Assessment of the validity of the measurement of newborn and maternal health-care coverage in hospitals (EN-BIRTH): an observational study

Louise Tina Day, Qazi Sadeq-ur Rahman*, Ahmed Ehsanur Rahman, Nahya Salim, Ashish KC, Harriet Ruysen, Tazeen Tahsina, Honorati Masanja, Omkar Basnet, Georgia R Gore-Langton, Sojib Bin Zaman, Josephine Shabani, Anjani Kumar Jha, Vladimir Sergeevich Gordeev, Shafiqul Ameen, Donat Shamba, Bijay Jha, Dorothy Boggs, Tanvir Hossain, Kizito Shirima, Ram Chandra Bastola, Kimberly Peven, Abu Bakkar Siddique, Godfrey Mbaruku†, Rajendra Paudel, Angela Baschieri , Aniqa Tasnim Hossain, Stefanie Kong, Asmita Paudel, Anisuddin Ahmed, Simon Cousens, Shams El Arifeen‡, Joy E Lawn‡, EN-BIRTH validation collaborative group§, EN-BIRTH Expert Advisory Group§*

Summary

Background Progress in reducing maternal and neonatal deaths and stillbirths is impeded by data gaps, especially regarding coverage and quality of care in hospitals. We aimed to assess the validity of indicators of maternal and newborn health-care coverage around the time of birth in survey data and routine facility register data.

Methods Every Newborn-BIRTH Indicators Research Tracking in Hospitals was an observational study in five hospitals in Bangladesh, Nepal, and Tanzania. We included women and their newborn babies who consented on admission to hospital. Exclusion critiera at admission were no fetal heartbeat heard or imminent birth. For coverage of uterotonics to prevent post-partum haemorrhage, early initiation of breastfeeding (within 1 h), neonatal bag-mask ventilation, kangaroo mother care (KMC), and antibiotics for clinically defined neonatal infection (sepsis, pneumonia, or meningitis), we collected time-stamped, direct observation or case note verification data as gold standard. We compared data reported via hospital exit surveys and via hospital registers to the gold standard, pooled using random effects meta-analysis. We calculated population-level validity ratios (measured coverage to observed coverage) plus individual-level validity metrics.

Findings We observed 23 471 births and 840 mother–baby KMC pairs, and verified the case notes of 1015 admitted newborn babies regarding antibiotic treatment. Exit-survey-reported coverage for KMC was 99·9% (95% CI 98·3–100) compared with observed coverage of 100% (99·9–100), but exit surveys underestimated coverage for uterotonics (84·7% [79·1–89·5]) *vs* 99·4% [98·7–99·8] observed), bag-mask ventilation (0·8% [0·4–1·4]) *vs* 4·4% [1·9–8·1]), and antibiotics for neonatal infection (74·7% [55·3–90·1] *vs* 96·4% [94·0–98·6] observed). Early breastfeeding coverage was overestimated in exit surveys (53·2% [39·4–66·8) *vs* 10·9% [3·8–21·0] observed). "Don't know" responses concerning clinical interventions were more common in the exit survey after caesarean birth. Register data underestimated coverage of uterotonics (77·9% [37·8–99·5] *vs* 99·2% [98·6–99·7] observed), bag-mask ventilation (4·3% [2·1–7·3] *vs* 5·1% [2·0–9·6] observed), KMC (92·9% [84·2–98·5] *vs* 100% [99·9–100] observed), and overestimated early breastfeeding (85·9% (58·1–99·6) *vs* 12·5% [4·6–23·6] observed). Inter-hospital heterogeneity was higher for register-recorded coverage than for exit survey report. Even with the same register design, accuracy varied between hospitals.

Interpretation Coverage indicators for newborn and maternal health care in exit surveys had low accuracy for specific clinical interventions, except for self-report of KMC, which had high sensitivity after admission to a KMC ward or corner and could be considered for further assessment. Hospital register design and completion are less standardised than surveys, resulting in variable data quality, with good validity for the best performing sites. Because approximately 80% of births worldwide take place in facilities, standardising register design and information systems has the potential to sustainably improve the quality of data on care at birth.

Funding Children's Investment Fund Foundation and Swedish Research Council.

Introduction

Investment in global health measurement has particularly focused on outcomes, notably deaths. Accurate data are urgently needed to track the progress towards the Sustainable Development Goals (SDGs) to end the annual 5·1 million preventable stillbirths and newborn deaths, plus 0·3 million maternal deaths by 2030.[1–4] Despite nearly 80% of births worldwide now being in health-care facilities,[5] many avoidable deaths occur, notably intrapartum stillbirths and in preterm newborn babies.[6,7] Many evidence-based, high-impact interventions for maternal and newborn health are

Lancet Glob Health 2021; 9: e267–79

Published Online December 14, 2020 https://doi.org/10.1016/S2214-109X(20)30504-0

This online publication has been corrected. The corrected version first appeared at thelancet.com/lancetgh on February 16, 2021

See **Comment** page e221

*Joint first authors

†Deceased

‡Joint senior authors

§Group members listed at the end of the Article

Maternal, Adolescent, Reproductive, & Child Health Centre, London School of Hygiene & Tropical Medicine, London, UK (L T Day MRCPCH, H Ruysen MSc, G R Gore-Langton MSc, V S Gordeev PhD, D Boggs MSc, K Peven MPH, A Baschieri PhD, S Kong MSc, S Cousens MA, J E Lawn FMedSci)**; Maternal and Child Health Division, International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh** (Q Sadeq-ur Rahman MSc, A E Rahman MPH, T Tahsina MSc, S B Zaman MPH, S Ameen MBBS, T Hossain MSc, A B Siddique MSc, A T Hossain MSc, A Ahmed MSc, S E Arifeen DPH)**; Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania** (N Salim PhD)**; Department of Health Systems, Impact Evaluation and Policy, Ifakara Health Institute, Dar es Salaam, Tanzania** (N Salim, H Masanja PhD, J Shabani MSc, D Shamba MSc, K Shirima HDiploma, G Mbaruku PhD)**; International Maternal and Child Health,**

Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden (A KC PhD); Research Division, Golden Community, Lalitpur, Kathmandu, Nepal (O Basnet BBS, R Paudel MSc, A Paudel BN); Nepal Health Research Council, Kathmandu, Nepal (Anjani Kumar Jha MD, Bijay Jha BCom); The Institute of Population Health Sciences, Queen Mary University of London, London, UK (V S Gordeev); Pokhara Academy of Health Science, Pokhara, Nepal (R C Bastola); Ministry of Health and Population, Kathmandu, Nepal (R C Bastola); and Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, London, UK (K Peven)

Correspondence to: Dr Louise T Day, Maternal, Adolescent, Reproductive, & Child Health Centre, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK louise-tina.day@lshtm.ac.uk

Research in context

Evidence before this study

Increasing coverage and quality of care around the time of birth is fundamental to reducing 5·3 million maternal, fetal, and newborn deaths and disability every year and achieving the Sustainable Development Goals. National and global tracking of maternal and newborn health has mostly used measures of contact coverage (eg, skilled birth attendance), with little attention paid to indicators of content or quality of care. We searched Ovid MEDLINE for articles published in English after Jan 1, 2010, with the search terms "valid" AND "maternal or newborn or neonate or labour or childbirth or delivery or intrapartum" AND "coverage or indicator or measure or track or numerator or denominator" and restricted to low-income and middle-income settings. Of the 598 papers identified, 17 met our inclusion criteria, among which most observational studies focused on indicator measurement validation in survey with two for routine register data. We found no additional relevant documents in the grey literature. Previous studies have reported low validity for measures around the time of birth. High-priority newborn interventions, such as neonatal bag-mask ventilation and kangaroo mother care (KMC), have not undergone validity testing in surveys or register data. Early initiation of breastfeeding has only been validated in surveys. WHO and UNICEF's *Every Newborn* Action Plan prioritised validation of data on indicators of maternal and newborn coverage of care to improve measurement of content of care and especially through routine data systems.

Added value of this study

The *Every Newborn*-Birth Indicators Research Tracking in Hospitals (EN-BIRTH) study was done in five hospitals in Bangladesh, Nepal, and Tanzania, and aimed to assess the validity of data for five indicators of maternal and newborn health-care coverage in two data sources: routine facility registers and women's report at exit survey. These data were compared against a gold standard of direct observation or case note verification. EN-BIRTH was about 10 times larger than previous facility-based validation studies and included more than 25 000 cases—either observed births or KMC or newborn admission case notes. This study is the first to assess the validity of routine register data for most of these indicators, and of survey data for the hospital newborn care indicators. We found that exit survey reports had low accuracy for uterotonic coverage and for early initiation of breastfeeding, consistent with previous smaller studies. Population-based household surveys already capture early initiation of breastfeeding coverage but research is needed to improve accuracy. We also found survey-reported data on bag-mask ventilation had low accuracy and neonatal infection treatment with antibiotics had low sensitivity among the target group. KMC coverage was accurately reported at exit survey by women who had practised KMC. Thus, further work is required to assess whether KMC report remains reliable after the typical household survey recall period of 2–5 years, and also to ascertain the extent to which mothers who did not practise KMC misreport having done so. Routine registers in some hospitals were found to be highly complete, but accuracy varied between hospitals, even with the same register design. Register accuracy for uterotonics was excellent in two hospitals, and KMC sensitivity was excellent in two hospitals and good in two hospitals. One hospital had good register accuracy for bag-mask ventilation. For early initiation of breastfeeding, register accuracy was poor in all four study hospitals with a register column.

