Validating Indicators of the Quality of Maternal Health Care in Kenya

Authors: Ann Blanc, Charlotte Warren, Katharine McCarthy, James Kimani, Jackline Kivunaga, Brian Mdawida, Charity Ndwiga, Saumya RamaRao

Objective: To validate a set of maternal and newborn health indicators that can be measured in facility and population-based surveys.

Methods: We compared women's self-reports of care received during labor and delivery against direct observation by a third party in two Kenyan hospitals (n=666). Self-reported data was collected via interviews prior to hospital discharge. The area under the receiver operating curve (AUC) and inflation factor (IF) was calculated for each indicator.

Results: Four of 47 indicators met both validation criteria (AUC>0.60 and 0.75<IF<1.25): main provider was a nurse/midwife, support companion present at birth, cesarean operation, and low birthweight infant. Twenty-five indicators met one criterion only, including skilled birth attendance.

Conclusions: Few indicators met the validation criteria, partly due to close to perfect reporting of routine practices. Validity results are influenced by context and question wording; low validity is associated with indicators related to the timing or sequence of events.

Introduction

Despite a 33 percent decline in maternal mortality from 1990 to 2011, persistent challenges remain [1]. Nearly 275,000 maternal deaths occurred globally in 2011, the vast majority of which took place in developing countries [1]. With the 2015 target date for the Millennium Development Goals (MDG) fast approaching, it is increasingly evident that most developing countries will require more time and a renewed action plan to achieve the goal of reducing maternal mortality (MDG5) [1]. In order to catalyze progress, efforts must be informed by the best available evidence. Yet despite the critical need for quality information to ensure concerted global action and the effective mobilization of resources to priority areas, assessing progress on MDG5 has been limited by a lack of reliable maternal health data [2]. Measurement challenges are particularly salient in developing country settings characterized by inadequate health systems with irregular and incomplete data reporting.

Given the difficulty of measuring maternal mortality, international and national agencies have relied on tracking proxy indicators, such as the proportion of births attended by a skilled birth attendant and the proportion of births delivered in health facilities, to measure progress [3]. Such indicators are routinely assessed in population-based household survey programs, such as the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS), implemented in low and middle income countries [4]. Despite widespread use, the majority of indicators proposed including skilled birth attendance, have not been sufficiently validated [5,6]. In fact, numerous researchers have noted the lack of correlation between these indicators and maternal mortality levels [5,7-9]. A crucial question is− can women accurately report on the quality and content of maternal health care received during their labor and delivery? Or are other factors, such as the misclassification of "skilled" providers, affecting noted disparities in maternal health outcomes. Such information is essential to informing national level action regarding the coverage and quality of maternal health care.

In response to the call for improvements in the measurement of maternal health care, a growing, but still limited, body of research has examined the validity of maternal and newborn health indicators. While these studies are useful for providing insight into the accuracy of self-reports, to our knowledge no study has yet reported on the validity of the skilled birth attendant indicator. Furthermore, the few validation studies that have taken place have generally compared maternal self-reports with hospital records, which may be incomplete or inaccurate, or have been conducted in high income settings, where maternal mortality rates are generally low [10].

To address this gap, the current study sought to identify and validate a set of quality of maternal health care indicators in two districts of Kenya. Specifically, a facility-based design

was used to compare women's self-reports of maternal care received against a gold standard of third party observations during labor and delivery. A primary research objective was validation of the skilled birth attendance indicator. We provide recommendations for the enhancement of data collection regarding maternal health care through the selection of key indicators that have the potential to be valid in routine population-based and facility-based data collection, a core component of accelerating progress in the global maternal health agenda.

Methods

Study Sites

Validation exercises were conducted in participating health facilities in two districts of Kenya: the New Nyanza Provincial General Hospital, now known as *Jaramogi Oginga Odinga Teaching and Referral Hospital* (JOOTRH), Kisumu District, formerly Nyanza Province; and Kiambu County Hospital, Kiambu District, formerly Central Province.[1] Both study facilities are high-volume public hospitals providing comprehensive obstetric care to women with both normal pregnancies who were self-referred for care and women with high-risk pregnancies referred from other health facilities. According to the 2009 Demographic and Health Survey, nationally, 43% of births in the past 5 years were delivered in a health facility [11]. Facility-based delivery is less likely among older women, those who have lower education, lower income, higher order births, or reside in rural areas [11]. The proportion of births delivered in a facility in Nyanza Province (44%) approximates the national level, while in Central Province it is substantially higher (73%). Educational attainment among women in Central and Nyanza Provinces corresponds with national rates [11]. Specifically, the majority of women aged 15-49 have completed at least some primary school in both Central (58%) and Nyanza Province (65%), as compared with 57% of women nationally. In both provinces, the percentage of deliveries in public sector institutions exceeds those in the private sector by more than three-times [11].

Data Collection

Data collection took place from July to September 2013. Participants included women aged 15 to 49 who underwent labor and delivery at a participating study facility and who were able to provide consent. Due to Ministry of Health guidelines, women who were more than six centimeters dilated were not approached for consent. All women who met study eligibility criteria and were admitted to the antenatal ward were invited to participate. Informed

[1] In August 2010, a new constitution was enacted which replaced Kenya's former 8 provinces into 47 semi-autonomous sub-counties. During the study, use of the new regional classification had not yet begun. Described geographical areas are comparable to the 2008-2009 DHS report for Kenya.

consent was obtained from all participants and their attending providers prior to participation.

Births were documented by trained researchers who observed providers from the maternity admission room and labor and delivery rooms. Study observers were registered nurse/midwives with at least three years of experience in a maternal, newborn and child health (MNCH) unit and who received detailed training on how to document maternal and immediate newborn postnatal care using the indicator checklist. Observations included events related to the birth itself as well as interactions between the women and provider, before, during and up to one hour after delivery. In the event that clarification was needed, observations were supplemented by checking facility records and by asking providers. We considered third party observations as the 'reference standard' as they were likely to reflect all facets of caregiving.

Exit interviews with participating women utilized an interviewer-administered questionnaire and took place prior to hospital discharge. The interview questionnaires were translated into Kiswahili, Dholuo and Kikuyu and underwent minor modifications to improve local understanding. Interviews were conducted in the language of the woman's preference. Data were directly entered into mobile electronic devices using Epi-data software and were exported to Stata Version 12 for statistical analysis.

