BMJ open
accessible medical research

How accurate are medical record data in Afghanistan's maternal health facilities? An observational validity study

Edward I Broughton,[1,2] Abdul Naser Ikram,[3] Ihsanullah Sahak[3]

To cite: 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. doi:10.1136/bmjopen-2013-002554

▸ Prepublication history for this paper are available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2013-002554).

Received 8 January 2013
Revised 7 March 2013
Accepted 21 March 2013

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 2.0 Licence; see http://bmjopen.bmj.com

[1]Department of Research and Evaluation, University Research Co., LLC, Bethesda, Maryland, USA
[2]International Health, Johns Hopkins School of Public Health, Baltimore, Maryland, USA
[3]Department of Research and Evaluation, University Research Co., LLC, Kabul, Afghanistan

Correspondence to
Dr Edward I Broughton; ebroughton@urc-chs.com

ABSTRACT

Objectives: Improvement activities, surveillance and research in maternal and neonatal health in Afghanistan rely heavily on medical record data. This study investigates accuracy in delivery care records from three hospitals across workshifts.

Design: Observational cross-sectional study.

Setting: The study was conducted in one maternity hospital, one general hospital maternity department and one provincial hospital maternity department. Researchers observed vaginal deliveries and recorded observations to later check against data recorded in patient medical records and facility registers.

Outcome measures: We determined the sensitivity, specificity, area under the receiver operator characteristics curves (AUROCs), proportions correctly classified and the tendency to make performance seem better than it actually was.

Results: 600 observations across the three shifts and three hospitals showed high compliance with active management of the third stage of labour, measuring blood loss and uterine contraction at 30 min, cord care, drying and wrapping newborns and Apgar scores and low compliance with monitoring vital signs. Compliance with quality indicators was high and specificity was lower than sensitivity. For adverse outcomes in birth registries, specificity was higher than sensitivity. Overall AUROCs were between 0.5 and 0.6. Of 17 variables that showed biased errors, 12 made performance or outcomes seem better than they were, and five made them look worse (71% vs 29%, p=0.143). Compliance, sensitivity and specificity varied less among the three shifts than among hospitals.

Conclusions: Medical record accuracy was generally poor. Errors by clinicians did not appear to follow a pattern of self-enhancement of performance. Because successful improvement activities, surveillance and research in these settings are heavily reliant on collecting accurate data on processes and outcomes of care, substantial improvement is needed in medical record accuracy.

ARTICLE SUMMARY

Article focus

- We investigate the accuracy in delivery care records from three hospitals across workshifts.
- We determined the sensitivity, specificity, area under the receiver operator characteristics curves, proportions correctly classified and the tendency to make performance seem better than it was.

Key messages

- Medical record accuracy was generally poor.
- Errors by clinicians did not appear to follow a pattern of self-enhancement of performance.
- Substantial improvement is needed in medical record accuracy.

Strengths and limitations of this study

- Some indicators have very high or low compliance score, decreasing the usefulness of some sensitivity or specificity measures.
- Clinician behaviour may have changed from normal due to the Hawthorne Effect.

INTRODUCTION

Quality improvement (QI) in healthcare often relies on teams of providers performing self-assessments of compliance with standards of care. These often take the form of medical record audits to determine if what is reported as completed in the written record follows the standards of care in force in the specific setting. This is often the most efficient method of data collection for performance indicators, and is therefore frequently used in resource-constrained settings.[1] Some have found health-provider self-assessment to be effective in improving performance in circumstances where higher level monitoring and supervision are unavailable.[2] Information from such assessment is crucial in designing QI interventions, to identify performance gaps that require attention and allow the QI team to monitor its progress in improving the process of healthcare delivery.[3] It is therefore essential that these data be an accurate and valid representation of actual performance.

The USAID Health Care Improvement Project (HCI) has been implementing

BMJ Open: first published as 10.1136/bmjopen-2013-002554 on 24 April 2013. Downloaded from http://bmjopen.bmj.com/ on March 30, 2023 by guest. Protected by copyright.

collaborative QI interventions in hospitals in Kabul since November 2009. In the beginning, HCI staff started data collection and gradually delegated it to QI teams who generally collect information from hospital records on compliance with quality performance standards. These data are shared with officials from the Afghanistan Ministry of Public Health (MoPH) and used to track and evaluate the progress of QI efforts. However, problems with medical records have been noted in this setting,[4] and there are concerns that the patient charts and facility registers may not accurately reflect the true clinical picture due to resource constraints and very high patient loads.