Implications of all the available evidence

For care around the time of birth, surveys are important for estimating population-based contact coverage and family-led behaviours such as breastfeeding, but our findings do not support adding questions about clinical interventions to surveys, with the possible exception of admission for KMC. Given that approximately 80% of the world's births are now in facilities, routine registers can provide data on intervention delivery more rapidly and at lower cost than surveys. Optimising register design, filling, and flow into national routine information systems requires investment in implementation research. Caesarean birth negatively affected the accuracy of survey-reported and register-recorded data. Further research is required regarding the measurement implications of increasing caesarean section rates. Reliable data are necessary, but not sufficient to improve care around the time of birth—health workers, policy makers, and politicians must also value and use these data to drive change.

delivered in health-care facilities, yet gaps in quality of care and gaps in data result in missed opportunities.[7,8] The SDGs aspire to achieve universal health coverage, which will be difficult without addressing the crucial measurement gaps regarding effective coverage and quality around the time of birth.[9]

For the **Demographic and Health Surveys programme** see https://dhsprogram.com/

For **UNICEF's Multiple Indicator Cluster Surveys** see http://mics.unicef.org/

Coverage is defined as the proportion of individuals receiving an intervention or practice (numerator) among the population in need of that intervention or practice (denominator). National and global tracking of coverage has primarily focused on survey measurement of contacts with the health system (eg, institutional birth), with few indicators capturing content (eg, interventions) or quality of care.[10,11] Maternal and newborn coverage of care in low-income and middle-income countries (LMICs) is mainly tracked through large-scale population-based household surveys, notably the Demographic and Health Surveys (DHS) programme and UNICEF's Multiple Indicator Cluster Surveys (MICS). Although these surveys importantly measure population-based contact coverage, previous research has found mixed validity for the content of care around the time of birth (eg, breastfeeding).[12–16] DHS has done more than 400 surveys in 90 countries. With over 400 questions in the DHS core questionnaire, focus has been on the need to validate additional questions before adding. Overall, few survey indicators have been validated, and those relating to clinical care for small and ill newborn babies. have not previously been validated.[17]

The shift to most births worldwide being in facilities, paired with rapid improvements for routine national health management information systems (HMIS) including digitalisation, have potential to transform measurement of coverage and quality of care for women and newborn babies, including in high-burden settings.[17] Health workers document details of admitted women and newborn babies in routine facility registers, usually in parallel to individual patient case notes. These registers are the primary source for aggregate data that flow into routine HMIS. The quality of HMIS data, or mistrust of quality, impedes full use by policy makers.[18–20] Previous studies of routine paper-based registers in facilities in LMICs have reported on data availability,[21,22] but only two small observational studies have examined some aspects of register measurement validity for care around the time of birth.[23,24]

Transforming measurement and use of data to track coverage and quality of care is one of five strategic objectives in the *Every Newborn* Action Plan, led by WHO and UNICEF, implemented in more than 92 countries.[25] Validation of coverage measurement for interventions and practices (content of care) was prioritised.[26] Core indicators regarding high-impact maternal and newborn care recommended by WHO were selected as outlined in the measurement improvement roadmap for *Every Newborn*[17,26] and Ending Preventable Maternal Mortality monitoring framework.[27] The *Every Newborn*-Birth Indicators Research Tracking in Hospitals (EN-BIRTH) study was designed to address these evidence gaps by assessing measurement validity for high-priority indicators of newborn and maternal health coverage to inform their use in routine HMIS and population-based surveys for national and global tracking. None of these indicators had previously been validated in routine register data and few in survey data.[28]

Criterion validity testing compares measurement against an objective gold standard to assess whether indicators measure what they intend to, and can provide accurate evidence to inform programmes.[29,30] The EN-BIRTH study protocol outlined four objectives: (1) to assess the validity of numerator measurement, (2) to compare denominators, (3) to evaluate content and quality of care, and (4) to assess barriers and enablers to routine register measurement.[28] We report results on the EN-BIRTH study's first objective for five prioritised core coverage indicators (appendix p 4). Validity testing of other core or additional indicators will be reported separately.

Methods

Study design and participants

EN-BIRTH was a mixed-methods observational study that compared directly observed or verified interventions and practices (considered to be the gold standard) with coverage measured by two different data sources: women's report at exit survey after discharge and hospital routine register records (appendix p 5). The contexts and methods are detailed in a previously published protocol.[28] Five public hospitals providing comprehensive emergency obstetric and neonatal care and including the interventions of interest in the contexts of high mortality burden in sub-Saharan Africa and south Asia were identified in Bangladesh, Nepal, and Tanzania (appendix p 6). Inclusion and exclusion critieria are given in the panel.

Participants gave voluntary informed written consent before recruitment for observation and again for exit surveys. This study was granted ethics approval by institutional review boards in all three countries and by the London School of Hygiene & Tropical Medicine, London, UK (appendix p 7). All collaborating partners have signed data sharing and transfer agreements.

Procedures

In each hospital, consenting participants were recruited in three service delivery contexts to collect data on five selected interventions (panel). Trained researchers in each hospital obtained informed consent and collected participant data prospectively. Data collection varied by site from 7 months to 12 months to achieve the required precision-based sample size (appendix p 8). Clinical observers worked in shifts, covering 24 h each day. Separate groups of data collectors were responsible for observation and verification, exit survey, and routine register data extraction. We used an Android-tablet-based electronic data capture system that was custom built and designed to restrict access by data collector group but linked records at an individual level.[28,31] All data were stored locally on the tablet and synchronised to a country database server managed in Microsoft SQL. A centralised

See Online for appendix

Panel: Service delivery contexts and selected interventions

Labour and delivery ward

(1) Uterotonics to prevent post-partum haemorrhage, (2) early initiation of breastfeeding, and (3) neonatal resuscitation by bag-mask ventilation. Eligible women and their newborn babies were observed while admitted on the labour ward. Exclusion criteria at admission no fetal heartbeat heard or imminent birth.

Kangaroo mother care ward or corner

(4) Skin-to-skin contact or kangaroo mother care position between mother and baby. All mother and baby pairs who were admitted to the kangaroo mother care ward or corner were eligible for observation and exit survey.

Newborn care ward or paediatrics ward

(5) Antibiotic treatment for neonatal infection. Infection diagnosis and name of injectable antibiotic treatment was verified from case notes because observation was not feasible.[30] Eligible neonates were those admitted with clinically defined infection (sepsis, pneumonia, or meningitis). Antibiotic treatment was not documented in routine registers in these hospitals.