Prior to participant enrollment, the study was reviewed and approved by ethical review committees of both the Population Council and the Kenya Medical Research Institute (KEMRI).

Indicator Selection

To identify indicators to be validated, a landscaping scan was conducted from April to July 2012. The scan focused on indicators currently in use or proposed for use, including both population-based and facility-based indicators. Indicators were identified by performing a key word search of electronic databases including: PUBMED, POPLINE, JSTOR and EMBASE. Additional searches were conducted of publications of organizations known for their involvement in measuring maternal health care: such as WHO, UNICEF, UNFPA, MCHIP, AMDD and IMMPACT and by searching reference lists of identified papers and reports. Key search terms included: maternal health, safe motherhood, quality of care, indicator, valid, skilled attendant, neonatal, perinatal, obstetric, intrapartum. No studies were excluded based on language or date of publication.

From an identified 2,505 documents, 71 provided information on indicators for assessing quality in maternal healthcare. Using this listing and a consultation with maternal health

experts, a set of 95 indicators was selected for validity testing (Supplemental Table 1). These indicators were selected on the basis of their wide use or potential to assess the critical elements of maternal and newborn care.

Analysis

A target sample size of 600 was calculated in order to detect at least a ten-percentage point difference between direct observations and women's self-report with 95 percent confidence. A difference of this magnitude was considered programmatically meaningful for the primary outcome of assessing how accurately women can identify the cadre of their birth attendant. An additional 20% was included to offset anticipated attrition in a separate study to re-interview women approximately one year following delivery.

Data analysis proceeded in two phases. In stage 1, we calculated the sensitivity and specificity of indicators of women's self-reported care by constructing two-by-two tables for each indicator that had at least five counts per cell. We then plotted the sensitivity, or true positive rate, of each indicator against its false positive rate (or 1 – specificity). To summarize the accuracy of each indicator, we then quantified the area under the receiver operating curve. In practice, the area under the curve (AUC) represents the "average accuracy of a diagnostic test" on a zero to one scale [12,13]. An AUC of 1.0, reflects an indicator with perfect classification accuracy while an AUC of 0 reflects zero accuracy.

To assess the population-based validity of indicators, we also estimated each indicator's inflation factor (IF). Using an equation by Vecchio, each indicator's estimated sensitivity and specificity was applied to its true prevalence (i.e., observer report) to estimate the prevalence that would be obtained using a population-based survey [14]. By comparing the ratio of the estimated survey-based prevalence to its true prevalence, we estimated the degree to which each indicator would be over or under-estimated if assessed using a population-based survey [15].

Together, the AUC and IF inform indicator validity at both the individual and population level. We used a priori benchmark criteria for 'valid' indicators of an AUC>0.6 and IF between 0.75 and 1.25. These criteria were previous benchmarks used in the literature for measuring the quality of maternal health care [10,16]. For statistics reported in the text we also report the margin of error.

Results

Sample Descriptive Frequencies

A total of 1039 women admitted for labor at participating study facilities were recruited to participate. Of those, a total of 676 women were observed during labor (Kiambu=395, Kisumu=281). As only women in early labor could be approached for consent, data collectors sought to consent all eligible women in the antenatal ward admission room, which at times included women admitted for purposes other than delivery. As a result, approximately one-third of women were not observed because they either did not progress into labor or they progressed rapidly into labor and full observation was not possible. See **Figure 1** for a detailed flow chart of participant enrollment.

[Insert Figure 1. Participant Response Rates.]

Descriptive statistics on participants' background characteristics are presented in **Table 1**. All participants resided in the surrounding Kiambu and Kisumu districts. Women who delivered in the Kisumu facility were on average 0.9 years younger in age (0.9 ± 0.4, $p<0.02$) and had on average 0.2 mean units greater educational attainment (0.2 ± 0.1, $p<0.01$, assessed on four point scale where 1=no education and 4= greater than secondary school). There were no differences among other key demographics including previous births, marital status and type of delivery (vaginal or cesarean section).

Thirteen percent of women delivered by caesarean section. Women who delivered by caesarean section did not differ from those who delivered vaginally by age, parity, educational status or marital status, but were more likely to experience a complication during their delivery (38%) compared to those who delivered vaginally (7%) (Pearson Chi-Square: 78.67; $p<0.01$).

[Insert Table 1. Descriptive Characteristics of Study Sample.]

Validation Results

The full list of indicators selected for validity testing is presented in supplemental Table S1. The table also describes the matched prevalence of each indicator by women's self-report and observer report (i.e., the 'true' prevalence). Women and observers were given the option to respond "Don't Know" to all indicators. "Don't Know" responses was generally minimal and were excluded from analysis.

Four indicators where the proportion of women who responded "Don't Know" exceeded 5% are reported in **Table 2**. Two of these indicators relate to the immediate postnatal period: whether the newborn was immediately dried after birth and whether the newborn

was immediately wrapped in a towel. Having a cesarean section was significantly associated with responding "Don't Know" to both of these questions. Specifically, women had 15 times the odds of not knowing if their newborn was immediately dried after birth (OR: 15.3 ± 4.8, $p<0.01$), and nearly 3 times the odds of not knowing if their newborn was immediately wrapped in a towel (OR: 2.7 ± 1.1, $p<0.02$) compared to women who had a vaginal delivery

The coverage of many maternal and immediate newborn practices was either near universal, or rarely occurred, depending on whether the intervention was preventative or harmful. A total of 47 indicators had adequate cell size for validity analysis (**Table 3**).

[Insert Table 3: Validation Results for Selected Indicators.]

A total of 4 indicators met both validation criteria. These indicators were: the main provider during delivery was a nurse/midwife, a support companion was present during the birth, cesarean operation, and low birthweight infant (<2,500 grams). An additional 25 indicators met either the AUC (13 indicators) or IF (12 indicators) criteria alone. We describe these results in relation to three areas of interest: (1) skilled birth attendance, and key elements of (2) maternal care and (3) immediate newborn care.

