pregnancy status and outcome reporting. A pregnancy status was considered captured by the HDSS if it was reported any time during the duration of the pregnancy episode, and up to one month following its completion. This reference period extended one month beyond the EDD in order to account for late-term pregnancies and inaccuracies in the estimation of the EDD.

The HDSS reference period for pregnancy outcome reporting was wider, as there are more likely to be inaccuracies in reported dates of pregnancy outcomes. The date of a pregnancy status report in the HDSS is simply the date that the household interview took place. However, information on pregnancy outcomes is provided by village reporters, or household interviews which often take place long after the pregnancy has come to term. Comparatively, these reporting mechanisms are more prone to rounding errors and recall bias. Therefore, a pregnancy outcome report was considered captured by the HDSS if it was reported within 6 months of the EDD.

In bivariate analyses, chi-square significance tests were used to identify characteristics associated with under-reporting of pregnancy statuses and outcomes in the HDSS. Explanatory variables of interest included age, HIV status, parity, marital status, education level, duration of residence in HDSS, and months since EDD. Variables that were significantly associated with under-reporting of pregnancy statuses and outcomes were further analysed through multivariable logistic regression. In a stepwise backwards elimination approach, variables were removed from the model if they did not contribute significantly to model fit, as measured by AIC.

Un-reported pregnancies identified in the first part of the analysis were used to conduct a sensitivity analysis on HDSS estimates of mortality under five. Estimates of neonatal, infant, and under-five mortality in the HDSS were calculated for 2018-2019. This is the approximate period for which record-linked ANC data was available. HDSS individual-level data with exact dates of births, deaths, and censoring was used to assign events and exposure time to age-groups defined by weeks for the first month of life, months for the first year, trimesters for the second year, and years until exact age 5. Standard demographic rate calculations of events over exposure time were used to estimate age-specific mortality rates, which were transformed into cumulative probabilities of dying. Neonatal, infant, and under-five mortality were estimated using the cumulative probabilities of dying in the first 28 days, year, and five years of life, respectively.

It was assumed that increasing proportions of the un-reported pregnancies resulted in the death of a child, and estimates of neonatal, infant, and under-five mortality were recalculated accordingly. In cases where the un-reported pregnancy was assumed to have ended in a death, this event was considered to have taken place at the mid-point of the age interval for neonates, and midway through the first year of life for infants and children under five. In cases where the missing pregnancy did not result in a death, this individual contributed 28 days of exposure time to the calculation of the neonatal mortality rate, and one year of exposure time to the calculations for infant and under-five mortality. These augmented estimates of neonatal, infant, and under-five mortality were compared to standard HDSS estimates for the period 2018-2019.

To further investigate the impact of pregnancy reporting on under-five mortality estimation, HDSS mortality estimates were then calculated for the subset of individuals that were observed prior to birth in pregnancy status reports. These were compared to estimates calculated using data from all other residents (who were not observed prior to birth in pregnancy status reports). Individuals were considered to have been observed as a pregnancy status if their mother reported a pregnancy status within the eight months preceding their birth. Estimates of mortality under five from these two groups were compared for two-year periods from 2010 to 2018.

Results

It was possible to assess HDSS pregnancy status and outcome reporting for 1,266 pregnancies. Approximately 46% of these pregnancies were captured in an HDSS pregnancy status report, and 44% were captured in a pregnancy outcome report. There were 400 (32%) un-reported pregnancies which were absent from both pregnancy status and outcome reporting.

Table 2: Cross-tabulation of pregnancy status and outcome reporting for individuals with linked records in the HDSS.

	Pregnancy outcome		
	Yes	No	Total
Pregnancy status			
Yes	271	309	580 (46%)
No	286	400	686 (54%)
Total	557 (44%)	709 (56%)	1266 (100%)

Table 3 summarises the characteristics of individuals by presence or absence of pregnancy status and outcome reports in the HDSS. Chi-square tests identified the variables of age, gravidity, parity, and residency start as significantly associated with both pregnancy status and pregnancy outcome reporting. Marital status was significantly associated with only pregnancy status reporting. The gestational age at an individual's first ANC clinic visit and months elapsed since their EDD were only associated with pregnancy outcome reporting.