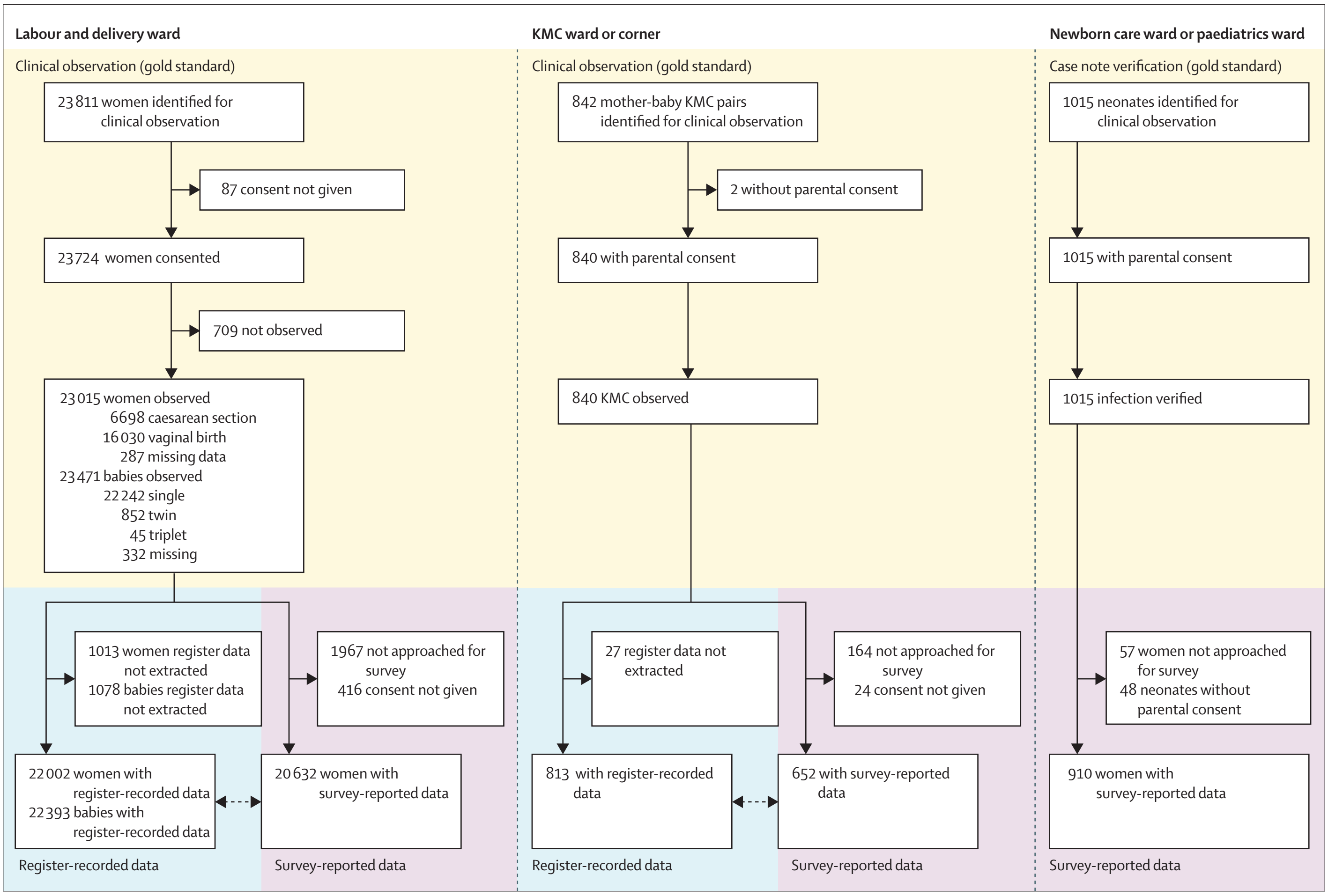

Figure 1: **Flow diagram for EN-BIRTH datasets**
Dashed line indicates methods were compared. KMC=kangaroo mother care.

web-based dashboard was developed to monitor the progress of data capture for selected interventions.

Observation data regarding the interventions and practices were time-stamped, captured in real time by touching a specific data element: once for "observed done" and twice for "observed not done" to override the default "not observed". Women's responses to close-ended questions in the exit surveys regarding the interventions and practices were captured in real time and recorded as "yes", "no", or "don't know, don't remember".

Health workers record each admitted individual woman and baby on one row in the routine registers and record data elements in columns that are either specific columns for that data element or non-specific columns (eg, other details). Research data collectors extracted intervention and practice data from these existing register records after hospital discharge, captured as "yes", "no", "not recorded", or "not readable". Registers were assessed according to two dimensions of data quality, completeness and external consistency with observation. If the register column was blank for an intervention, the column was extracted as "not recorded" in Tanzania and Nepal, but in Bangladesh as "no" (intervention not given) to align with register filling instructions. Hence calculating data element completeness in the Bangladesh registers was not possible. Bangladesh, Nepal, and Tanzania each had different routine register designs, all of which were paper based. The labour ward register design in Bangladesh changed during the study because of national standardisation. In the tables and figures we present data from the revised registers. Original and revised register design and data are shown in the appendix (pp 9–10). In Muhimbili, Tanzania, additional data elements were captured in a long-standing informal perinatal register.

Statistical analysis

Data were anonymised before pooling for analysis using Stata (version 14.2) To assess coverage for the selected indicators, we calculated observed, survey-reported, and register-recorded coverage (appendix p 11). We excluded participants with missing data from their relevant sample. For numerator validation, we used the

	Observed coverage	Survey-recorded coverage	"Don't know" responses	Sensitivity	Specificity	Percent agreement*
Labour and delivery ward† (n=23 471)						
Uterotonics	99·4% (98·7–99·8)	84·7% (79·1–89·5)	8·7% (4·5–14·1)	84·9% (79·6–89·6)	32·5% (21·2–44·6)	84·7% (79·4–89·4)
Early breastfeeding	10·9% (3·8–21·0)	53·2% (39·4–66·8)	0·6% (0·1–1·3)	76·9% (70·7–82·7)	50·0% (32·3–67·7)	53·8% (40·2–67·2)
Neonatal resuscitation (bag-mask ventilation)	4·4% (1·9–8·1)	0·8% (0·4–1·4)	5·9% (2·4–10·7)	9·3% (4·7–15·0)	99·5% (99·2–99·8)	96·0% (93·1–98·1)
Kangaroo mother care ward or corner (n=840)						
Kangaroo mother care	100% (99·9–100)	99·9% (98·3–100)	0·0% (0·0–0·1)	100% (99·8–100)	‡	100% (99·8–100)
Newborn care ward or paediatrics ward (n=1015)						
Antibiotic injection for neonatal infection§	96·7% (94·0–98·6)	74·7% (55·3–90·1)	16·9% (7·4–29·2)	75·9% (55·6–91·6)	‡	75·3% (56·4–90·2)
Named antibiotic for neonatal infection§	96·7% (94·0–98·6)	12·3% (3·5–25·1)	16·9% (7·4–29·2)	12·7% (3·7–25·6)	‡	16·1% (8·0–26·2)

Further individual-level validity statistics and site-specific results by mode of birth are given in the appendix (p 17). *(true positives + true negatives) / n. †Data are for all modes of birth. ‡Specificity not reported because all true negatives not captured. §Verified from case notes.

Table 1: **Individual-level validity testing for survey-recorded coverage versus observed coverage**

simplest denominator (total women observed, total births [livebirths plus stillbirths] observed, kangaroo mother care (KMC) mother–baby pairs observed, or newborn babies treated for infection). A comparison of denominators, including true denominators of clinical need where relevant, will be analysed subsequently. Often, population-based surveys (eg, DHS or MICS) measure coverage from "yes" responses, therefore "don't know" and "no" responses might both be used to suggest no coverage. For registers, monthly data aggregation typically involves counting column ticks, so that "not recorded" is treated as "intervention not given". In our analyses we present these typical scenarios and compare the effect on validity of excluding "don't know" and "not recorded" responses (appendix p 12).

We calculated the absolute differences between observed coverage and exit-survey-reported and register-recorded coverage. Cut-off ranges were adapted from data quality review methods (overestimate or underestimate by 0–5%, 6–10%, 11–15%, 16–20%, and >20%) and used to generate heatmaps.[32] To assess population-level validity for all indicators across both platforms (survey and registers with specific columns) we calculated validity ratios, similar to verification factors in data quality review methods.[32] Validity ratios can be applied when interventions and practices are intended for all women or newborn babies or a small target group. A ratio higher than 1·0 implies overestimation of survey-reported or register-recorded coverage compared with observed, and a ratio lower than 1·0 implies an underestimate. Ratio measure cut-offs used 0·05 increments, defining "excellent" as 0·95–1·05. Measures of individual-level validity were calculated as follows: when two-way table column totals were 10 or more we calculated sensitivity, and if relevant, specificity, negative predictive value, positive predictive value, area under the curve (AUC), and inflation factor. Otherwise we present percent agreement.[28,30] We excluded participants with missing data on a pairwise basis.

All calculations were first done separately for each participating hospital and exact 95% CIs were calculated using the binomial distribution. We then combined the hospital-specific results using a random effects meta-analysis approach (Stata metan command). We calculated I^2 and τ^2 to assess heterogeneity between hospitals. In addition to the protocol planned analyses, because of the increasing proportion of caesarean sections globally, and the many caesarean births in this study, we did analyses stratified by mode of birth (vaginal birth [normal vaginal or vacuum extraction and forceps combined] and caesarean section) to assess any effect on measurement.

STARD guidelines were followed throughout (appendix p 13).

To assess the reliability of our gold standard observation, we calculated Cohen's κ coefficient for the 5% of the sample observed by both supervisors and data collectors.[28] We included all indicators in our analyses, discussing κ scores below high or substantial cut offs, less than 0·71 for observation and less than 0·91 for data extraction, in study limitations. To assess any change in recording practices in routine registers due to study observer presence, we compared absolute differences between completeness of extracted study data with register data from 1 year pre-study collected retrospectively.[22] We also calculated κ coefficients for a 5% sample of double-extracted study register data.

EN-BIRTH is registered with Research Registry, 4833.

Role of the funding source

The funder of the study attended the study design workshop but had no role in data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and had final responsibility for publication submission decision.