Skilled Attendance at Birth

A key objective of the study was to assess whether women could accurately identify the category of provider who assisted with their deliveries. The majority (81%) of births were attended by a nurse/midwife as the primary provider, followed by medical resident (9%), student nurse (5%), doctor/ob-gyn (3%), and other providers such as medical intern or clinical officer (<1%, respectively). Two types of providers during delivery had sufficient cell counts for robust analysis: nurse/midwife and student nurse. The nurse/midwife indicator met both the AUC (0.80 ± 0.03) and IF (0.93) criteria while the student nurse indicator met neither the AUC (0.57 ± 0.04) nor IF (0.45) criteria. The student nurse indicator had notably low sensitivity (16%), with the majority of misclassified reported cases mistaken for nurse/midwifes (65%). Although robust analysis was not possible on less common types of providers, cross-tabulation results reflect a high degree of individual level misclassification (see **Table 4**). For example, medical residents and nurse/midwives were each most likely to be misclassified by women as doctors. Results suggest women may be able to report on common types of providers, but have greater difficulty discerning among more narrow distinctions of provider skill-level.

TABLE 4. Cross-tabulation: Main provider during delivery.*

Self-Report (Number)	**Observer Report** (Number)							
	Doctor (Ob-gyn)	Medical resident	Medical Intern	Nurse/ Midwife	Clinical Officer	Student Nurse	Other	**Total**
Doctor (Ob-gyn)	16	46	6	46	2	7	0	123
Medical resident	0	1	0	0	0	0	0	1
Medical intern	0	0	0	1	0	1	0	2
Nurse/midwife	2	7	1	450	3	17	3	483
Clinical officer	1	0	0	12	0	1	0	14
Student nurse	0	1	0	8	0	5	0	14
Support person	0	0	0	1	0	0	0	1
No one	0	0	0	2	0	0	0	2
Other	0	2	0	2	0	0	0	4
DK	1	3	0	0	0	0	0	4
Total	20	60	7	522	5	31	3	648

*Note: Above table presents descriptive results. Validation analysis presented in text was restricted to matched data and excluded 'Don't Know' responses.

It is important to note that in Kenya, doctors/ob-gyns and nurses/midwives are considered 'skilled' in that they are each trained and legislated to perform critical and lifesaving interventions such as administration of antibiotics, uterotonics and anticonvulsant drugs and removal of retained products of conception [17]. For the purposes of this study medical residents were also considered 'skilled' in that they have completed their residency training but have not yet done their post-graduate specialization. Identifying 'skilled' rather than 'unskilled' birth attendance is of particular programmatic importance; we assessed a composite skilled birth attendance indicator. Although this indicator had high sensitivity (95%) it had low specificity (15%), reflecting a high false positive rate and came close to but did not meet the AUC (0.55 ± 0.04). However, at the aggregate level the imbalance of false positives and underreported true cases cancel out to produce an IF close to 1 (1.02), suggesting the indicator would generate an acceptable estimate of skilled birth attendance coverage at the population level.

Other Core Indicators of Maternal Care

Of the other core indicators of maternal care tested, two met both criteria for both the AUC and IF: whether a support person was present during the birth (AUC: 0.73± 0.07, IF: 0.77) and whether the delivery was a cesarean (AUC: 0.96 ± 0.01, IF: 1.01). A third indicator, receiving episiotomy, was very close to meeting both criteria (AUC: 0.87 ± 0.03, IF: 1.26).

Fifteen core indicators on maternal care met at least one criterion: 10 for AUC and 5 for IF. Indicators that met the AUC criteria only were: whether the woman received an injection, IV medication, or tablets (oral or rectal) to either bring on labor, to strengthen labor (i.e., uterotonic for induction or augmentation of labor), or for either induction or augmentation of labor, episiotomy, whether in the first physical examination following delivery, the provider took the woman's temperature or blood pressure, whether the woman received pain relief medication, and whether hemorrhage, prolonged labor (>12 hours), or no complications were experienced. Indicators on whether the woman's blood pressure was taken upon admission, whether she was encouraged to ambulate during labor or to assume different positions in labor, whether more than one provider assisted with the birth, and whether during the first physical examination following delivery the provider checked to see if the womb was becoming firm (i.e., check for involution) met the IF criteria only.

A key indicator of interest was women's report on whether she received an injection in the thigh or buttocks, medication through a tube in her arm, or tablets (placed in mouth or rectum) in the first few minutes after birth (i.e., uterotonic for postpartum hemorrhage). Nearly all women received a uterotonic for the prevention of hemorrhage (99%), which limited the validity analysis. Despite this, cross-tabulation results suggest the potential of this indicator. Specifically, of the 562 women observed to receive a uterotonic, 555 (98%) correctly reported someone performed one of the interventions necessary to receive the drug. However, results indicate women are unlikely to report on the time in which the uterotonic was received with accuracy. For example, an indicator on whether the uterotonic was received after the delivery of the placenta, had low sensitivity (54%) and specificity (41%), and did not meet the AUC (0.47± 0.04) or IF (25.1). While robust analysis of prophylactic uterotonic within 3 minutes of delivery was not possible, cross-tabulation data reflect high sensitivity (97%), but notably low specificity (3%), suggesting a high rate of false positive reporting by women. These results should be interpreted with care as full validity analysis was not possible due to insufficient variation in indicator prevalence.

Core Indicators of Immediate Newborn Care

Only one indicator of immediate newborn care, whether the newborn was low birthweight (<2,500 grams), met both the AUC (0.85± 0.03) and IF (0.87) criteria. In the exit interview, women were asked to give the gram weight of their newborn at birth. Mothers' reports of their newborn weight were collapsed into low (<2,500 grams) or normal weight (>2,500 and <5,000) [18]. Validation results should be interpreted with caution as project staff also observed that nearly all women were given a card that listed their baby's weight. Although the analysis was restricted to mothers who self-reported newborn weight (rather than read it off the card), this practice likely enhanced the salience and recall of birthweight information.

Whether the newborn was breastfed in the first hour after birth met the AUC only (0.63 ± 0.04), and whether the newborn was given to the mother immediately after birth met the IF

(1.04) only. Neither an indicator of whether the newborn was placed immediately skin-to-skin with the mother after birth, a key element of newborn thermal care, nor a composite indicator of three essential elements of newborn care (if the newborn was immediately dried, placed skin-to-skin with the mother, and breastfed within the first hour of birth), met either validation criteria. Of note was the high sensitivity (79%) and low specificity (28%) of the skin-to-skin item, and its corresponding effect in the composite indicator.