The first and final logistic regression models for pregnancy status reporting and pregnancy outcome reporting are shown in Table 4. Individuals who were married or formerly married (i.e. separated, divorced, or widowed) were more likely to have pregnancy status reports in the HDSS than individuals who identified as single. Individuals with higher gravidity were also more likely to have their pregnancy status captured in the HDSS, while older individuals were less likely. For pregnancy outcome reporting, individuals whose EDD occurred earlier in the period were more likely to have reported outcomes in the HDSS.

Table 3: Socio-demographic and clinical characteristics of record-linked ANC patients by pregnancy status and pregnancy outcome report in the HDSS.

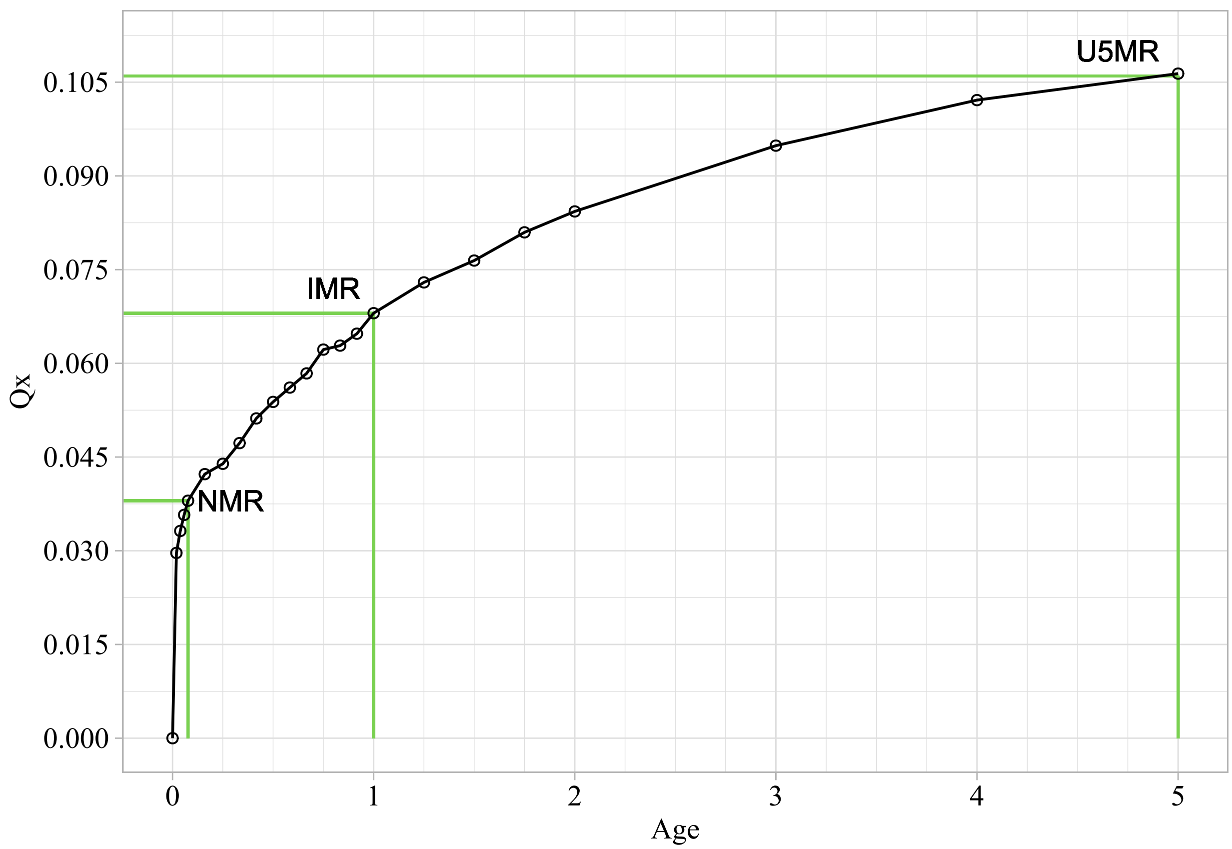
		Pregnancy status report			Pregnancy outcome report		
Variable	Value	Yes (%)	No (%)	P-value	Yes (%)	No (%)	P-value
Age	9-14	4 (36.36)	7 (63.64)	0.002	4 (36.36)	7 (63.64)	0.017
	15-19	73 (33.8)	143 (66.2)		79 (36.57)	137 (63.43)	
	20-24	186 (49.34)	191 (50.66)		172 (45.62)	205 (54.38)	
	25-29	156 (50.49)	153 (49.51)		139 (44.98)	170 (55.02)	
	30-34	99 (47.14)	111 (52.86)		101 (48.1)	109 (51.9)	
	35-39	55 (46.22)	64 (53.78)		58 (48.74)	61 (51.26)	
	40+	7 (29.17)	17 (70.83)		4 (16.67)	20 (83.33)	
Education level	None	0 (0)	4 (100)	0.218	2 (50)	2 (50)	0.665
	Primary	81 (20.66)	311 (79.34)		173 (44.13)	219 (55.87)	
	Secondary	39 (19.4)	162 (80.6)		99 (49.25)	102 (50.75)	
	Tertiary	1 (4.55)	21 (95.45)		11 (50)	11 (50)	
	.	459 (70.94)	188 (29.06)		272 (42.04)	375 (57.96)	
Gestational age	1st trimester	60 (45.8)	71 (54.2)	0.919	37 (28.24)	94 (71.76)	<0.001
	2nd trimester	275 (46.22)	320 (53.78)		206 (34.62)	389 (65.38)	
	3rd trimester	229 (44.99)	280 (55.01)		293 (57.56)	216 (42.44)	
	.	16 (51.61)	15 (48.39)		21 (67.74)	10 (32.26)	