This study examined the accuracy of patient medical record data from patient charts and ward registries generated from vaginal deliveries in two hospitals in Kabul and one in Parwan, Afghanistan. There have been few such studies in maternal health settings in high-income settings from which the conclusion was that accuracy was mixed.[5] Studies from low-resource settings are fewer in number and do not offer strong conclusions on the medical record accuracy.[6 7] We could find no study on the accuracy of medical records conducted in Afghanistan to date.

Three specific research questions were addressed:

1. To what extent are the data reported in the medical records representative of what happened during childbirth?
2. Does the accuracy of medical record data vary between facilities? Does the accuracy of medical records vary among the three workshifts in which the delivery occurs?
3. What is the level of compliance to standards of clinical practice seen in the deliveries observed?

METHODS

Study design

This observational cross-sectional study was conducted in three hospitals in Afghanistan, one dedicated maternity facility in Kabul, one maternity department of a general hospital in Kabul and the maternity department of one provincial hospital close to the capital. Three medical doctors were trained in observing deliveries taking place in participating facilities and recording their observations on a written checklist. They checked their observations against the data entered into the corresponding patient medical records and facility registers 24 or more hours after the observed delivery to ensure adequate time for the records to be completed by the attending clinician.

Sampling

The sampling frame was any vaginal delivery that took place in one of the three maternity facilities on the days in which the observations took place. Three observers were assigned to each of the operational delivery rooms for the three shifts in a 24 h period of consecutive days until an adequate sample size was achieved. The same three observers were used in each of the three hospitals. Deliveries were excluded if they occurred outside the delivery rooms (those occurring in other rooms in the hospital or before arrival at the facility) and deliveries that progressed to caesarian sections. The sample size calculation was based on a level of agreement between observations and the patient medical record/registers of 50% and the ability to detect a 15% difference between agreement in the referral hospitals and the general hospitals with an α of 0.05 and a power of 0.8. This yielded a minimum sample size of 186 in each facility and from each workshift. We aimed for approximately 200 in each group.

Data collection

Performance of the following 17 tasks was recorded from the observations and then checked against the patient medical record. These tasks were chosen because they are all considered standard practices and part of the clinical guidelines for vaginal delivery by the Afghanistan MOPH.

- Active management of third stage of labour (AMTSL): administration of a uterotonic, controlled cord contraction and uterine massage (performance of all three elements of AMTSL for the case)
- Uterotonic administration carried out in the first minute following delivery
- Controlled traction of the umbilical cord
- Uterine massage following delivery
- Drying and wrapping of the newborn
- Umbilical cord care
- Breastfeeding within the first hour following delivery
- Measuring maternal blood loss in 30 min after delivery
- Monitoring of woman's pulse rate at 30 min after delivery
- Monitoring of woman's blood pressure at 30 min after delivery
- Monitoring of uterine contraction at 30 min after delivery
- Monitoring of woman's pulse rate at 60 min after delivery
- Monitoring of woman's blood pressure at 60 min after delivery
- Monitoring of uterine contraction at 60 min after delivery
- Inspection for laceration
- Newborn eye care
- Apgar score at 5 min after delivery

The following data were recorded during observations and then checked against the birth register:

- Woman's diagnosis of postpartum haemorrhage (PPH; blood loss >500 ml);
- Neonatal asphyxia;
- Neonatal death within 1 h of delivery;

BMJ Open: first published as 10.1136/bmjopen-2013-002554 on 24 April 2013. Downloaded from http://bmjopen.bmj.com/ on March 30, 2023 by guest. Protected by copyright.

- Stillbirth;
- Maternal death within 6 h of delivery.

Medical records were considered correctly classified only if they were completed and agreed with what the research assistant observed. For example, if the observer saw that uterotonic was administered in the first minute following delivery, but it was reported as not administered or the information on uterotonic administration was completely missing from the chart, then this was considered incorrectly classified.

The hospital and the workshift in which the delivery occurred were also recorded.