A closer examination of the construction of the skin-to-skin indicator highlights the importance of question wording on women's reporting. Specifically, women were asked, "*Did someone place the baby on your chest, against your skin, immediately after delivery of the baby*" (79% self-report prevalence, 16% true prevalence). Women who responded "Yes" were subsequently asked "*Was your baby wrapped in a towel while lying against your chest or lying naked against your skin.*" A cross-tabulation of these two questions indicates that of the women who reported their newborn was placed skin-to-skin, 85% subsequently indicated the baby was first wrapped in a towel, and then placed against their chest. The distinction is significant, as the appropriate standard for newborn thermal care is direct on-skin contact, with a towel subsequently draped over the mother and newborn. These results reveal that only 15% of women who responded "Yes" to the one-item indicator, were accurate in their understanding of the newborn being placed against the mother's skin. A two-item skin-to-skin indicator comprised of women who indicated that their newborn was placed on her skin and was naked (i.e., not first wrapped in a cloth), better approximated the true prevalence (29% self-report prevalence; 16% true prevalence). However, while specificity was substantially higher (70%), sensitivity was reduced (27%). Neither the two-item skin-to-skin indicator nor the composite essential newborn care indicator using this construct met both validation criteria.

Discussion

This study provides insight into the validity of 47 assessed indicators on the quality of maternal and immediate newborn health care received in two hospital-based facilities in Kenya. Across phases of care, we found indicators related to concrete, observable aspects of care or which reflected pain or concern may have been particularly salient for women and potentially enhanced reporting accuracy. These results are consistent with previous studies which have found particularly 'distinctive' events, such as cesarean operation [10,16,19] and having a support person present [16], have high overall validity.

That a low overall number of indicators met both validation criteria, in part, reflects that many interventions were routine practice for preventative health care that almost always occurred while other harmful practices rarely occurred, and there was not sufficient variation for robust analysis. For many indicators of beneficial or preventative care we found most

women accurately reported receiving the care (i.e., high indicator sensitivity). One interpretation of these results is that the majority of women correctly classified receiving routine preventative care. For example, an indicator on receiving a uterotonic for the prevention of postpartum hemorrhage (i.e., if an injection, IV medication or tablets were received in the first few minutes following birth), was accurately classified by nearly all women. Although the near universal implementation of the practice limited robust analysis for such indicators, cross-tabulation results suggest that some aspects of routine care can be validly reported. It is important to bear in mind that these results do not indicate whether women are aware of the purpose or name of the intervention, just that someone intervened.

Indicators of widely implemented routine care had high specificity that was often also associated with low specificity. This pattern likely is due to the fact that there were few instances in which standard preventative services were not received. Therefore, unless there was almost perfect negative classification by women, self-reports reflected low specificity. An alternate interpretation is that the observed pattern of high sensitivity and low specificity for many preventative care practices reflects social desirability bias among women, or the assumption that since they delivered in a hospital setting, they should have received appropriate care, rather than individual knowledge. This finding has also been described in a study of women's reporting of maternal and child health care in China [10].

The potential for social desirability bias may be relevant for indicators on skilled birth attendance at an individual level. Indicators measuring whether a provider was 'skilled' (a doctor, medical resident, or nurse/midwife) had high sensitivity and low specificity for both labor and delivery. Similarly, results indicate that women tended to underreport the presence of less skilled providers, such as student nurses, and overreport the presence of a doctor/ob-gyn. An additional explanation for the positive bias is differences in how women conceptualized key terms such as who their 'main' provider was. It is possible that women conceptualized their 'main' provider as the attendant with the highest rank and who may have been deemed 'in-charge' of her care, while observers identified the primary provider as the attendant who administered the majority of the care to the woman.

Study findings also suggest that the validity of some indicators may be highly dependent on context and question wording. Low validity indicators (i.e., indicators that did not meet either the AUC or IF) were particularly related to the timing or sequence of events. For example, in some cases newborns were placed in a warmer and subsequently breastfed. It is possible that women conceptualized the first hour after birth as the first hour after labor room procedures ended, rather than the exact time period used by observers. A two-item indicator about whether the newborn was placed skin-to-skin on the mother's chest immediately after delivery greatly reduced women's overestimation of the practice compared to a one-item indicator. That neither a one nor two-item skin-to-skin indicator met either validation criteria contrasts with findings from a recent study that interviewed mothers in

Mozambique [16]. However, women in the Mozambique study were interviewed 8 to 10 months following delivery. The results of this study are consistent with findings that women had difficulty reporting whether their newborn was placed skin-to-skin in a qualitative study of delivery and newborn care among women in Bangladesh and Malawi [20].

In addition, women who had a c-section were much less likely to be able to report on immediate newborn care than women with normal deliveries given high 'Don't Know' responses. This suggests that it may be worth excluding women with c-sections from questions about newborn care in routine household surveys.

A number of indicators met the AUC or IF criteria only. An important aspect of our results is that individual-level misclassification does not inherently signify that indicator measurement at the aggregate level will be inaccurate [15]. In studies where the goal is to estimate the approximate population-based coverage of an indicator, for example, discrepancies in false positives and false negatives may balance out to produce a close approximation of population level coverage (i.e., indicators that meet the IF criteria alone). For example, in this study setting, individual false positive reporting of 'skilled' birth attendance balanced out with underreporting of true cases to approximate true coverage. Knowing if an indicator's IF is large can also inform when corrective methods may need to be used to limit false positive reporting (e.g., use of a two-item indicator). This may be of particular use in settings where certain practices are rare, and the potential for over-estimation is hence large. Knowledge of whether an indicator is likely to be overestimated can also have significant programmatic implications. For example, settings where skilled birth attendance is markedly overreported could signify countries have not made as much progress in accelerating the proportion of births attended by a skilled provider as assumed.