Results

Between July 3, 2017, and May 30, 2018, among 23 015 women on the labour and delivery ward, 23 471 births were observed, 20 632 (89·6%) women had an exit survey, and 22 002 (95·6%) had register data

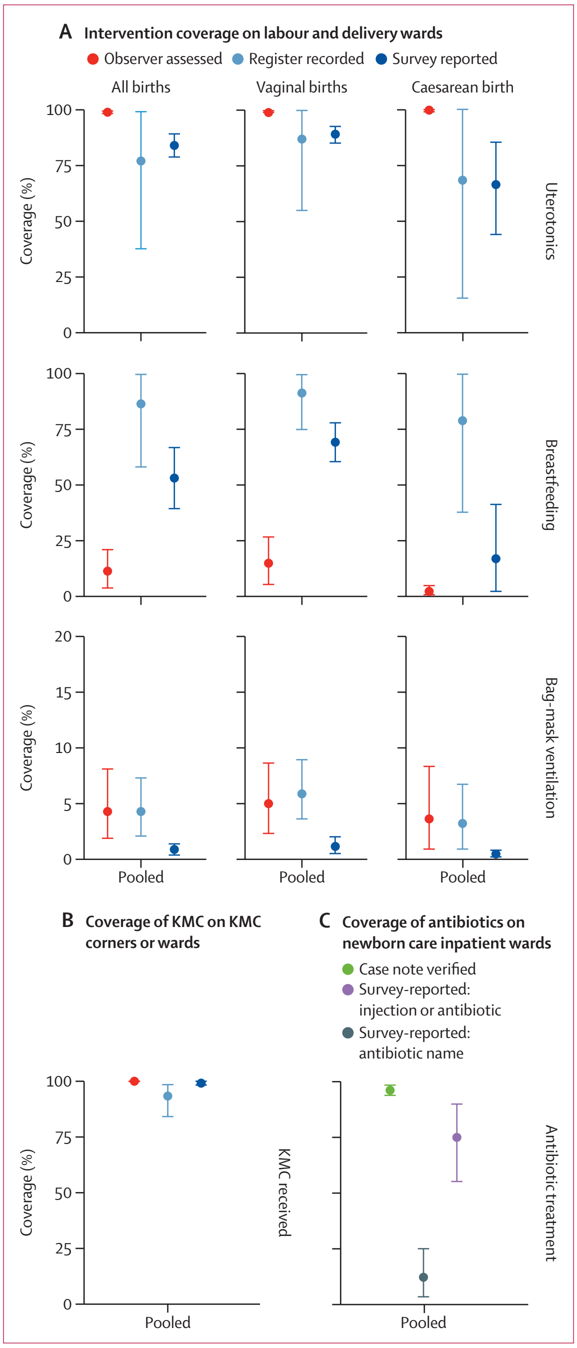

Figure 2: **Coverage for five selected indicators measured by observation, register, and survey, overall and by mode of birth**
Labour and delivery ward (n=23 015; A), KMC (n=840; B), verified neonatal infection (n=1015; C). Error bars show 95% CI. Pooled using random effects meta-analysis. Site-specific results by mode of birth are given in the appendix (pp 21–22). KMC=kangaroo mother care.

extracted (figure 1). Exit survey interviews were done at mean 1·4 days (SD 2·9) after birth (appendix p 8). Consent was not granted by 87 (0·4%) women for observation and 416 (1·8%) for exit survey. 1967 (8·5%) women left the hospital before the survey could be done.

Birth outcomes and background characteristics are shown in the appendix (pp 15–16). Women younger than 19 years comprised 4·6–17·2% of the sample and secondary education completion varied between sites (34·8–61·2%). Caesarean sections were done for 6698 participants, ranging from 7·0% in Temeke, Tanzania, to 72·8% in Azimpur, Bangladesh (appendix p 15–16). The proportion of in-facility stillbirths ranged from three per 1000 to 49 per 1000 total births and the proportion of newborn babies with low birthweight (<2500 g) ranged from 7·4% to 26·5% (appendix p 16).

Survey-reported coverage of uterotonics for prevention of post-partum haemorrhage was 84·7% (95% CI 79·1–89·5), compared with observed coverage of 99·4% (98·7–99·8; table 1). Exit survey heterogeneity was low (τ^2 0·027), with 8·7% (4·5–14·1) of responses given as “don’t know” (appendix p 24). At the individual validity level, sensitivity of survey-reported coverage of uterotonics was 84·9% (79·6–89·6) and specificity 32·5% (21·2–44·6; table 1). Exclusion of “don’t know” responses resulted in sensitivity increasing to 93·5% (91·1–95·5) and specificity decreasing to 17·3% (7·4–29·4; appendix p 24). Survey-reported coverage of uterotonics after caesarean birth was 66·3% (44·0–85·3) compared with observed coverage of 99·5% (98·9–99·9), and after vaginal birth was 89·3% (85·3–92·8) compared with observed coverage of 99·6% (99·1–99·9; figure 2, appendix p 21).

Register-recorded coverage of uterotonics for prevention of post-partum haemorrhage was 77·9% (95% CI 37·8–99·5) with high heterogeneity (τ^2 0·729) between hospitals, compared with observed coverage of 99·2% (98·6–99·7; table 2; appendix pp 17, 24). In the hospital in Pokhara, Nepal, the register had no column to capture uterotonics, all other hospitals used specific columns. Temeke, Tanzania underestimated coverage by 1·7% and Azimpur overestimated by 0·5% (appendix pp 17, 24, 29). However, different hospitals in these countries using the same registers underestimated coverage, by 78·2% in Kushtia and by 33·9% in Muhimbili. At the individual validity level, sensitivity was 78·0% (37·8–99·5) and specificity 22·8% (1·7–53·6). Exclusion of “not recorded” records resulted in sensitivity increasing to 86·1% (48·5–100·0) and specificity decreasing to 3·5% (0·0–17·2). Register-recorded coverage of uterotonics after caesarean birth was 68·5% (15·5–100·0) compared with observed coverage of 99·4% (98·7–99·9), and after vaginal birth was 86·6% (55·0–100·0) compared with observed coverage of 99·4% (98·7–99·9; figure 2; appendix p 24).

Validity ratios for coverage of uterotonics by exit survey were “very good” in Temeke, Tanzania, “good” in Kushtia, Bangladesh, and “moderate” in the remaining three sites. Register validity ratios for coverage of uterotonics were “excellent” in Azimpur and Temeke, but “poor” in Kushtia, and Muhimbili (figure 3).

Among newborn babies who were observed for more than 1 h after birth (n=6304), exit surveys substantially

	Observed coverage	Register-recorded coverage	Not recorded	Not readable	Sensitivity	Specificity	Percent agreement*
Labour and delivery ward (n=23 471)†							
Uterotonics	99·2% (98·6–99·7)	77·9% (37·8–99·5)	3·1% (0·0–19·1)	0·1% (0·0–0·1)	78·0% (37·8–99·5)	22·8% (1·7–53·6)	77·2% (37·7–99·3)
Early breastfeeding	12·5% (4·6–23·6)	85·9% (58·1–99·6)	7·6% (1·1–19·2)	0·1% (0·0–0·2)	97·6% (83·9–100)	6·4% (0·0–29·2)	24·6% (8·5–45·7)
Neonatal resuscitation (bag-mask ventilation)	5·1% (2·0–9·6)	4·3% (2·1–7·3)	65·4% (15·3–99·2)	0·3% (0·2–0·6)	23·6% (7·3–45·2)	96·8% (94·2–98·7)	93·2% (88·0–97·0)
Kangaroo mother care ward or corner (n=840)							
Kangaroo mother care	100% (99·9–100)	92·9% (84·2–98·5)	0·9% (0·2–2·0)	0·0% (0·0–0·3)	93·0% (84·5–98·5)	‡	93·0% (84·5–98·5)

Further individual-level validity statistics and site-specific results by mode of birth are given in the appendix (p 17). *(true positives + true negatives) / n. †Data are for all modes of birth. ‡Specificity not reported because all true negatives not captured. Antibiotic treatment for neonatal infections was not documented in routine registers in these hospitals.