Ethical considerations

The study was approved by the institutional review boards of both the Afghan MoPH and the University Research Co., LLC. Data collectors observing deliveries were all female medical doctors specialised in obstetrics and gynaecology and dressed in scrubs as appropriate for the delivery room. As the settings for the study were teaching hospitals, there are often personnel observing patient care without being part of that care. Also, the nature of the delivery rooms in all three facilities allowed unobtrusive observations with no interference to clinical care. Data were anonymised for analysis. Delivering mothers were informed verbally of the nature of the study and gave written consent or thumb-print for those who were illiterate. Participant health workers who were observed also signed a written consent form before participating in the study. If the observers saw any practice dangerous to the delivering mother or the neonate, they informed the clinician delivering the care. Permission from hospital administrators and MoPH officials was obtained prior to starting the study.

Data analysis

Results were entered into an Excel database with double entry to ensure accuracy. Analyses were conducted using STATA V.11. p Values were calculated for statistical significance. We calculated sensitivity (proportion of cases where performance to standard was accurately reported) and specificity (proportion of cases where non-performance to standard was accurately reported). We also calculated the area under the receiver operator characteristics (AUROCs), combining the proportion of true and false positives to give an indication of the usefulness of the medical record, where 1.0 is a perfect indicator, while 0.5 is a test no better than a guess. We recorded the overall compliance with the indicator and raw agreement between observers and the medical records. Self-enhancement errors are the proportion of discordances between the medical records and observers where the medical record shows a positive result: either compliance with an indicator such as AMTSL or the non-occurrence of an adverse outcome such as PPH. The p value is for the test of whether or not this proportion is 50% as would be expected if errors occurred at random. For example, in table 1, 60% of the discordances were when the medical record indicated that AMTSL was completed, but the observer reported that it was not actually done. This proportion is not significantly different to the 50% expected if the errors occurred at random as determined by Fisher's exact test (p=0.138).

RESULTS

A total of 600 observations were completed with close to equal distribution across the three shifts and three hospitals (table 2). Below are presented the results for all variables in all hospitals (table 1) as well as the results from five indicators selected to represent high, medium and low compliance/occurrence divided by hospitals and workshift (tables 3 and 4). Full tables including all variables reported by hospitals and workshift are available online, but not included here due to their size.

Overall

There was high compliance with the three elements of AMTSL, measurement of blood loss and uterine contraction at 30 min, cord care, drying and wrapping newborns and Apgar scores. There was low compliance with taking the mothers' vital signs following delivery, especially 1 h after delivery. In many cases of compliance with quality-of-care indicators, specificity was lower than sensitivity; while in reporting adverse outcomes of stillbirths, neonatal death, asphyxia and PPH, specificity was higher than sensitivity. Of the 16 variables in the medical charts and birth registries for which there was a statistically significant indication of biased errors, 11 were of the type that made the clinicians' performance or the clinical outcomes seem better than they actually were and five were of the type that made them look worse (71% vs 29%, p=0.143). There were no maternal deaths observed or recorded in the medical records among the women sampled.

Hospitals

All hospitals had high compliance with the three elements of AMTSL and high sensitivity in the medical records. However, specificity was low in all three. Compliance with breastfeeding within the first hour after delivery was variable; sensitivity was high in the maternity hospital and low in the general and provincial hospitals. Monitoring of uterine contraction 1 h after delivery had low compliance in all hospitals, with high sensitivity and low specificity. The proportion of stillbirths was highest in the provincial hospital and lowest (50% lower proportion, p=0.247) in the general hospital. Sensitivity and specificity were both relatively high for this indicator. There was variability in compliance with neonatal eye care with the maternity hospital having high sensitivity and low specificity, and the general and provincial hospital having moderate sensitivity and specificity. The AUROC for all indicators was less than 0.6 except for stillbirths, which was above 0.92 in all hospitals (table 3).

BMJ Open: first published as 10.1136/bmjopen-2013-002554 on 24 April 2013. Downloaded from http://bmjopen.bmj.com/ on March 30, 2023 by guest. Protected by copyright.