While a major strength of this study was the use of a reference standard, i.e., direct observation by a third-party observer in a health facility, the study had a few notable limitations. Firstly, validation results rely on the assumption that the observer report was valid. To ensure the reference standard was an accurate reflection of care received, we used experienced nurse/midwives as third-party observers, conducted rigorous training on the data collection protocol and encouraged observers to check medical records or ask for clarification when needed. Additionally, validation results are reflective of women seeking facility-based delivery only, and may not be generalizable to women who deliver at home. For example, in Kenya increasing levels of education, income, and decreasing age and parity is correlated with facility delivery [11]. Our results may therefore be more reflective of a population with these characteristics. Another limitation is that, due to budget limitations, the study took place in public, government facilities only. The high overall standard of care and low variation in hospital practices limited the ability to robustly analyze all indicators. Future validation exercises should take place in a range of health institutions. The fact that

both study facilities received referrals from surrounding health facilities, may somewhat limit the extent to which patients may differ from those attending other types of health facilities.

Finally, our results inform a 'best case' scenario in terms of recall accuracy because women were interviewed shortly following delivery. Of interest is how reporting accuracy changes over time. To explore this question, we are conducting a follow-up study to re-interview women in their home community approximately one year after delivery. The validity of indicators over time will be informative for DHS and MICS household survey programs, which typically have a recall period of several years. This follow-up study also includes a qualitative component, which will explore women's understanding of key terms and concepts, highlighted by the present findings (e.g., 'main' provider, 'immediate' and 'skin-to-skin').

Conclusion

The measurement of the quality of maternal health care received in developing country settings often relies on data from surveys of women. More information is needed on how accurately women can recall events surrounding delivery and thus how valid these indicators are. The primary indicator of interest – delivery by a skilled birth attendant – met validation criteria for reporting at the aggregate level only. Indicator properties established here provide insight into contexts where indicator use is appropriate, and where adaptive procedures in data collection or question construction may be warranted. To extend the generalizability of these results, we recommend future validation studies be conducted in other developing country settings. Validation exercises should take place in a range of health care settings, to obtain greater variation in standards of care.

References

1. Lozano R, Wang H, Foreman KJ, Rajaratnam JK, Naghavi M, et al. (2011) Progress towards Millenium Development Goals 4 and 5 on maternal and child mortality: an updated systematic analysis. Lancet 378: 1139-1165.
2. WHO, UNICEF, UNFPA, The World Bank (2010) Trends in maternal mortality: 1990 to 2008. Geneva: World Health Organization.
3. Starrs A (1997) The Safe Motherhood Action Agenda: Priorities for the next decade. New York: Inter-Agency Group for Safe Motherhood, Family Care International.
4. Hodgins S (2013) Achieving better maternal and newborn outcomes: coherent strategy and pragmatic, tailored implementation. Global Health: Science and Practice 1: 146-153.
5. Graham WJ, Bell JS, Bullough CH (2001) Can skilled attendance at delivery reduce maternal mortality in developing countires? Studies in Health Services Organisation and Policy 17: 97-130.
6. Koblinksy M, Conroy C, Kursehy N, Stanton ME, S. J (2000) Issues in programming for safe motherhood. Arlington, VA: John Snow, Inc.
7. Scott S, Ronsmans C (2009) The relationship between birth with a health professional and maternal mortality in observational studies: a review of the literature. Tropical Medicine and International Health 14: 1523-1533.
8. Lohela TJ, Campbell OM, Gabrysch S (2012) Distance to care, facility delivery and early neonatal mortality in Malawi and Zambia. PLoS One 7: e52110.
9. Hounton S, Menten J, Ouedraogo M, Dubourg D, Meda N, et al. (2008) Effects of a skilled care initiative on pregnancy-related mortality in rural Burkina Faso. Tropical Medicine and International Health 13: 53-60.
10. Liu L, Li M, Yang L, Ju L, Tan B, et al. (2013) Measuring coverage in MNCH: A validation study linking population survey derived coverage to maternal, newborn and child health care records in rural China. PLOS One 8: e60762.
11. Kenya National Bureau of Statistics (KNBS) and ICF Macro (2010) Kenya Demographic and Health Survey 2008-09. Calverton, Maryland: KNBS and ICF Macro.
12. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 201: 621-625.
13. Eng J (2005) Receiver operating characteristic analysis: a primer. Academic Radiology 12: 909-916.
14. Vecchio TJ (1966) Predictive value of a single diagnostic test in unselected populations. New England Journal of Medicine 274: 1171-1173.
15. Campbell H, Biloglav Z, Rudan I (2008) Reducing bias from test misclassification in burden of disease studies: use of test to actual positive ratio- new test parameter. Croatian Medical Journal 49: 402-414.

16. Stanton SK, Rawlins B, Drake M, dos Anjos M, Cantor C, et al. (2013) Measuring coverage in MNCH: Testing the validity of women's self-report of key maternal and newborn health interventions during the peripartum period in Mozambique. PLOS One 8: e60694.
17. Adegoke A, Utz B, Msuya SE, van den Broek N (2012) Skilled birth attendants: who is who? A descriptive study of definitions and roles from nine sub saharan african countries. PLOS One 7: e40220.
18. United Nations Children's Fund and World Health Organization (2004) Low Birthweight: Country and regional and global estimates. New York: UNICEF.
19. Tuncalp O, Stanton C, Castro A, Adanu R, Heymann M, et al. (2013) Measuring coverage in MNCH: Validating women's self-report of emergency cesarean sections in Ghana and the Dominican Republic. PLOS One 8: e60761.
20. Yoder P, Rosato M, Mahmud R, Fort A, Rahman F, et al. (2010) Women's recall of delivery and neonatal care: a study of terms, concepts and survey questions. Calverton, Maryland: ICF Macro.

Figures & Tables

Figure 1: Participant response rates.

Table 1: Sample Background Characteristics by Facility Location.

Table 2: Indicators with High 'Don't Know' Responses, Unmatched Data.

Table 3: Validation Results for Selected Indicators.

Supporting Information: Table S1: Full List of Indicators for Validation Exercises and Measured Coverage.

Figure 1: Participant response rates.