Gravidity	1	85 (30.69)	192 (69.31)	<0.001	114 (41.16)	163 (58.84)	0.014
	2	104 (43.15)	137 (56.85)		95 (39.42)	146 (60.58)	
	3	135 (56.72)	103 (43.28)		102 (42.86)	136 (57.14)	
	4	114 (51.82)	106 (48.18)		93 (42.27)	127 (57.73)	
	5+	142 (48.97)	148 (51.03)		153 (52.76)	137 (47.24)	
HIV status	Negative	454 (46.19)	529 (53.81)	0.798	431 (43.85)	552 (56.15)	0.733
	Positive	116 (47.35)	129 (52.65)		111 (45.31)	134 (54.69)	
	.	10 (26.32)	28 (73.68)		15 (39.47)	23 (60.53)	
Marital status	Married	480 (50)	480 (50)	<0.001	436 (45.42)	524 (54.58)	0.080
	Sep./Div./Wid.	19 (50)	19 (50)		11 (28.95)	27 (71.05)	
	Single	80 (30.08)	186 (69.92)		110 (41.35)	156 (58.65)	
	.	1 (50)	1 (50)		0 (0)	2 (100)	
Months since EDD	(0,3]	114 (42.54)	154 (57.46)	0.110	2 (0.75)	266 (99.25)	<0.001
	(3,6]	183 (48.8)	192 (51.2)		84 (22.4)	291 (77.6)	
	(6,9]	153 (42.27)	209 (57.73)		250 (69.06)	112 (30.94)	
	(9,15]	130 (49.81)	131 (50.19)		221 (84.67)	40 (15.33)	
Parity	0	85 (30.69)	192 (69.31)	<0.001	113 (40.79)	164 (59.21)	0.029
	1	104 (42.98)	138 (57.02)		97 (40.08)	145 (59.92)	
	2	137 (57.56)	101 (42.44)		101 (42.44)	137 (57.56)	
	3	112 (51.14)	107 (48.86)		93 (42.47)	126 (57.53)	
	4	78 (47.85)	85 (52.15)		89 (54.6)	74 (45.4)	
	5+	63 (50)	63 (50)		63 (50)	63 (50)	
	.	1 (100)	0 (0)		1 (100)	0 (0)	
Residency start	before 2000	41 (28.08)	105 (71.92)	<0.001	60 (41.1)	86 (58.9)	<0.001
	2000-2004	61 (37.42)	102 (62.58)		75 (46.01)	88 (53.99)	
	2005-2009	93 (54.71)	77 (45.29)		90 (52.94)	80 (47.06)	
	2010-2014	152 (52.23)	139 (47.77)		125 (42.96)	166 (57.04)	
	2015-2017	176 (53.01)	156 (46.99)		164 (49.4)	168 (50.6)	
	2018-present	57 (34.76)	107 (65.24)		43 (26.22)	121 (73.78)	
Total		580 (45.80)	686 (54.20)		557 (44.00)	709 (56.00)	

Table 4: Logistic regression results for pregnancy status reporting and pregnancy outcome reporting.