Table 1 Overall results for all indicators

	Compliance/ occurrence (%)	Sensitivity (%)	Specificity (%)	Correctly classified (%)	AUROC	Self-enhancement errors	p Value
From patient charts							
AMTSL	94.0	96.1	8.3	90.8	0.52	60.0	0.138
Oxytocin	94.2	98.9	2.9	93.3	0.51	85.0	<0.001*
Cord traction	99.5	96.5	0.0	96.0	0.48	12.3	<0.001**
Uterine massage	100.0	97.2	Na	97.2	Na	0.0	<0.001**
Dry and wrap newborn	97.5	38.3	86.7	39.5	0.62	0.6	<0.001**
Cord care	97.5	52.8	60.0	53.0	0.56	2.1	<0.001**
Immediate breastfeeding	49.3	37.2	67.1	52.3	0.52	35.0	<0.001
Blood loss at 30 min	93.3	98.8	2.5	92.3	0.51	84.8	<0.001*
HR at 30 min	1.5	100	4.4	5.8	0.52	100	<0.001*
BP at 30 min	3.0	96.8	2.8	17.7	0.50	99.4	<0.001*
Uterine contraction at 30 min	99.3	99.3	4.1	91.5	0.52	92.2	<0.001*
HR at 60 min	0.8	100	4.0	4.8	0.52	100	<0.001*
BP at 60 min	3.0	88.9	3.8	6.3	0.46	99.6	<0.001*
Uterine contraction at 60 min	12.5	97.3	4.0	15.7	0.51	99.6	<0.001*
Laceration	4.3	11.5	99.3	95.5	0.55	85.2	0.000*
Apgar	99.2	99.0	20.0	98.3	0.59	60.0	0.527
Eye care	80.5	76.2	37.6	68.7	0.57	38.8	<0.001**
From birth registers							
Postpartum haemorrhage	1.5	22.2	100.0	98.8	0.61	100.0	0.003*
Asphyxia	6.0	33.3	99.1	95.2	0.66	82.8	<0.001*
Neonatal death	0.7	75.0	99.2	99.0	0.87	16.7	0.103
Still birth	3.5	90.5	99.0	98.7	0.95	25.0	0.157

*Statistically significant, supporting hypothesis of errors showing higher performance.
**Statistically significant, supporting hypothesis of errors showing lower performance.
AUROC, area under the receiver operator characteristics curve; BP, blood pressure; HR, heart rate.

Workshift

Compliance, sensitivity and specificity varied less among the three shifts than among the three hospitals. The greatest variation in compliance was in uterine contraction at 1 h after delivery, while the greatest variation in AUROC was in neonatal eye care. Errors analysed by workshift did not appear to follow a pattern of errors of self-enhancement of performance (table 4).

Table 2 Number of observations by workshift and hospital

	Hospital			
Workshift	Maternity (M)	General (G)	Provincial (P)	Total
Morning	63	51	80	194
Evening	63	66	70	199
Night	74	85	48	207
Total	200	202	198	600

Postpartum haemorrhage

Although the study was not powered to determine whether there were differences in the way women diagnosed with PPH were treated, we included a separate analysis of those nine cases. There were slightly higher proportions of women with PPH who had their uterine contraction, blood loss and vital signs measured, but those proportions were still low (table 5).

DISCUSSION

There have been substantial investments in improving the quality of care with the goal of achieving better maternal

BMJ Open: first published as 10.1136/bmjopen-2013-002554 on 24 April 2013. Downloaded from http://bmjopen.bmj.com/ on March 30, 2023 by guest. Protected by copyright.

Table 3 Results from all shifts by hospital

Hospital	Compliance/ occurrence (%)	Sensitivity (%)	Specificity (%)	Correctly classified (%)	AUROC	Self-enhancement errors	p Value
AMTSL							
M	93	100	0	93	0.50	100	<0.001*
G	92	90	13	84	0.51	44	0.480
P	97	98	17	95	0.57	56	0.739
Immediate breastfeeding							
M	49	94	8	50	0.51	94	<0.001*
G	23	6	98	77	0.52	6	<0.001**
P	76	10	94	30	0.52	2	<0.001**
Uterine contraction @ 60 min							
M	6	100	3	9	0.51	100	<0.001*
G	19	97	4	22	0.51	99	<0.001*
P	12	96	5	16	0.51	99	<0.001*
Still birth							
M	4	86	99	99	0.92	67	0.564
G	2	100	98	99	0.99	100	0.083
P	5	89	99	99	0.94	50	0.157
Eye care							
M	64	98	12	67	0.55	97	<0.001*
G	92	69	63	68	0.66	9	<0.001**
P	86	68	89	71	0.78	5	<0.001**

*Statistically significant, supporting hypothesis of errors showing higher performance.
**Statistically significant, supporting hypothesis of errors showing lower performance.
AMTSL, active management of third stage of labour; AUROC, area under the receiver operator characteristics curve; G, general hospital; M, maternity hospital; P, provincial hospital.

and neonatal outcomes in several hospitals throughout Afghanistan in the previous several years, including the three hospitals participating in this study.[8–10] Compliance with the quality standards measured by the indicators was generally high, particularly for AMTSL and several elements of essential newborn care. These indicators are often used to monitor the processes of maternal and neonatal care, and it is therefore expected that compliance should be reasonably high.