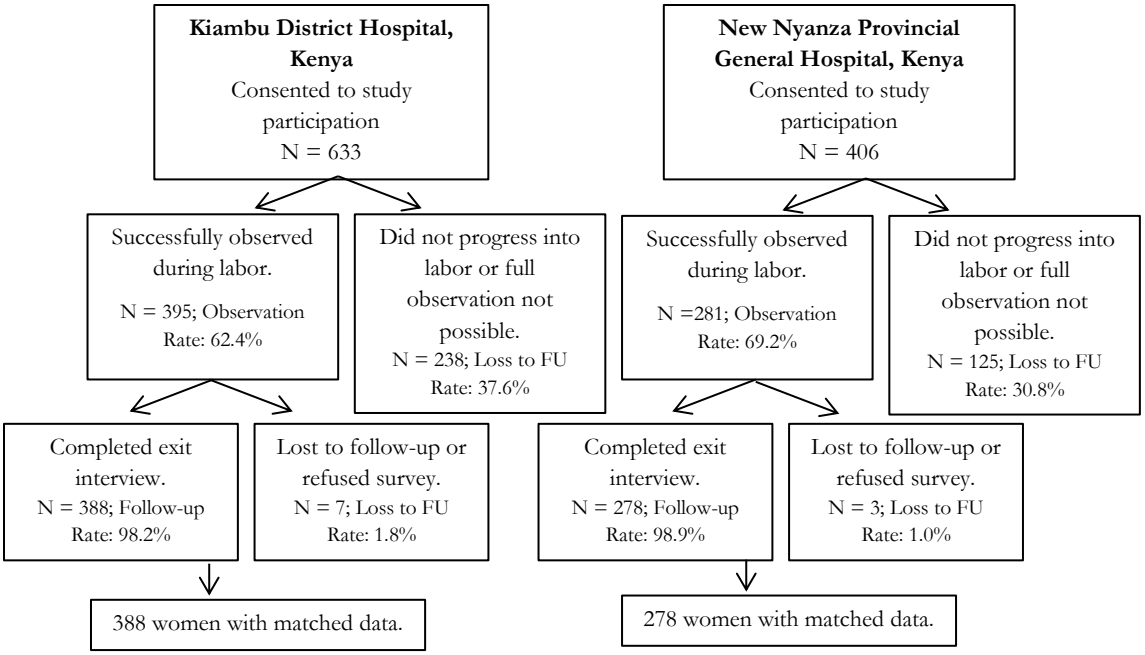

TABLE 1. Sample Background Characteristics by Facility Location.

	Total Sample (%)	**Kiambu (%)**	**Kisumu (%)**	**P-Value**
Age				p=0.021
15-19	14.5	11.9	18.1	
20-24	42.0	43.1	40.6	
25-29	28.9	29.4	28.6	
30-34	8.5	8.8	9.1	
35-39	5.7	6.2	4.7	
40-45	0.4	0.5	0.4	
Prior Parity (Total # Live Births)				p=0.408
0	50.2	49.7	51.3	
1	26.5	28.8	22.9	
2	14.0	13.2	14.9	
3	6.0	5.2	7.3	
4 or more	3.3	3.1	3.6	
Educational attainment				p=0.007
None	10.2	10.3	10.2	
Primary	43.6	45.9	41.2	
Secondary	29.2	33.2	24.1	
Higher	16.2	10.6	24.5	
Marital status				p=0.128
Single, never married	14.7	9.8	21.5	
Married	78.1	80.2	75.2	
Living together	5.3	7.5	2.2	
Separated	1.8	2.6	0.7	
Widowed	0.2	0.0	0.4	
Type of delivery				p=0.679
Vaginal	86.6	87.0	85.9	
Cesarean section	13.4	13.0	14.1	

TABLE 2. Indicators with High 'Don't Know' Responses, Unmatched Data.

Indicator	% "Don't Know"
Woman Self-Report (>5% Don't Know)	
Did the health provider(s) wash his/her hands with soap and water or use antiseptic before examining you?	29.5
Was your baby wrapped in a towel or cloth immediately after birth?	20.6
Was your baby dried off with a towel or cloth immediately after his/her birth, within a few minutes of delivery?	8.4
In your first physical examination after delivery, did a health provider do a perineal exam?	9.8
Observer Report (>1% Don't Know)	
Was anything besides breast milk given to the baby to drink within the first hour after birth?	5.1
Baby bathed within first hour after birth	2.4

TABLE 3. Validation Results for All Indicators With at least 5 Counts per Cell, Matched Data.

Indicator	N Matched data	Reported Prev (%) Matched data	True Prev (%) Matched data	Sensitivity of Self Report	Specificity of Self-Report	Population Survey Estimate	AUC (>0.60)	IF (0.75 to 1.25)	Recommend? (Y/N) *List Criteria*
Initial Client Assessment									
Woman referred to facility because of a problem	655	8.24	7.94	25.00	93.20	8.24	0.5910	1.04	IF
Provider washes hands with soap and water or uses antiseptic before initial examination	467	73.09	26.55	83.87	32.94	71.52	0.5841	2.69	
Takes blood pressure	654	93.43	87.00	87.73	23.26	86.30	0.5549	0.99	IF
Provider Respectful Care									
Woman allowed to drink liquids or eat	624	66.83	41.99	72.52	37.29	66.83	0.5491	1.59	
Encourages/assists woman to ambulate during labor	644	86.96	77.48	90.18	24.14	86.96	0.5716	1.12	IF
Encourages/assists woman to assume different positions in labor	649	14.33	58.24	19.05	92.25	14.33	0.5565	0.25	
Woman allowed to have a support person during labor and delivery	648	8.80	9.10	23.73	92.70	8.80	0.5821	0.97	IF
Support companion present during birth	644	3.73	4.81	48.39	98.53	3.73	0.7346	0.77	**Yes**
First Stage of Labor									
Induces labor with uterotonic	630	10.79	4.60	68.97	92.01	10.80	0.8049	2.35	AUC
Augments labor with uterotonic	625	39.20	22.40	72.86	70.52	39.20	0.7169	1.75	AUC
Uterotonic received (labor induction or augmentation)	619	43.78	27.14	77.98	68.96	68.96	0.7347	1.61	AUC
Membranes ruptured (labor induction or augmentation)	650	3.08	42.31	4.00	97.60	3.08	0.5080	0.07	
Skilled Birth Attendance									
Skilled main provider* labor	649	89.98	92.60	90.52	16.67	89.99	0.5359	0.97	IF
Main provider labor nurse/midwife	638	80.41	93.73	81.10	30.00	80.28	0.5555	0.86	IF
Other provider labor nurse/ midwife	654	50.46	69.88	48.36	59.90	45.87	0.5413	0.66	
Other provider labor student nurse	654	5.35	12.23	13.75	95.82	5.35	0.5478	0.44	
Skilled main provider* delivery	644	94.25	92.86	94.98	15.22	94.25	0.5510	1.02	IF
Main provider delivery doctor (ob-gyn)/ medical resident	644	19.25	11.80	82.89	89.26	19.25	0.8608	1.63	AUC