	Dependent variable:			
	Pregnancy status report		Pregnancy outcome report	
	(1)	(2)	(3)	(4)
Age (ref. = 25-29)				
9-19	0.079	0.127	−0.255	
	(−0.387, 0.546)	(−0.334, 0.589)	(−0.798, 0.288)	
20-24	0.220	0.245	0.094	
	(−0.109, 0.549)	(−0.082, 0.572)	(−0.344, 0.532)	
30-34	−0.258	−0.274	−0.021	
	(−0.623, 0.107)	(−0.638, 0.090)	(−0.516, 0.475)	
35+	−0.572**	−0.598***	−0.193	
	(−1.020, −0.122)	(−1.040, −0.154)	(−0.818, 0.432)	
Gestational age			0.009	
			(−0.010, 0.029)	
Gravidity	0.339	0.145***	0.168	
	(−0.309, 0.987)	(0.049, 0.241)	(−0.646, 0.983)	
Marital status (ref. = Single)				
Married	0.578***	0.706***		
	(0.154, 1.000)	(0.326, 1.090)		
Sep./Div./Wid.	0.708*	0.795**		
	(−0.046, 1.460)	(0.053, 1.540)		
Parity	−0.189		−0.087	
	(−0.835, 0.457)		(−0.896, 0.722)	
Residency start	0.012		−0.014	
	(−0.005, 0.030)		(−0.034, 0.007)	
Months since EDD	0.00004		0.020***	0.020***
	(−0.001, 0.001)		(0.018, 0.022)	(0.018, 0.022)
Constant	−25.700	−1.180***	23.200	−4.080***
	(−60.900, 9.480)	(−1.680, −0.692)	(−17.900, 64.300)	(−4.520, −3.630)
Observations	1,263	1,264	1,234	1,266
Log Likelihood	−845.000	−847.000	−531.000	−548.000
Akaike Inf. Crit.	1,713.000	1,711.000	1,082.000	1,100.000
Note:				*p<0.1; **p<0.05; ***p<0.01

HDSS estimates of the probability of dying for detailed ages within the first five years of life were calculated for the period 2018-2019. The probability of dying within the first 28 days of life was estimated as 0.038. This can be understood as a neonatal mortality rate of 38 deaths per 1,000 live births. The infant and under-five mortality rates were 68 and 106 deaths per 1,000 live births, respectively.