About 1.5% of women were diagnosed with PPH. While no published data of the prevalence of PPH in Afghanistan could be found, it is better than the PPH prevalence of 2.6% found in Mali, which also has a high maternal death rate similar to that in Afghanistan.[11 12] There were no maternal mortalities among the 600 cases observed in this study. If the maternal mortality ratio of 327/100 000 reported from the Afghanistan Mortality Survey of 2010[13] was observed in this hospital, about two deaths would have been predicted. However, the maternal mortality ratio is likely to be lower in these hospitals than the country as a whole because of the access to emergency obstetric care that is unavailable to many women in the rest of Afghanistan.[14] The four infant deaths observed were lower than the 42 expected, if the national neonatal mortality ratio was seen in the 600 deliveries observed.[13] However, only the immediate postpartum period was observed rather than the 28 days as per the definition of neonatal morality; again, a lower occurrence of death was expected in this setting compared with the country as a whole.

Sensitivity was high for indicators of compliance with standards of care with the exception of breastfeeding in the first hour after delivery, cord care and drying and wrapping of the newborn. Failing to record these accurately when they were actually carried out is not likely to have a major impact on the safety of the care provided, but does lead to the underestimation of the level of quality for newborn care. Specificity was lower than 10% in all compliance indicators except the three neonatal care indicators listed above. This showed that clinicians recorded having performed a task that observers reported they did not do, which in this case makes the quality of care appear better than it truly is. The clinical implication on patient safety of such errors may not be of great consequence. For example, if the medical record indicates that a specific uterotonic was administered, when in reality it was not, and that woman is later diagnosed with PPH, an additional dose of uterotonic may be administered. This may or may not change the clinical outcome for that patient. For the indicators of asphyxia, PPH and laceration, for which sensitivity was low, failing to accurately identify cases may have a detrimental effect on clinical decision making, potentially leading to an increase in the risk of

BMJ Open: first published as 10.1136/bmjopen-2013-002554 on 24 April 2013. Downloaded from http://bmjopen.bmj.com/ on March 30, 2023 by guest. Protected by copyright.

Table 4 Results for all hospitals by workshift

Indicator	Shift	Compliance/ occurrence (%)	Sensitivity (%)	Specificity (%)	Correctly classified (%)	AUROC	Self-enhancement errors	p Value
AMTSL	Morning	94	95	0	90	0.48	55	0.824
	Evening	96	96	13	93	0.54	50	1.000
	Night	92	97	12	90	0.54	71	0.078
Immediate breastfeeding	Morning	55	36	69	51	0.52	28	0.000**
	Evening	51	35	68	51	0.52	32	0.000**
	Night	42	41	65	55	0.53	79	0.000**
Uterine contraction @ 60 min	Morning	19	95	4	22	0.50	99	<0.001**
	Evening	9	100	3	12	0.51	100	<0.001**
	Night	10	100	5	14	0.52	100	<0.001**
Stillbirth	Morning	6	92	99	99	0.96	50	1.000
	Evening	2	75	98	98	0.87	80	0.625
	Night	2	100	99	99	1.00	100	0.500
Eye care	Morning	83	75	52	71	0.63	29	0.002**
	Evening	84	82	58	78	0.70	30	0.010
	Night	74	71	17	57	0.44	50	1.000

*Statistically significant supporting hypothesis of errors showing higher performance.
**Statistically significant, supporting hypothesis of errors showing lower performance.
AMTSL, active management of third stage of labour; AUROC, area under the receiver operator characteristics curve.

adverse outcomes for resuscitated neonates or mothers because they miss follow-up observation and care indicated by these diagnoses.