Indicator	N Matched data	Reported Prev (%) Matched data	True Prev (%) Matched data	Sensitivity of Self Report	Specificity of Self-Report	Population Survey Estimate	AUC (>0.60)	IF (0.75 to 1.25)	Recommend? (Y/N) *List Criteria*
Main provider delivery nurse/ midwife	644	75.00	81.06	86.21	72.95	75.00	0.7958	0.93	**Yes**
Main provider delivery student nurse	644	2.17	4.81	16.13	98.53	2.18	0.5733	0.45	
Other provider delivery doctor (ob-gyn)	651	9.52	3.07	35.00	91.28	9.53	0.6314	3.10	AUC
Other provider delivery nurse/ midwife	650	48.15	64.31	49.52	54.31	48.15	0.5192	0.75	IF
Other provider delivery student nurse	647	5.26	8.96	13.79	95.59	5.25	0.5469	0.59	
Second and Third Stage Labor									
Episiotomy performed	545	22.94	18.17	82.83	90.36	22.94	0.8659	1.26	AUC
Uterotonic received following delivery of placenta	552	59.06	2.36	53.85	40.82	59.05	0.4733	25.1	
Multiple providers assisted with birth	629	64.55	54.69	70.35	42.46	64.55	0.5640	1.18	IF
Immediate Newborn Postnatal Care									
Baby given to mother immed. after birth	611	59.90	57.61	66.48	49.03	59.91	0.5776	1.04	IF
Baby placed immed. skin to skin on mother	602	78.90	16.28	78.57	21.03	78.90	0.498	4.85	
Baby placed immed. skin to skin on mother (2 item)[#]	596	29.19	16.20	26.80	70.34	29.20	0.4857	1.80	
Breastfeeding within first hr of birth	551	76.41	52.99	88.36	37.07	76.41	0.6271	1.44	AUC
3 elements of essential newborn care (immed. dried, on mother's skin, breastfed within first hr)	506	71.5	9.29	70.21	28.32	71.54	0.4927	7.70	
3 elements of essential newborn care (immed. dried, 2 item on mother's skin[#], breastfed within first hr)	501	29.44	9.18	7.14	90.13	9.62	0.4864	1.05	IF
Low birthweight newborn (<2,500g)	579	6.74	7.77	71.11	98.69	6.73	0.8490	0.87	**Yes**
Immediate Postnatal Care									
Palpates uterus 15 minutes after delivery of placenta	557	88.33	70.20	88.75	12.65	88.33	0.5070	1.26	
First post-delivery exam, provider ask/checks for bleeding	627	62.04	90.59	59.86	16.95	62.04	0.3840	0.68	
First post-delivery exam, provider examines perineum	554	56.14	87.36	57.85	55.71	56.14	0.5678	0.64	
First post-delivery exam, provider takes temperature	638	60.03	40.28	75.10	50.13	60.03	0.6261	1.49	AUC
First post-delivery exam, provider takes blood pressure	642	74.61	48.29	88.06	37.95	74.61	0.6301	1.55	AUC
First post-delivery exam, provider checks for involution	615	64.23	78.21	64.03	35.07	64.23	0.4955	0.82	IF

Indicator	N Matched data	Reported Prev (%) Matched data	True Prev (%) Matched data	Sensitivity of Self Report	Specificity of Self-Report	Population Survey Estimate	AUC (>0.60)	IF (0.75 to 1.25)	Recommend? (Y/N) *List Criteria*
Woman asked for pain relief medication during stay	638	32.13	10.50	35.82	68.30	32.13	0.5206	3.06	
Woman received pain relief medication	640	59.38	17.50	85.71	46.21	59.38	0.6596	3.39	AUC
Maternal Outcomes									
Cesarean section (C/S) performed	651	13.52	13.36	93.10	98.76	13.52	0.9593	1.01	**Yes**
Reason for C/S- prolonged/obstructed labor	76	32.89	67.11	39.22	80.00	32.90	0.5961	0.49	
Complications (any)	654	44.80	11.00	62.50	57.39	44.80	0.5994	4.07	
Hemorrhage	654	11.16	4.59	33.33	89.90	11.17	0.6162	2.43	AUC
Prolonged labor	654	23.70	3.67	50.00	77.30	23.70	0.6365	6.46	AUC
None	654	51.53	88.99	53.78	66.67	51.53	0.6022	0.58	AUC

Notes: Recommended indicators meet both AUC and IF validation criteria.

* Skilled provider includes doctor (ob-gyn), medical resident or nurse/midwife

Indicator constructed from two skin-to-skin items: (1) baby placed against mother's chest after delivery and (2) baby was lying naked against the mother's chest.

Table S1. Full List of Indicators Assessed and Measured Coverage[a,b]

Indicator	Self-Report Prevalence (Matched data)	True Prevalence (Matched data)	At least 5 counts/cell?
Initial Client Assessment			
Type of facility where gave birth	98.00	100.00	N
Referred to facility because of a problem	8.24	7.94	Y
HIV status checked	25.19	94.67	N
Offered HIV test	8.33	1.67	N
Receives HIV test	8.56	9.33	N
Provider washes hands with soap and water or uses antiseptic before any initial examination	73.09	26.55	Y
Takes blood pressure	93.43	87.00	Y
Takes urine sample	5.66	1.38	N
Checks fetal heart rate with fetoscope/ ultrasound	95.75	99.7	N
Wears high-level disinfected or sterile gloves for vaginal examination	99.85	99.85	N
Provider Respectful Care			
Encourages/assists woman to ambulate during labor	86.96	77.48	Y
Woman allowed to drink liquids/eat	66.83	41.99	Y
Woman allowed to have a support person present during labor and delivery	8.78	9.10	Y
Encourages/assists woman to assume different positions in labor	14.33	58.24	Y
A support person is present at birth	3.73	4.81	Y
First Stage of Labor			
Induces labor by uterotonic (IV, IM, tablet) [d]	10.79	4.60	Y
(Of women whose labor was induced) Uterotonic route for induction of labor - Tablet (oral or vaginal) [d]	80.95	100.00	N
Augments labor with uterotonic (by IV line, IM injection, or tablet) [d]	39.20	22.40	Y
(Of women whose labor was augmented) Augmentation of labor by IV line (Push, Drip, Drip plus IM) [d]	92.45	100.00	N
Uterotonic received (to induce or augment labor) [d]	43.78	27.14	Y
Membranes ruptured (to induce or augment labor) [d]	3.08	42.31	Y
Skilled Birth Attendance- Main Provider			
Skilled main provider labor[c, d]	89.98	92.60	Y
Main provider labor- doctor or medical resident [d]	9.55	0.46	N
Main provider labor- doctor (ob-gyn)	9.72	0.31	N
Main provider labor- medical resident	0.00	0.16	N
Main provider labor- medical intern	0.16	1.88	N
Main provider labor- nurse/midwife	80.41	93.73	Y
Main provider labor- clinical officer	2.04	0.63	N
Main provider labor- facility support/ staff aide	0.15	0.31	N