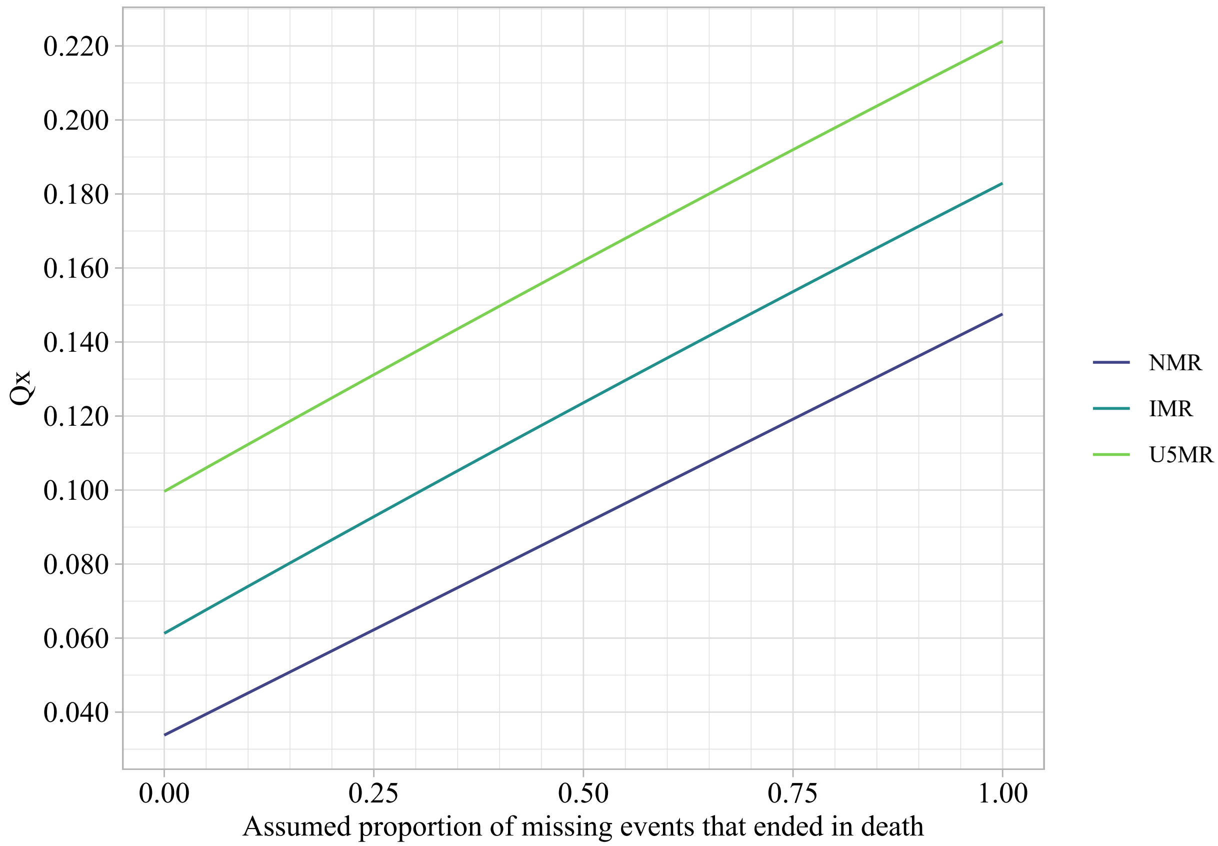
Figure 1: The cumulative probability of dying under age five in the HDSS for the period 2018-2019. The levels of neonatal mortality (NMR), infant mortality (IMR), and under-five mortality (U5MR) are indicated.



HDSS estimates of mortality under five were subject to a sensitivity analysis using information from missing pregnancies. The results of the sensitivity analysis are presented in Figure 2. There were 400 pregnancies in the ANC data that were neither reported as a pregnancy status nor outcome in the HDSS. If none of these pregnancies resulted in the death of a child, the neonatal mortality rate was estimated as 34 deaths per 1,000 live births. This is slightly lower than the standard HDSS estimate of 38, given the addition of these pregnancies to the denominator of the mortality rate. The infant mortality rate and under-five mortality rate were also lower than standard HDSS estimates at 61 and 100 deaths per 1,000 live births, respectively.