Specificity and the proportion of cases correctly classified were generally higher in records taken from registers compared with those take from medical charts while sensitivity was lower in the registers. This is possibly because the registers are generally used to record low-frequency events and clinicians may think it more important to capture the occurrence of those events than their absence.

Analysis of whether or not women with PPH were treated differently was included because, given the high volume of deliveries attended for the small number of staff, we thought that clinicians may be rationing their time taking vital signs only of those women whose vital signs were very important in the overall management of their condition. While a higher proportion of women with PPH were observed to have their vital signs checked at 30 and 60 min, it was far from being complete monitoring in all indicators for these cases. These lower levels of monitoring could have detrimental consequences to clinical care and outcomes.

There were few significant differences among hospitals in terms of the accuracy of their medical records. The two largest differences were in the recording of immediate breastfeeding and infant eye care, both of which showed the maternity hospital substantially outperformed the general and provincial hospitals. Given the relative consistency in performance on the other measures, the reason for this large variation is unclear.

While some hospitals in urban centres in Afghanistan are overstaffed, there tends to be very few female staff overall; given that maternal services are almost exclusively provided by women, maternity facilities are generally understaffed.[15] Maternity hospitals are also reported to have infection control problems and chronic shortages of material resources.[16]

Few other studies have examined accuracy in the documentation of patient status and care using expert observations of medical procedures. In a study of surgical complications in the Netherlands, ten Broek *et al*[17] found sensitivity and specificity of documenting a specific complication as 85.1% and 72.4%, respectively, compared with the gold standard of observation of the surgery. Another study found a discrepancy of around 30% in identifying patients at risk for undernutrition between observations carried out by researchers and records of the evaluation in the patients' medical charts.[18] We found no benchmark study using observations of deliveries to test the accuracy or completeness of medical records.

The three participating hospitals were selected because one is a national maternity referral hospital and the other two were considered representative of a large general hospital and a provincial facility. Like many facilities in Afghanistan, the three have been involved in improvement interventions since 2009 that have focused on maternal and newborn health. The study was not designed to be representative of all the hospitals in Afghanistan and the performance in the participating

BMJ Open: first published as 10.1136/bmjopen-2013-002554 on 24 April 2013. Downloaded from http://bmjopen.bmj.com/ on March 30, 2023 by guest. Protected by copyright.

Table 5 Comparison of vital signs, blood loss and uterine contraction monitoring between postpartum haemorrhage cases and controls

	At 30-min postpartum				At 60 min postpartum			
Postpartum haemorrhage	HR (%)	BP (%)	UC (%)	BL (%)	HR (%)	BP (%)	UC (%)	BL (%)
No=591	8 (1.4)	92 (15.6)	542 (91.7)	551 (93.2)	4 (0.7)	16 (2.7)	71 (12.0)	76 (12.9)
Yes=9	1 (11.1)	3 (33.3)	9 (100)	9 (100)	1 (11.1)	2 (22.2)	4 (44.4)	4 (44.4)
p Value	0.017	0.147	0.367	0.419	<0.001	<0.001	0.003	0.006

BL, blood loss; BP, blood pressure; HR, heart rate; UC, uterine contraction.

facilities is likely to be higher than that in hospitals that have not undergone the same level of improvement activities as these three.

It is expected that the accuracy of medical records may not be as high as desired, and it has been noted by other authors that their quality is poor.[4] However, several organisations conducting QI work in this setting rely to a great extent on medical records to monitor the progress of improvement in care processes and outcomes.[8–10 19] These records have also been used for the surveillance of maternal healthcare and outcomes.[4 20] While the efficiency of reviewing medical records for monitoring and evaluating improvement interventions and for surveillance is very attractive to implementers, this should be weighed against the poor accuracy of this resource and may lead to suboptimal policy to the detriment of patients and the health system.

In situations where resources allow it, procedures to establish the validity of the medical records should be implemented. Those working to improve the quality of care who rely heavily on medical records should stress to frontline clinicians the importance of accurately recording clinical activities in patient charts. Providing training on clinical record keeping, allowing adequate time and staff support and fostering an atmosphere of not assigning punishment or blame for errors in clinical practice may lead to more accurate medical records.[21 22] Given the importance of the accuracy of medical records to the success of improvement efforts, implementers should use the same approaches to addressing record keeping as they do for improving the processes and outcomes of clinical care. Those involved with surveillance based on medical records should take into account the inaccuracies found in this study when interpreting their own results.