Indicator	Self-Report Prevalence (Matched data)	True Prevalence (Matched data)	At least 5 counts/cell?
Main provider labor- student nurse	2.31	2.77	N
Main provider labor- support companion	0.62	1.71	N
Skilled main provider delivery[c,d]	94.25	92.86	Y
Main provider delivery- doctor (ob-gyn) or medical resident[†]	19.25	11.80	Y
Main provider delivery- doctor (ob-gyn)	19.10	3.00	N
Main provider delivery- medical resident	0.15	8.85	N
Main provider delivery- medical intern	0.31	1.09	N
Main provider delivery- nurse/midwife	75.00	81.06	Y
Main provider delivery- clinical officer	2.17	0.78	N
Main provider delivery- student nurse	2.17	4.81	Y
Skilled Birth Attendant- Other Providers Present			
Other provider(s) labor			
Other provider labor- doctor (ob-gyn)	9.02	1.99	N
Other provider labor- medical resident	0.15	3.98	N
Other provider labor- medical intern	0.92	28.29	N
Other provider labor- nurse/midwife	50.46	69.88	Y
Other provider labor- student nurse	5.35	12.23	Y
Other provider labor- clinical officer	3.82	9.79	N
Other provider(s) delivery			
Other provider delivery- doctor (ob-gyn)	9.52	3.07	Y
Other provider delivery- medical resident	0.15	4.92	N
Other provider delivery- medical intern	0.61	14.59	N
Other provider delivery- nurse/midwife	48.15	64.31	Y
Other provider delivery- student nurse	5.26	8.96	Y
Other provider delivery- clinical officer	4.79	5.41	N
More than one provider assisted with birth[d]	64.55	54.69	Y
Second & Third Stage of Labor			
Episiotomy performed	22.94	18.17	Y
Uterotonic administered within few minutes of delivery (via injection, IV medication, or oral/rectal tablets)	96.80	98.75	N
Uterotonic received 1-3 mins after birth	96.92	81.52	N
Uterotonic received after delivery of placenta	59.06	2.36	Y
Applies controlled cord traction	97.50	98.93	N
Performs uterine massage after delivery of placenta	88.35	98.57	N
Position of mother at birth- on back	94.73	99.84	N
Health provider wore gloves during delivery of baby	100.00	99.82	N
Immediate Newborn Care (babies breathing at birth)			
Baby immediately dried with towel/cloth	96.13	99.49	N
Baby given to mother immediately after birth	59.90	57.61	Y
Baby placed immediately skin to skin on mother's abdomen	78.90	16.28	Y

Indicator	Self-Report Prevalence (Matched data)	True Prevalence (Matched data)	At least 5 counts/cell?
Baby immediately skin to skin on mother (2 item indicator) [e]	29.19	16.20	Y
Babies on skin covered with dry towel on mothers abdomen	42.31	100.00	N
Babies not on skin wrapped with towel	90.70	91.86	N
Breastfeeding within first hour of birth	76.41	52.99	Y
Something other than breastmilk given to baby within first hour of delivery	1.92	1.05	N
Baby bathed within the first hour after birth [d]	2.81	0.05	N
Baby weighed	99.84	100.00	N
Low birth-weight baby (<2,500 g) [d]	6.74	7.77	Y
High birth-weight baby (>=4,500 g) [d]	1.03	1.03	N
3 elements of newborn care (immed. dried + on skin + breastfed in first hour) [d]	71.54	9.29	
3 elements of newborn care (immed. dried, 2 item skin-to-skin [e], breastfed in first hour) [d]	29.4	9.18	Y
Immediate Postnatal Care			
Palpates uterus 15 minutes after delivery of placenta	88.33	70.20	Y
Provider did at least one post-delivery health check [d]	95.99	94.92	N
In first post-delivery exam, provider checks for bleeding	62.04	90.59	Y
In first post-delivery exam, provider examines perineum	56.14	87.36	Y
In first post-delivery exam, provider takes temperature	60.03	40.28	Y
In first post-delivery exam, provider takes blood pressure	74.61	48.29	Y
In first post-delivery exam, provider checks for involution	64.23	78.21	Y
Maternal and Infant Outcomes			
Cesarean section (C/S) performed	13.52	13.36	Y
Decision for C/S taken after labor started	90.79	100.00	N
C/S performed after labor started	90.79	100.00	N
Provider decided C/S would be done	82.50	100.00	N
Reason for C/S- prolonged/obstructed labor	32.89	67.11	Y
Complications- Any [d]	44.80	11.00	Y
Eclampsia	10.86	0.31	N
Hemorrhage	11.16	4.59	Y
Prolonged labor (>12 hours)	23.70	3.67	Y
None	51.53	88.99	Y
Blood products given	15.28	18.06	N
Woman asked for pain relief medication while at facility	32.13	10.50	Y
Woman received pain relief medication	59.38	17.50	Y
Stillborn delivery [d]	0.92	1.38	N

[a] Text in blue notes indicators where there was not sufficient cell counts for robust analysis (n<5 per cell).
[b] Excludes 'Don't Know' responses.
[c] Skilled provider is doctor (ob-gyn), nurse/midwife or medical resident
[d] Indicator constructed in analysis to dichotomize women's responses to related question.
[e] Indicator constructed from two skin-to-skin items: (1) baby placed against mother's chest after delivery and (2) baby was naked on skin.