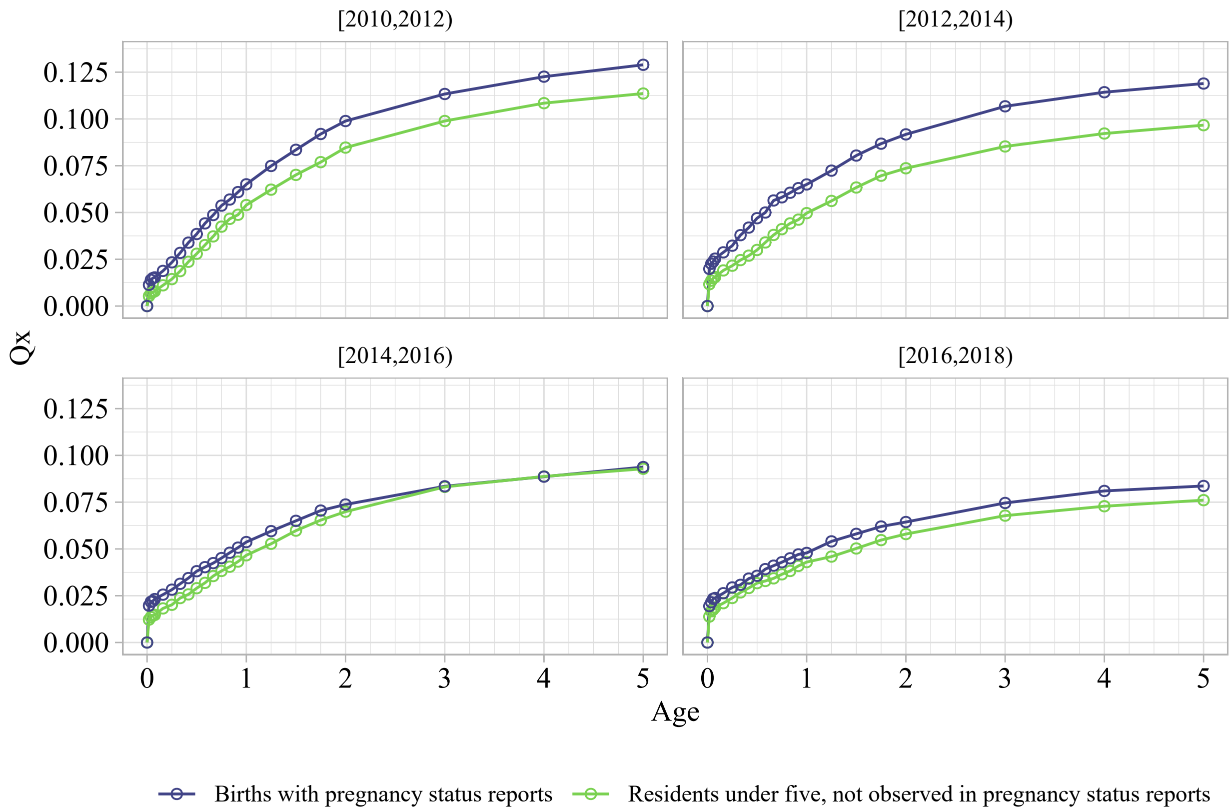
Under the assumption that some of these missing pregnancies ended in the death of a child, the mortality estimates began to increase. If 50% of unreported pregnancies ended in death, neonatal mortality would increase to 91, infant mortality to 124, and under-five mortality to 162. If all of the missing pregnancies resulted in the death of a child, neonatal, infant, and under-five mortality would further increase to 148, 183, and 221 deaths per 1,000 live births, respectively.

Figure 2: Sensitivity analysis of HDSS estimates of neonatal, infant, and under-five mortality for the period 2018-2019. Increasing proportions of the 400 un-reported pregnancies (identified through record linkage with ANC clinics) were assumed to result in the death of a child, and mortality estimates were re-calculated accordingly.



The next part of the analysis further investigated the impact of pregnancy status reporting on mortality estimation in the HDSS. HDSS estimates of mortality under five were calculated for births that were observed in pregnancy status reports, and compared to estimates for all other residents. The estimates of mortality under five for these two groups for two-year periods from 2010 to 2018 are presented in Figure 3. Across all periods, the estimates of mortality under age two are higher for births that were observed in pregnancy status reports. Estimates calculated using data from residents that were not observed in pregnancy status reports yields lower levels of mortality for these ages. The largest disparity between estimates of overall under-five mortality for the two groups is found in the period from 2012 to 2014, where the values differ by more than 20 deaths per 1,000 live births. The estimates of under-five mortality for each group are more similar for recent periods.

Figure 3: HDSS estimates of mortality under five for births with pregnancy status reports, and individuals that were not observed prior to birth in pregnancy status reports.



Discussion

There were 1,266 pregnancies in ANC registers that were linked to individuals in the Kisumu HDSS. Of these 1,266, 557 (44%) had pregnancy outcomes reported in the HDSS, and 709 (56%) did not. Among the 709 without reported pregnancy outcomes, 309 had pregnancy status reports in the HDSS. The HDSS will attempt to follow-up on these pregnancy status reports in future interview rounds, and it is likely that the pregnancy outcome reporting percentage will increase from 44%. However, the more concerning number are the 400 pregnancies that had neither pregnancy status nor outcome reports in the HDSS. Given that these pregnancies were missed by HDSS surveillance while they were in progress, fieldworkers will not be prompted to inquire about their outcome in subsequent interview rounds. If these pregnancies resulted in live births, the children will likely be registered in the HDSS at some point in the future. However, if these pregnancies resulted in adverse outcomes or the death of a child, such events may never be captured by the HDSS.