LIMITATIONS

Compliance was mostly high for the quality measures and occurrence was low for the adverse outcomes such as stillbirths. While this is a positive result for the clinicians, it does not make for an optimal study of the quality of the medical records of clinical processes and outcomes. For example, cord traction conducted to standards following delivery of the neonate occurred in 597 of the 600 observed deliveries (99.5%). This left only three opportunities of 600 deliveries for clinicians to accurately record not conducting cord contraction to compliance with standards. Missing any or all of these few opportunities gives a low or zero specificity and therefore an AUROC close to 0.5. This was the case for several indicators even though their proportions correctly classified were high. However, with indicators where the compliance or occurrence was not at the extremes, such as the 49.3% compliance with immediate breastfeeding and the 80.5% compliance with neonatal eye care, the results for the AUROC were still not very high and not greatly different to the results obtained for the other indicators. The proportions correctly classified for those indicators were correspondingly low. A larger sample size would have lessened this effect.

The Hawthorne Effect, defined as the change in the behaviour being observed due to the known presence of the observer,[23 24] may have improved compliance with quality of care indicators. Clinician participants were initially aware of the observer because they were required to sign the informed consent form. However, they did not know that the accuracy of the medical records would also be checked. Also, the delivery rooms where observations took place are large open areas and clinicians are used to operating where many people observe their activities. We do not consider it likely that the Hawthorne Effect had a significant influence over the accuracy of the medical records.

We did not distinguish between data that were incorrectly reported and data that were missing from the chart. The reason was because, regardless of whether clinical information is missing or incorrectly recorded, the patients' care may be compromised and the medical record cannot be trusted as a reflection of reality. Had we considered only the accuracy of the non-missing data in the medical records, they would have appeared to be of better quality for clinical care than they actually were. However, it could be argued that we should have considered missing and erroneously recorded data separately.

Observations from the researchers were considered the 'gold standard'. These were three medical doctors with extensive experience in maternal and neonatal clinical care, and they received training on how to conduct their observations, including a trial of observing deliveries. However, there was no check in this study of intratester or intertester reliability of these three observers. If observers did make errors, there is no reason these would have biased the results for the accuracy of the records one way or the other.

CONCLUSION

Compliance was high in some indicators of maternal and neonatal health quality of care, but low for others. The accuracy of medical records in capturing clinical activities

BMJ Open: first published as 10.1136/bmjopen-2013-002554 on 24 April 2013. Downloaded from http://bmjopen.bmj.com/ on March 30, 2023 by guest. Protected by copyright.

and outcomes was generally poor. The success of activities to improve the quality of care in these settings is heavily reliant on collecting accurate data on processes and outcomes of care, substantial attention needs to be paid to improving medical record accuracy.

Acknowledgements This study was supported by the American people through the US Agency for International Development (USAID) and its Health Care Improvement Project (HCI). HCI is managed by the University Research Co., LLC (URC) under the terms of Contract Number GHN-I-03-07-00003-00. We thank the directors of the participating hospitals for their generous cooperation. We also thank Stacie Gobin for her assistance with analysis.

Contributors EIB and ANI conceived the idea for the study and were responsible for the design of the study. EIB drafted the protocol with input from ANI and IS. ANI organised the submission to the Afghanistan IRB. EIB submitted to the US IRB. ANI and IS were responsible for organising data collection and data entry and cleaning. ENB analysed the data and produced all the tables. EIB, ANI and IS drafted the paper with ENB taking the lead. All three authors responded to peer reviewers' questions, edited the final draft and approved the final manuscript.

Funding US Agency for International Development.

Competing interests None.

Ethics approval University Research Co., LLC and Afghanistan Ministry of Public Health.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Additional data from indicators not included in the main manuscript due to size limitations and redundancy are available from the corresponding author by e-mailing ebroughton@urc-chs.com. These include extended versions of table 1 (with five additional indicators for a total of 26 indicators) and tables 3 and 4 (with all 26 indicators disaggregated by hospital and workshift, respectively).