Table 2: **Individual-level validity testing for register-recorded coverage versus observed coverage**

Ratio	Hospital ward	Selected indicator	Azimpur, Bangladesh Tertiary	Kushtia, Bangladesh District	Pokhara, Nepal Regional	Temeke, Tanzania Regional	Muhimbili, Tanzania National	All sites pooled (95% CI)
Survey to observed	A. Labour ward	1 Uterotonics to prevent post-partum haemorrhage	0·81	0·88	0·82	0·92	0·81	0·85 (0·80–0·91)
		2 Early initiation of breastfeeding	27·40	6·70	10·13	2·68	1·43	2·92 (1·58–5·38)
		3 Neonatal Resuscitation - Bag-mask ventilation	0·60	0·30	0·14	0·10	0·14	0·20 (0·12–0·34)
	B. KMC ward	4 KMC	0·89	1·00	1·00	1·00	1·00	1·00 (0·99–1·01)
	C. Neonatal ward	5.1 Neonatal infection antibiotics - injection	0·83	0·60	0·49	0·92	0·96	0·74 (0·55–1·00)
		5.2 Neonatal infection antibiotics - name	0·05	0·26	0·03	0·22	0·15	0·11 (0·05–0·25)
Register to observed	A. Labour ward	1 Uterotonics to prevent post-partum haemorrhage	1·01	0·22		0·98	0·66	0·61 (0·45–0·84)
		2 Early initiation of breastfeeding	50·40	9·84		3·67	2·30	4·29 (7·22–7·25)
		3 Neonatal resuscitation - Bag-mask ventilation	1·14	1·27		0·75	0·55	0·85 (0·59–1·23)
	B. KMC ward	4 KMC	1·00	0·98		0·85	0·85	0·92 (0·82–1·03)
	C. Neonatal ward	5 Neonatal infection antibiotics/injection						

<0·80 or >1·20 poor; 0·80 to 0·84 or 1·16 to 1·20 moderate; 0·85 to 0·89 or 1·11 to 1·15 good; 0·90 to 0·94 or 1·06 to 1·10 very good; 0·95 to 0·99 or 1·00 to 1·05 excellent; Data not captured in specific column in register

Figure 3: **Heatmap for selected indicator validity ratios**
Validity ratios study by sites and pooled (heatmap cut offs 5%, 10%, 15%, and 20%). Pooled using random effects meta-analysis. KMC=kangaroo mother care. Cut-off ranges adapted from WHO Data Quality Review, Module 2.[32]

overestimated the coverage of breastfeeding initiated within 1 hour after birth compared with observed coverage (table 1). Exit survey heterogeneity was low (τ^2 0·101), with 0·6% (95% CI 0·1–1·3%) of responses “don’t know” (appendix pp 17, 31, 36). Survey-reported coverage of early breastfeeding after caesarean birth was 17·1% (2·3–41·3), compared with observed coverage of 2·4% (1·2–3·9) and 69·5% (60·5–77·9) after vaginal birth, compared with observed coverage of 14·4% (5·4–26·7; figure 2, appendix pp 31, 36).

Register-recorded coverage also substantially over-estimated early breastfeeding compared with observed coverage (table 2), with high heterogeneity between hospitals (τ^2 0·423; appendix pp 31, 36). Register-recorded coverage of early breastfeeding after caesarean birth was 78·3% (95% CI 37·8–99·7), compared with observed coverage of 2·2% (0·9–4·0) and after vaginal birth was 91·4% (74·9–99·5) compared to observed 17·3% (8·0–29·1; figure 2, appendix pp 31, 36). Validity ratios for both survey and registers were categorised as “poor” in all sites (figure 2).

Exit survey-reported coverage underestimated bag-mask ventilation compared with observed coverage, using a total birth denominator (table 1). Exit survey heterogeneity was low (τ^2 0·003) with 5·9% (95% CI 2·4–10·7) of responses “don’t know” (appendix pp 18, 38, 43). Sensitivity was 9·3% (4·7–15·0) and specificity was 99·5% (99·2–99·8; table 1). Exclusion of “don’t know” responses resulted in sensitivity increasing slightly to 12·5% (6·5–19·9) with no decrease in specificity (appendix pp 38, 43). Survey-reported coverage of bag-mask ventilation after caesarean birth was 0·4% (0·2–0·8), compared with observed coverage of 3·7% (0·9–8·3), and after vaginal birth was 1·1% (0·5–2·0), compared with observed coverage of 5·0% (2·3–8·6) (figure 2, appendix pp 38, 43).

Register-recorded coverage underestimated bag-mask ventilation, with low heterogeneity (τ^2 0·017), compared

A Labour and delivery ward

		Azimpur, Bangladesh Tertiary	Kushtia, Bangladesh District	Pokhara, Nepal Regional	Temeke, Tanzania Regional	Muhimbili, Tanzania National
1. Uterotonics to prevent post-partum haemorrhage						
Register design: column allotted data element		Specific column	Specific column	No column	Specific 2 columns	Specific 2 columns
Completeness	Data element recorded in register	Not possible	Not possible		99·2%	68·6%
External Consistency	Indicator: observed coverage %	98·9%	99·8%		99·3%	98·4%
	Indicator: measured coverage, register recorded %	99·4%	21·6%		97·6%	64·5%
	Measurement gap: register recorded and observed	0·6% underestimate	78·2% underestimate		1·7% underestimate	34·0% underestimate
2. Early Initiation of breastfeeding (observed > 1 h)						
Register design: column allotted data element		specific column	specific column	No column	specific 2 columns	specific 2 columns
Completeness	Data element recorded in register	Not possible	Not possible		97·7%	76·6%
External Consistency	Indicator: observed coverage %	1·8%	9·8%		26·0%	19·1%
	Indicator: measured coverage, register recorded %	91·7%	96·8%		95·3%	43·8%
	Measurement gap: register recorded and observed	89·9% overestimate	87·0% overestimate		69·3% overestimate	24·7% overestimate
3. Neonatal resuscitation (bag-mask ventilation)						
Register design: column allotted data element		Specific column	Specific column	No column	Specific 2 columns	Specific 2 columns
Completeness	Data element recorded in register	Not possible	Not possible		91·1%	55·7%
External Consistency	Indicator: observed coverage %	0·8%	6·1%		7·1%	9·0%
	Indicator: measured coverage, register recorded %	0·9%	7·7%		5·3%	5·0%
	Measurement gap: register recorded and observed	0·1% overestimate	1·6% overestimate		1·8% underestimate	4·0% underestimate

B Kangaroo mother care ward or corner

		Azimpur, Bangladesh Tertiary	Kushtia, Bangladesh District	Pokhara, Nepal Regional	Temeke, Tanzania Regional	Muhimbili, Tanzania National
4. Kangaroo mother care						
Register design: column allotted data element		Specific column	Specific column	Non-specific	Specific 2 columns	Specific 2 columns
Completeness	Data element recorded in register	100·0%	98·5%	93·0%	99·1%	98·8%
External Consistency	Indicator: observed coverage %	100·0%	99·9%	99·9%	99·8%	99·5%
	Indicator: measured coverage, register recorded %	100·0%	97·8%	21·2%	84·8%	85·2%
	Measurement gap: register recorded and observed	0·0% accurate	2·1% overestimate	78·7% underestimate	15·0% underestimate	14·3% underestimate

C Newborn care ward or paediatrics ward

	Azimpur, Bangladesh Tertiary	Kushtia, Bangladesh District	Pokhara, Nepal Regional	Temeke, Tanzania Regional	Muhimbili, Tanzania National
5. Neonatal Infection Antibiotic Treatment					
Register design	No column	No column	No column	No column	No column

No column for data element · Non-specific column for data element · Specific column
>20% poor · 16–20% moderate · 11–15% good · 6–10% very Good · 0–5% excellent

Figure 4: **Routine register design for EN-BIRTH study sites and accuracy of data quality dimensions**
Cut-off ranges adapted from WHO Data Quality Review, Module 2.[32] Completeness calculations were denoted "not possible" for Bangladesh registers because the register was designed to be left blank if the intervention or practice was not done. An expanded version of this figure is shown in the appendix (p 9).

with observed coverage (table 2; appendix pp 18, 38). Register column design varied (figure 4). In Bangladesh, the column was ticked when bag-mask ventilation was done and was otherwise left blank, and coverage was slightly overestimated by 0·1–1·6% (figure 4). In Tanzania, a numerical code ("3" for bag-mask ventilation) or "no" was written in the column, and completeness was 55·7% for Muhimbili and 91·1% for Temeke (appendix pp 38, 43). The Pokhara register did not capture this data element. Sensitivity was 23·6% (95% CI 7·3–45·2) and specificity was 96·8% (94·2–98·7; table 2). Exclusion of "not recorded" records resulted in sensitivity increasing to 53·6% (28·1–78·1) and specificity decreasing to 77·7% (57·9–92·5; appendix pp 38, 43). Register-recorded coverage of bag-mask ventilation after caesarean birth was 3·2% (0·9–6·7), compared with observed coverage of 4·1% (0·6–10·4), and after vaginal birth register-recorded coverage of bag-mask ventilation was 5·9% (3·6–8·9), compared with observed coverage of 5·9% (3·3–9·2; figure 2, appendix pp 38, 43). Survey validity ratios were all categorised as "poor" (figure 3). For register validity ratios, Azimpur was "good" but all other sites were "poor" (figure 3).

In KMC wards or corners, 840 mother–baby pairs, including 91 babies who were born outside of the study hospital, were observed, with 652 exit surveys done and 813 routine register records extracted (figure 1). 815 (97%) of KMC babies had a birthweight of 2000 g or less (WHO recommendation or KMC). Background characteristics are shown in the appendix (p 45). Pre-discharge mortality was low (0·0–1·8%; appendix p 45).