Marital status was found to be a strong predictor of pregnancy status reporting in the logistic regression models. Pregnancies to married or formerly married individuals had increased odds of being captured by the HDSS, compared to the pregnancies of individuals who were single. This is likely influenced by the HDSS's reliance on proxy respondents who report events on behalf of the entire household. The proxy respondent is typically the head of the household. If a woman is married, she likely resides in a household where her husband is the proxy respondent. While women in rural sub-Saharan Africa will often conceal their pregnancy from the broader community, disclosing a pregnancy to a sexual partner is done soon after the first missed menses (10). This is seen as necessary in order to confirm the partner's paternity status, and secure his support in preparing for the pregnancy (10). Thus, the proxy respondent is much more likely to be aware of the pregnancy status of his wife, as opposed to another individual in the household.

Women aged 35 and over were found to have reduced odds of pregnancy status reporting when compared to the reference group of women aged 25-29. In qualitative research in eastern Uganda, interview subjects identified the advanced age of the mother as a contributing cause of pregnancy loss and stillbirth (11). If such is the general perception among women in Kisumu, it is possible that older women go to greater lengths to conceal their pregnancy until they are certain it will be carried to term. This may be done to avoid stigma or being suspected of induced abortion if the pregnancy is lost. The motivations for pregnancy concealment reflect the high level of social risk faced by women in such settings (10).

For pregnancy outcome reporting, the final logistic regression model only contained the covariate for months since EDD. The odds of having a pregnancy outcome report in the HDSS were higher for individuals whose pregnancies occurred longer ago. This provides more time for the pregnancy outcome to be captured by a village reporter or household interview, and subsequently entered into the HDSS database. Among pregnancies with at least 9-15 months between the EDD and April 30, 2019, approximately 85% had pregnancy outcome reports in the HDSS. This proportion drops to 69% for pregnancies with a 6-8 months following the EDD, and 22% for 3-5 months. The significance of this relationship indicates a need to revisit the analysis at a later date. Once all pregnancy outcomes have had at least 9 months to be reported in the HDSS, it is likely that other important predictors of pregnancy outcome reporting will be identified.

The sensitivity analysis on childhood mortality estimates in the HDSS dimensioned the potential bias that can arise from pregnancy under-reporting. It was assumed that an increasing proportion of the un-reported pregnancies resulted in the death of a child, and indicators of under-five mortality were calculated accordingly. Instances of mortality among the un-reported pregnancies had a large impact on the mortality estimates for the one-year period from 2018 to 2019. This is partly due to the relative rarity of mortality as an event, and the short length of the period.

Nevertheless, the potential for downward bias in HDSS estimates of mortality from missing pregnancy reports is concerning. Mortality among un-reported pregnancies is likely higher than the general population. Children that died at an early age are more likely to be omitted from reports of births than children who survived (18). The regression analysis also found higher levels of pregnancy under-reporting for single women. As children born to single mothers are at higher risk for under-five mortality, this suggests even greater potential for downward bias in mortality estimates (19). Furthermore, it is important to note that the record linkage with ANC clinics in Kisumu HDSS was only conducted in the Gem subcounty. There are two other subcounties in the HDSS for which linked data was not available. Thus, the missing pregnancies that were identified through record linkage are only a fraction of the true total that are missing from the HDSS.

These findings all suggest that under-five mortality is likely being under-estimated in the HDSS. This is supported by the comparison between estimates of under-five mortality for births with pregnancy status reports in the HDSS and estimates for all other residents. Levels of mortality were higher across almost all ages under five for the births with pregnancy status reports in the HDSS. In many ways, pregnancy status reports indicate the start of an individual's time under observation in the HDSS. This allows them to be followed-up by HDSS data collectors in the same way as existing residents, and instances of mortality as less likely to be missed.

HDSS are valuable sources of empirical data for much of sub-Saharan Africa, however, information on newborns and young children is often unreliable. This is problematic, as accurate measurement of the levels, trends, and age patterns of under-five mortality is essential to tracking and accelerating progress towards its reduction. Improving HDSS estimates of childhood mortality depends on the exhaustive capture of pregnancy status and outcome information. This research demonstrates the potential of using record linkage with ANC clinics to evaluate pregnancy reporting completeness in HDSS and investigate bias in estimates of childhood mortality. The analysis will be revisited as more record-linked data becomes available. In future work record-linked data will also be leveraged to organize follow-up data collection on missing pregnancy outcomes in the HDSS.