REFERENCES

1. Franco L, Marquez L, Ethier K, *et al*. Results of collaborative improvement: effects on health outcomes and compliance with evidence-based standards in 27 applications in 12 countries. Bethesda, MD: Iniversity Research Co, LLC, 2009.
2. Kelley E, Kelley AG, Simpara CH, *et al*. The impact of self-assessment on provider performance in Mali. *Int J Health Plann Manage* 2003;18:41–8.
3. Vos L, Duckers ML, Wagner C, *et al*. Applying the quality improvement collaborative method to process redesign: a multiple case study. *Implement Sci* 2010;5:19.
4. Kandasamy T, Merialdi M, Guidotti RJ, *et al*. Cesarean delivery surveillance system at a maternity hospital in Kabul, Afghanistan. *Int J Gynaecol Obstet* 2009;104:14–17.
5. Lain SJ, Hadfield RM, Raynes-Greenow CH, *et al*. Quality of data in perinatal population health databases: a systematic review. *Med Care* 2012;50:e7–20.
6. Hermida J, Broughton EI, Franco L Miller. Validity of self-assessment in a quality improvement collaborative in Ecuador. *Int J Qual Health Care* 2011;23:690–6.
7. Ndira SP, Rosenberger KD, Wetter T. Assessment of data quality of and staff satisfaction with an electronic health record system in a developing country (Uganda): a qualitative and quantitative comparative study. *Methods Inf Med* 2008;47:489–98.
8. USAID Health Care Improvement Project. *USAID HCI Afghanistan newsletter*. 2nd edn. Kabul, Afghanistan: University Research Co., LLC, 2012.
9. Jhpiego. *Afghanistan country profile*. Baltimore, MD: Jhpiego, 2012.
10. Holmes W. Technical report: PHI Afghanistan, June–September, 2008. Kabul, Afghanistan, 2012.
11. Teguete I, Maiga AW, Leppert PC. Maternal and neonatal outcomes of grand multiparas over two decades in Mali. *Acta Obstet Gynecol Scand* 2012;91:580–6.
12. UNICEF. Mali Statistics. Geneva, 2012.
13. World Bank. *Afghanistan country overview, 2012*. Washington, DC: World Bank, 2012.
14. Hirose A, Borchert M, Niksear H, *et al*. Difficulties leaving home: a cross-sectional study of delays in seeking emergency obstetric care in Herat, Afghanistan. *Soc Sci Med* 2011;73:1003–13.
15. Broun D, Debionne E, Ghane S, *et al*. Afghanistan National Hospital Survey. Kabul, Afghanistan, 2004.
16. Williams JL, McCarthy B. Observations from a maternal and infant hospital in Kabul, Afghanistan—2003. *J Midwifery Womens Health* 2005;50:e31–5.
17. ten Broek RP, van den Beukel BA, van Goor H. Comparison of operative notes with real-time observation of adhesiolysis-related complications during surgery. *Br J Surg* 2013;100:426–32.
18. Simmons SF, Lim B, Schnelle JF. Accuracy of minimum data set in identifying residents at risk for undernutrition: oral intake and food complaints. *J Am Med Dir Assoc* 2002;3:140–5.
19. Kim YM, Tappis H, Zainullah P, *et al*. Quality of caesarean delivery services and documentation in first-line referral facilities in Afghanistan: a chart review. *BMC Pregnancy Childbirth* 2012; 12:14.
20. Dott MM, Orakail N, Ebadi H, *et al*. Implementing a facility-based maternal and perinatal health care surveillance system in Afghanistan. *J Midwifery Womens Health* 2005;50:296–300.
21. Clayton HB, Sappenfield WM, Gulitz E, *et al*. The Florida Investigation of Primary Late Preterm and Cesarean Delivery: the accuracy of the birth certificate and hospital discharge records. *Matern Child Health J* 2012. 10.1007/s10995-012-1065-0.
22. Schnelle JF, Osterweil D, Simmons SF. Improving the quality of nursing home care and medical-record accuracy with direct observational technologies. *Gerontologist* 2005;45:576–82.
23. De Amici D, Klersy C, Ramajoli F, *et al*. Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. *Control Clin Trials* 2000;21:103–14.
24. Kohli E, Ptak J, Smith R, *et al*. Variability in the Hawthorne effect with regard to hand hygiene performance in high- and low-performing inpatient care units. *Infect Control Hosp Epidemiol* 2009;30:222–5.

BMJ Open: first published as 10.1136/bmjopen-2013-002554 on 24 April 2013. Downloaded from http://bmjopen.bmj.com/ on March 30, 2023 by guest. Protected by copyright.