Exit survey-reported coverage of KMC was 99·9% (95% CI 98·3–100), compared with observed coverage of 100% (99·9–100), and sensitivity was 100% (99·8–100), with zero "don't know" responses (table 1). KMC coverage was captured in specific national standardised KMC ward registers, except in Nepal, which used the general child register (figure 4). Register-recorded coverage from standardised specific KMC registers was 92·9% (84·2–98·5), with low heterogeneity (τ^2 0·065), compared with observed coverage of 100% (99·9–100) and sensitivity of 93·0% (84·2–98·5; table 2; appendix pp 19, 46, 48). Validity ratios for KMC by exit surveys were categorised as "excellent" for all sites, except Azimpur, which was "good", and for registers validity ratios were "excellent" in the Bangladesh sites and good in the Tanzania sites (figure 3).

In newborn care inpatient wards, among 1532 newborn babies, a diagnosis of clinically defined infection (sepsis, pneumonia, or meningitis) was documented in individual case notes for 1015 (66·6%) neonates and 910 exit surveys were done (figure 1). Background characteristics of these neonates are shown in the appendix (p 49). Coverage of antibiotic treatment verified by case notes (gold standard) was 96·7% (95% CI 94·0–98·6; table 1). Exit-survey reported coverage was 74·7% (55·3–90·1) when measured using a general question regarding injection or antibiotic use, with sensitivity of 75·9% (55·6–91·6) and high heterogeneity (τ^2 0·204; appendix pp 19, 50, 52). The proportion of "don't know" responses was high (table 1). When adding the question regarding the name of the antibiotic, survey-reported coverage dropped to 12·3% (3·5–25·1; table 1; appendix pp 20, 50, 52). Validity ratios for the general question (injection or antibiotic) ranged from excellent in Muhimbili to poor in Pokhara; for the more specific question (antibiotic name), ratios were poor in all sites (figure 3).

Analysis of inter-rater reliability for gold standard data showed high or substantial κ scores for most data elements but moderate scores for observed uterotonic coverage in Temeke, and early breastfeeding in Nepal and Temeke (appendix pp 53–54). Lower κ scores were found for both KMC and verification of antibiotics in Pokhara (appendix p 51). Inter-rater reliability for routine register data was lower than the high or substantial cut offs for all labour ward indicators in Kushtia, Temeke, and Muhimbili; and for KMC in Pokhara (appendix pp 55–56). Register completeness comparison, before and during the study, revealed decreases of more than 5% for bag-mask ventilation coverage in both Tanzanian sites and for uterotonic coverage in Muhimbili. Breastfeeding completeness increased in Muhimbili from 0 to 99·4% (appendix p 54).

Discussion

We examined validity of coverage indicators for selected maternal and newborn care indicators in two data systems: exit surveys and routine registers. Surveys are highly standardised in question design and interview technique. Registers are variable in design and filling techniques. We found much higher heterogeneity for register-recorded coverage compared with exit-survey reported coverage. Even with the same register design, accuracy varied between hospitals, with good validity for the highest performing sites. We stratified data by mode of birth and found that caesarean birth affected measurement in surveys and registers. With rising caesarean rates, especially in LMICs,[33] this finding needs further consideration. Register data that are aggregated for use are typically located in the labour and delivery ward, but with high caesarean section rates, specific registers in operating theatres might be necessary. For survey reports after caesarean birth, low accuracy might relate to not seeing an intervention happening (eg, whether or not the baby had bag-mask ventilation) but might also be a reflection on gaps in respectful care; women have a right to communication and informed consent regarding care for themselves or their newborn babies.[34,35]

Survey report had low accuracy for indicators of clinical intervention coverage led by health workers around the time of birth, notably neonatal care, and to a lesser extent for uterotonics. This study is the first to test the validity of survey reporting of indicators of care for small and ill newborn babies. We found that survey reporting of bag-mask ventilation coverage had low accuracy and reporting of neonatal infection treatment with antibiotics had low sensitivity among the target group. By contrast, KMC, which is led by the woman had high sensitivity in exit surveys, with potential for further testing, including report from women did not practice KMC. Use in population-based surveys would require sufficient sample size for the target group of small babies requiring KMC. For neonatal infection, even interviewing only women whose babies had been admitted, we found high proportions of "don't know" responses and underestimates of observed coverage. A high proportion of "don't know" responses suggests that the survey question is a poor way of measuring intervention coverage. We reported "don't know" as "no" in line with the practice of the DHS and

MICS for yes or no questions, but we note that sensitivity increased when "don't know" responses were excluded for uterotonic and bag-mask ventilation coverage. Our exit survey had a recall period of a few days, but most household surveys cover the previous 2–5 years; other studies have found recall decay even by 1 year.[36] Hence, these results are likely to be worse in a routine household survey.

Early breastfeeding rates were observed to be very low across study hospitals, particularly after caesarean. The early initiation of breastfeeding indicator is already measured in the DHS and MICS household surveys, and national and global tracking rely on these data.[37,38] Our results indicate that both women's report and routine register data substantially overestimate coverage. This finding is in agreement with previous survey validation studies.[12–15] One register validation study for early breastfeeding in a composite indicator of essential newborn care also showed overestimation.[24] We postulate three explanations for these overestimates: first, this finding might be due to the timing component (ie, breastfeeding early but not within 1 h) being misreported by the woman or health worker. Second, successful initiation of breastfeeding is a process that involves the baby being put to the breast, then attachment, and then sucking. Putting the baby to the breast is one important step and is the focus of the survey-reported question, but might not have been considered as initiation by the observer. The observers were trained to click the stamp on the tablet at the point of initiation and recording is likely to vary given the challenges of observing this complex and dynamic process. Third, the overestimate might be due to social desirability bias among women to over-report in surveys or professional desirability pressure on the health worker to over-record in registers. More work is needed to improve measurement of this crucial indicator, including exploring whether changing the timing component could increase accuracy in surveys and registers. In Muhimbili, before and during the study, register data completeness for breastfeeding increased from 0 to 99%. This finding was probably due to the data being extracted from the informal perinatal register before the study, rather than the formal labour ward register, and highlights how complex documentation systems affect measurement.

Registers have unrealised potential as a useful data source, shown by the high accuracy and sensitivity for indicators in some EN-BIRTH study hospitals. Although hospital registers can only capture a limited number of data elements, we found that register designs in Tanzania and Bangladesh already have the relevant numerator or count data for selected indicators of maternal and newborn health coverage.[22] Many data are already being collected by frontline health workers. Register data were highly complete, and although data collectors rarely indicated data were not readable, we found low inter-rater κ results across register recorded data. Extracting data for aggregation is a crucial step for data flowing to higher levels in the health system, and more research is needed to inform data extraction quality. The accuracy of register-recorded coverage varied between hospitals even with identical register design, reflecting variation in implementation and data culture.[39] In both Bangladesh hospitals, register-recorded coverage increased in the revised registers when specific columns for data elements replaced absent or non-specific columns in original registers; however, change in accuracy varied by indicator. Sensitivity increased (and specificity decreased) when "not recorded" data were excluded from analyses (eg, uterotonics and bag-mask ventilation). Further work on register design for high quality monthly data aggregation, as well as feedback after use, is needed.[40] Qualitative research might help to understand the differences in these five hospitals by exploring barriers and enablers to routine documentation, specifically for register design, register filling, and register use.[28]

EN-BIRTH study strengths include multi-country sites, rigorous observational design, and large sample size (about 10-times higher than that in previous validation studies[12–15,24]) targeting previously missed clinical care groups, especially small and ill newborn babies—a priority in universal health coverage measurement.[16] Errors were minimised in observation data by the tablet-based app design, which was custom-built and user friendly, with major effort invested in being able to navigate between the recording of simultaneous events for the woman and baby with minimal delay.[31] The tablet app also importantly captured time-stamped data for time-sensitive interventions, around the time of birth when complications might mean that women can die within hours and babies within minutes. Detailed analyses, including quality of care with timing data, are published separately.[41–45] Dual observation by supervisors showed high or substantial agreement for most data elements, with breastfeeding coverage scoring the least well across all sites, probably due to challenges of capturing the process. Information bias during data collection was reduced by using different data collector groups for observation, exit surveys, and register data extraction. The effect on register recording completeness from the presence of researchers was assessed by comparison with register data extraction before the study.[22] As per protocol, we did not base our assessment of validity using AUC cut offs because our data were all binary (yes or no) for the coverage estimates; thus, the AUC is simply the average of the sensitivity and specificity.[28] I^2 estimates the percentage of variation that is attributable to study heterogeneity. For an intervention like uterotonics that was almost universally applied, there is little total variation so that even small differences between sites result in a large I^2. We therefore chose to place more emphasis in τ^2, which provides an estimate of the magnitude of the between-site variation. A small value of τ^2 indicates little absolute variation between sites even when I^2 might appear large.