Record linkage between HDSS and routine programme data is an efficient and cost-effective manner of augmenting population health information. In sub-Saharan Africa, record linkage can greatly expand the uses of existing data, avoiding the high costs that come with designing and implementing new data collection systems (20). Such efforts have the potential to both improve our understanding of population health and our ability to accurately measure it.

References

1. Rosa W, editor. Transforming Our World: The 2030 Agenda for Sustainable Development. In: A New Era in Global Health. New York, NY: Springer Publishing Company; 2017.

2. Assefa N, Oljira L, Baraki N, Demena M, Zelalem D, Ashenafi W, et al. HDSS Profile: The Kersa Health and Demographic Surveillance System. International Journal of Epidemiology. 2016 Feb;45(1):94–101.

3. INDEPTH N. Iganga-Mayuge Health and Demographic Surveillance Site. 2013.

4. Kahn K, Collinson MA, Gomez-Olive FX, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt Health and Socio-demographic Surveillance System. International Journal of Epidemiology. 2012 Aug;41(4):988–1001.

5. Kishamawe C, Isingo R, Mtenga B, Zaba B, Todd J, Clark B, et al. Health & Demographic Surveillance System Profile: The Magu Health and Demographic Surveillance System (Magu HDSS). International Journal of Epidemiology. 2015 Dec;44(6):1851–61.

6. Rossier C, Soura A, Baya B, Compaore G, Dabire B, Dos Santos S, et al. Profile: The Ouagadougou Health and Demographic Surveillance System. International Journal of Epidemiology. 2012 Jun;41(3):658–66.

7. INDEPTH N. Delivering better health information, Driving better health policy: The past, the present, & the future. 2014.

8. Waiswa P, Moyer C, Kwesiga D, Arthur S, Sankoh O, Welaga P, et al. Status of birth and pregnancy outcome capture in Health Demographic Surveillance Sites in 13 countries. International Journal of Public Health. 2019 Jul;64(6):909–20.

9. Kadobera D, Waiswa P, Peterson S, Blencowe H, Lawn J, Kerber K, et al. 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. 2017 Jan;10(1):1356641.

10. 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. Social Science & Medicine. 2010 Nov;71(10):1764–72.

11. Kiguli J, Namusoko S, Kerber K, Peterson S, Waiswa P. Weeping in silence: Community experiences of stillbirths in rural eastern Uganda. Global Health Action. 2015 Dec;8(1):24011.

12. Organization WH. Provision of effective antenatal care: Integrated management of pregnancy and childbirth. 2006.

13. DHS, Kenya National Bureau of Statistics, Ministry of Health, Kenya, Council NAC, Institute KMR, National Council for Population and Development, et al. Kenya Demographic and Health Survey 2014. 2015.

14. Sankoh O, Byass P. The INDEPTH Network: Filling vital gaps in global epidemiology. International Journal of Epidemiology. 2012 Jun;41(3):579–88.

15. Odhiambo FO, Laserson KF, Sewe M, Hamel MJ, Feikin DR, Adazu K, et al. Profile: The KEMRI/CDC Health and Demographic Surveillance System–Western Kenya. International Journal of Epidemiology. 2012 Aug;41(4):977–87.

16. Kohler PK, Okanda J, Kinuthia J, Mills LA, Olilo G, Odhiambo F, et al. Community-Based Evaluation of PMTCT Uptake in Nyanza Province, Kenya. PLOS ONE. 9(10):e110110.

17. Rentsch CT, Kabudula CW, Catlett J, Beckles D, Machemba R, Mtenga B, et al. Point-of-contact Interactive Record Linkage (PIRL): A software tool to prospectively link demographic surveillance and health facility data. Gates Open Research. 2017 Nov;1:8.

18. Pullum TW, Schoumaker B, Becker S, Bradley SEK. An Assessment of DHS Estimates of Fertility and Under-Five Mortality. 2013 International Population Conference of the International Union for the Scientific Study of Population (IUSSP). 2013;34.

19. Clark S, Hamplová D. Single Motherhood and Child Mortality in Sub-Saharan Africa: A Life Course Perspective. Demography. 2013 Oct;50(5):1521–49.

20. Harron K. Introduction to Data Linkage. Administrative Data Research Network. 2016;38.