However, our study also had limitations. Despite large samples, we did not achieve ten or more column counts in the two-way tables for all indicators, either because interventions had very high coverage (eg, correctly provided for all women and babies) or very low prevalence (eg, small clinical target group). This affected our ability to report on individual-level validity metrics. By contrast with other measurement validation studies, EN-BIRTH chose to use vigorous subset double observation inter-observer κ calculations to assess possible between-site variation in validation results. In the protocol, we pre-defined ranges for what would be considered high or substantial agreement. κ scores were lower than expected for some indicators in some sites. We postulate that this finding might be a real reflection of inter-observer variation, and we suspect this to be the case, for example, regarding breastfeeding in Temeke, with the lowest percent agreement and a low κ. We note the dependence of κ on prevalence with paradoxically low κ scores, due to the imbalance in marginal totals or with perfect symmetry in the imbalance, as our results showed.[46] The tablet app was not available for pre-study extraction of register data, and the different data collection methods might account for some of the differences in completeness before and during the study.

The five hospitals were high-volume public comprehensive emergency obstetric and neonatal care hospitals in cities and results might not be generalisable to lower-level or rural facilities with lower volumes. The study sample might have been healthier than is typical in such hospitals because recruitment excluded antepartum intrauterine fetal death and women who were too ill to consent after admission. This observational study was not designed to capture true denominators in terms of the need of intervention required for coverage indicators—eg, for accurate diagnosis of neonatal sepsis in terms of microbiological culture or molecular diagnosis. Although blood culture is still considered the definitive diagnostic method, even with excellent laboratory capacity, only about a third of neonatal sepsis cases have a positive culture result. The focus of this study was validity of routine health system data, especially of the clinical diagnosis, and was not addressing the need for better laboratory systems, which is also crucial.[45,47]

Time pressure on health workers in these busy hospitals might have affected their ability to deliver and communicate care to women, affecting exit survey reports and register documentation.[48] We consider this real-world time pressure to be present in many such contexts. Women's reports were collected at hospital exit, close in time to the event and without the typical 2–5 year gap in household surveys. If women cannot report accurately at exit from the facility, they are unlikely to report more accurately later in a household survey, so exit survey findings are relevant when considering adding questions to household surveys.[49] The EN-SMILING study is following up with the cohort, with the potential to repeat interviews in 5 years to investigate the change in women's report.[28]

Policymakers and programme managers require information to inform investments and programmatic course correction. Surveys are important in most LMICs for population-based outcome data and contact coverage of care, but given the high rate of "don't know" responses and low accuracy for the reporting of clinical interventions, adding these indicators to surveys is not justifiable. Because approximately 80% of births worldwide are now in facilities, standardising register design and linked information systems have the potential to sustainably improve data quality for care at birth. Routine register data can be accurate and health workers' time investment would not be wasted if these data were better used. Well-designed, standardised registers are important, and could reduce the burden on health workers of duplicative, or non-valid, data collection.[22] If interventions and practices are defined with a timing element (eg, early initiation of breastfeeding), this inclusion needs consideration in register design. The timing component of the uterotonics coverage indicator is not yet clearly defined, which will limit comparability if routinely measured. EN-BIRTH tested validity of measurement in paper-based registers at the interface with women and their babies. A further phase will explore the feasibility of these indicators flowing up through the national routine data systems, many of which are being rapidly digitised.[28] Another important research gap is how to best measure experience of care, including respectful care in all settings, and the provision and experience of care for women and newborn babies in fragile and humanitarian settings.[35,50,51]

Valid routine data alone will not save lives. Data need to be used by health-care professionals caring for women and their babies, and by policy makers and governments to invest and transform care, enabling universal health coverage as a reality that can be measured and improved.

Contributors
JEL conceived the EN-BIRTH study, acquired the funding and led the overall design. QS-uR was the main lead for analyses, working closely with LTD and the EN-BIRTH team. LTD and JEL drafted the manuscript with the principal investigators: SEA, AKC, HM, AER, NS, and country teams. All authors made substantial contributions to conception, design, data collection, analysis, or interpretation of data. GRG-L did the analysis for register recording practice comparison before and during the study. All authors revised the manuscript and gave final approval for publication and agree to be accountable for the work. Collaborative authors including the Expert Advisory Group made contributions to conception, design, data collection, analysis, or interpretation of data. The authors' views are their own, and not necessarily from any of the institutions they represent.

EN-BIRTH validation collaborative group
Bangladesh—Ayub Ali, Bilkish Biswas, Rajib Haider, Abu Hasanuzzaman, Amir Hossain, Ishrat Jahan, Rowshan Hosne Jahan, Jasmin Khan, M A Mannan, Tapas Mazumder, Hafizur Rahman, Ziaul Haque Shaikh,

Aysha Siddika, Taslima Akter Sumi, and Taqbir Us Samad Talha. *Tanzania*—Evelyne Assenga, Claudia Hanson, Edward Kija, Rodrick Kisenge, Karim Manji, Fatuma Manzi, Namala Mkopi, Mwifadhi Mrisho, and Andrea Pembe. *Nepal*—Jagat Jeevan Ghimire, Rejina Gurung, Elisha Joshi, Avinash K Sunny, Naresh P KC, Nisha Rana, Shree Krishna Shrestha, Dela Singh, Parashu Ram Shrestha, and Nishant Thakur. *London School of Hygiene & Tropical Medicine*—Hannah Blencowe and Sarah G Moxon.

EN-BIRTH Expert Advisory Group

Agbessi Amouzou, Tariq Azim, Debra Jackson, Theopista John Kabuteni, Matthews Mathai, Jean-Pierre Monet, Allisyn Moran, Pavani Ram, Barbara Rawlins, Jennifer Requejo, Johan Ivar Sæbø, Florina Serbanescu, and Lara Vaz.

Declaration of interests

We declare no competing interests.

Data sharing

For the **study data** see https://datacompass.lshtm.ac.uk/955/

The datasets generated during and analysed during the study are available on the London School of Hygiene & Tropical Medicine Data Compass repository.

Acknowledgments

This study was funded by The Children's Investment Fund Foundation through a grant awarded to JEL at the London School of Hygiene & Tropical Medicine. The Swedish Research Council funded the Nepal site through Lifeline Nepal and Golden Community. We acknowledge the core funders for all the partner institutions. We credit the inspiration of the late Godfrey Mbaruku. We thank Claudia DaSilva, Veronica Ulaya, Mohammad Raisul Islam, Sudip Karki, and Rabina Karki for their administrative support and Sabrina Jabeen, Goutom Banik, Shahidul Alam, Tamatun Islam Tanha, and Moshiur Rahman for support during data collector training. We also thank Ann Blanc, Emily Carter, and Melinda Munos for sharing relevant technical inputs and expertise. We acknowledge the National Advisory Groups: Mohammod Shahidullah, Muhiuddin Osmani, Sabina Ashrafee Lipi, and Jahurul Islam in Bangladesh; Naresh P KC in Nepal; and Muhammad Bakari Kambi, Georgina Msemo, Talhiya Yahya, Claud Kumalija, Eliudi Eliakimu, and Honest Kimaro in Tanzania. This paper is published with permission from the Directors of Ifakara Health Institute, Muhimbili University of Health and Allied Sciences, International Centre for Diarrhoeal Disease Research, Bangladesh, and Golden Community. We are also very grateful to fellow researchers who peer-reviewed this paper. Finally, and most importantly, we also thank the women, their families, the health workers and data collectors (appendix p 57).

References

1 WHO. Trends in maternal mortality 2000 to 2017 estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division 2019. Geneva: World Health Organization, 2019. https://www.who.int/reproductivehealth/publications/maternal-mortality-2000-2017/en/ (accessed May 20, 2020).

2 UNICEF. Levels and trends in child mortality report 2019. https://www.unicef.org/reports/levels-and-trends-child-mortality-report-2019 (accessed May 18, 2020).

3 Lawn JE, Blencowe H, Oza S, et al. Every Newborn: progress, priorities, and potential beyond survival. *Lancet* 2014; **384:** 189–205.

4 WHO. Global strategy for women's, children's and adolescents health, 2016–2030. Geneva, World Health Organization, 2015. http://www.who.int/life-course/publications/global-strategy-2016-2030/en/ (accessed May 18, 2020).

5 UNICEF. The State of the World's Children 2019: statistical tables, https://www.unicef.org/media/63016/file/SOWC-2019.pdf (accessed June 11, 2020).

6 Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. *Lancet Glob Health* 2018; **6:** e1196–252.

7 Bhutta ZA, Das JK, Bahl R, et al. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? *Lancet* 2014; **384:** 347–70.

8 Miller S, Abalos E, Chamillard M, et al. Beyond too little, too late and too much, too soon: a pathway towards evidence-based, respectful maternity care worldwide. *Lancet* 2016; **388:** 2176–92.

9 Marsh AD, Muzigaba M, Diaz T, et al. Effective coverage measurement in maternal, newborn, child, and adolescent health and nutrition: progress, future prospects, and implications for quality health systems. *Lancet Glob Health* 2020; **8:** e730–36.

10 Carvajal-Aguirre L, Vaz LM, Singh K, et al. Measuring coverage of essential maternal and newborn care interventions: an unfinished agenda. *J Glob Health* 2017; **7:** 020101.

11 Marchant T, Bryce J, Victora C, et al. Improved measurement for mothers, newborns and children in the era of the Sustainable Development Goals. *J Glob Health* 2016; **6:** 010506.

12 Stanton CK, Rawlins B, Drake M, et al. 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* 2013; **8:** e60694.

13 Blanc AK, Warren C, McCarthy KJ, Kimani J, Ndwiga C, RamaRao S. Assessing the validity of indicators of the quality of maternal and newborn health care in Kenya. *J Glob Health* 2016; **6:** 010405.

14 Blanc AK, Diaz C, McCarthy KJ, Berdichevsky K. Measuring progress in maternal and newborn health care in Mexico: validating indicators of health system contact and quality of care. *BMC Pregnancy Childbirth* 2016; **16:** 255.

15 McCarthy KJ, Blanc AK, Warren CE, Kimani J, Mdawida B, Ndwidga C. Can surveys of women accurately track indicators of maternal and newborn care? A validity and reliability study in Kenya. *J Glob Health* 2016; **6:** 020502.

16 Munos MK, Stanton CK, Bryce J. Improving coverage measurement for reproductive, maternal, neonatal and child health: gaps and opportunities. *J Glob Health* 2017; **7:** 010801.

17 Moxon SG, Ruysen H, Kerber KJ, et al. Count every newborn; a measurement improvement roadmap for coverage data. *BMC Pregnancy Childbirth* 2015; **15** (suppl 2): S8.

18 Mate KS, Bennett B, Mphatswe W, Barker P, Rollins N. Challenges for routine health system data management in a large public programme to prevent mother-to-child HIV transmission in South Africa. *PLoS One* 2009; **4:** e5483.

19 Chiba Y, Oguttu MA, Nakayama T. Quantitative and qualitative verification of data quality in the childbirth registers of two rural district hospitals in Western Kenya. *Midwifery* 2012; **28:** 329–39.

20 Duffy S, Crangle M. Delivery room logbook - fact or fiction? *Trop Doct* 2009; **39:** 145–49.

21 Maternal and Child Survival Program. What data on maternal and newborn health do national health management information systems include? A review of data elements for 24 low- and lower middle income countries. 2018. https://www.mcsprogram.org/resource/what-data-on-maternal-and-newborn-health-do-national-health-management-information-systems-include (accessed May 18, 2020).

22 Day LT, Gore-Langton GR, Rahman AE, et al. Labour and delivery ward register data availability, quality, and utility - Every Newborn - birth indicators research tracking in hospitals (EN-BIRTH) study baseline analysis in three countries. *BMC Health Serv Res* 2020; **20:** 737.

23 Broughton EI, Ikram AN, Sahak I. How accurate are medical record data in Afghanistan's maternal health facilities? An observational validity study. *BMJ Open* 2013; **3:** e002554.

24 Bhattacharya AA, Allen E, Umar N, et al. Monitoring childbirth care in primary health facilities: a validity study in Gombe State, northeastern Nigeria. *J Glob Health* 2019; **9:** 020411.

25 WHO, UNICEF. Every Newborn: an action plan to end preventable deaths. Geneva: World Health Organization, 2014. https://apps.who.int/iris/handle/10665/127938 (accessed May 18, 2020).

26 WHO. WHO technical consultation on newborn health indicators: Every Newborn Action Plan metrics, Ferney Voltaire, France, 3–5 December 2014. Geneva: World Health Organization, 2015. https://apps.who.int/iris/handle/10665/184225 (accessed May 18, 2020).

27 Moran AC, Jolivet RR, Chou D, et al. A common monitoring framework for ending preventable maternal mortality, 2015–2030: phase I of a multi-step process. *BMC Pregnancy Childbirth* 2016; **16:** 250.

28 Shamba D, Day LT, Zaman SB, et al. Barriers and enablers to labour ward register data collection and use: EN-BIRTH multi-country validation study. *BMC Pregnancy Childbirth* 2020 (in press).

29 Benova L, Moller A-B, Moran AC. "What gets measured better gets done better": the landscape of validation of global maternal and newborn health indicators through key informant interviews. *PLoS One* 2019; **14:** e0224746.

30 Munos MK, Blanc AK, Carter ED, et al. Validation studies for population-based intervention coverage indicators: design, analysis, and interpretation. *J Glob Health* 2018; **8:** 020804.

31 Ruysen H, Rahman AE, Gordeev VS, et al. Electronic data collection for multi-country, hospital-based, clinical observation of maternal and newborn care: EN-BIRTH study experiences. *BMC Pregnancy Childbirth* 2020 (in press).

32 WHO. Data quality review: a toolkit for facility data quality assessment. Module 2: Desk review of data quality. Geneva: World Health Organization, 2017. https://apps.who.int/iris/handle/10665/259225 (accessed May 18, 2020).

33 Boerma T, Ronsmans C, Melesse DY, et al. Global epidemiology of use of and disparities in caesarean sections. *Lancet* 2018; **392:** 1341–48.

34 WHO. Standards for improving quality of maternal and newborn care in health facilities. Geneva: World Health Organization, 2016.

35 Bohren MA, Mehrtash H, Fawole B, et al. How women are treated during facility-based childbirth in four countries: a cross-sectional study with labour observations and community-based surveys. *Lancet* 2019; **394:** 1750–63.

36 McCarthy KJ, Blanc AK, Warren CE, Kimani J, Mdawida B, Ndwidga C. Can surveys of women accurately track indicators of maternal and newborn care? A validity and reliability study in Kenya. *J Glob Health* 2016; **6:** 020502.

37 Victora CG, Bahl R, Barros AJD, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *Lancet* 2016; **387:** 475–90.

38 Fanzo J, Hawkes C, Udomkesmalee E, et al. 2018 global nutrition report: shining a light to spur action on nutrition. 2018. https://globalnutritionreport.org/reports/ (accessed May 18, 2020).

39 Lippeveld T, Sauerborn R, Bodart C. Design and implementation of health information systems. 2000. https://apps.who.int/iris/handle/10665/42289 (accessed May 18, 2020).

40 Braa J, Heywood A, Sahay S. Improving quality and use of data through data-use workshops: Zanzibar, United Republic of Tanzania. *Bull World Health Organ* 2012; **90:** 379–84.

41 Ruysen H, Shabani J, Hanson C, et al. Uterotonics for prevention of postpartum hemorrhage: EN-BIRTH multi-country study. *BMC Pregnancy Childbirth* (in press).

42 Tahsina T, Hossain AT, Ruysen H, et al. Immediate newborn care and breastfeeding: EN-BIRTH multi-country study. *BMC Pregnancy Childbirth* (in press).

43 Ashish KC, Peven K, Ameen S, et al. Neonatal resuscitation: EN-BIRTH multi-country validation study. *BMC Pregnancy Childbirth* (in press).

44 Salim N, Shabani J, Peven K, et al. Kangaroo mother care: EN-BIRTH multi-country validation study. *BMC Pregnancy Childbirth* 2020 (in press).

45 Rahman AE, Hossain AT, Ameen S, et al. Antibiotic use for inpatient newborn care with suspected infection: EN-BIRTH multi-country validation study. *BMC Pregnancy Childbirth* 2020 (in press).

46 Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; **43:** 543–49.

47 Fitchett EJA, Seale AC, Vergnano S, et al. Strengthening the Reporting of Observational Studies in Epidemiology for Newborn Infection (STROBE-NI): an extension of the STROBE statement for neonatal infection research. *Lancet Infect Dis* 2016; **16:** e202–13.

48 MEASURE Evaluation. Easing the Data collection burden on healthcare providers. 2017. https://www.measureevaluation.org/resources/publications/tr-17-211 (accessed May 30, 2020).

49 Ameen S, Siddique AB, Peven K, et al. Survey of women's report for 33 maternal and newborn indicators: EN-BIRTH multi-country validation study. *BMC Pregnancy and Childbirth* 2020 (in press).

50 Tunçalp Ö, Were WM, MacLennan C, et al. Quality of care for pregnant women and newborns—the WHO vision. *BJOG* 2015; **122:** 1045–49.

51 Afulani PA, Phillips B, Aborigo RA, Moyer CA. Person-centred maternity care in low-income and middle-income countries: analysis of data from Kenya, Ghana, and India. *Lancet Glob Health* 2019; **7:** e96–109.