



How Accurate Are Medical Record Data in Afghanistan's Maternal Health Facilities? An observational Validity Study

Journal:	<i>BMJ Open</i>
Manuscript ID:	bmjopen-2013-002554
Article Type:	Research
Date Submitted by the Author:	08-Jan-2013
Complete List of Authors:	Broughton, Edward; University Research Co., LLC, Research and Evaluation; Johns Hopkins School of Public Health, International Health Ikram, Abdul; University Research Co., LLC, Sahak, Ihsanullah; University Research Co., LLC,
Primary Subject Heading:	Obstetrics and gynaecology
Secondary Subject Heading:	Health informatics, Epidemiology, Reproductive medicine, Health services research, Research methods
Keywords:	Maternal medicine < OBSTETRICS, Fetal medicine < OBSTETRICS, AUDIT, EPIDEMIOLOGY, Organisation of health services < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

1
2
3
4 **How Accurate Are Medical Record Data in Afghanistan's Maternal Health**
5 **Facilities? An observational Validity Study**
6
7
8
9

10
11 Corresponding author address: Edward I Broughton, MPH, PhD
12 University Research Co., LLC
13 7200 Wisconsin Ave Suite 600
14 Bethesda MD 20814

15
16
17 Ph: +1-301-941-8533
18 Fax: +1-301-941-8427
19

20
21
22 Co-authors Abdul Naser Ikram, MD, MPH
23 University Research Co., LLC (Afghanistan)
24 Research and Evaluation Department
25 Kabul, Afghanistan
26

27
28 Ihsanullah Sahak, MD
29 University Research Co., LLC (Afghanistan)
30 Research and Evaluation Department
31 Kabul, Afghanistan
32
33

34
35
36 Word Count: 3329
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 work-shifts.

Design/Setting

This observational cross-sectional 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. We determined sensitivity, specificity, areas under the receiver operator characteristics curves (AUROC), proportions correctly classified and the tendency to make performance seem better than it actually was.

Main Results

Six hundred observations across the three shifts and three hospitals showed high compliance with active management of the third stage of labor, measuring blood loss and uterine contraction at 30 minutes, 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% versus 29%, $P = 0.143$). Compliance, sensitivity and specificity varied less among the three shifts than among hospitals.

Conclusions

Medical records 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 is heavily reliant on collecting accurate data on processes and outcomes of care, substantial improvement is needed in medical record accuracy.

Word count: 274/300

Running title: Medical record accuracy in Afghanistan

Article summary

Article focus

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

Key Messages

- Medical records 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 very high or low compliance score, decreasing the usefulness of some sensitivity or specificity measures
- Clinician behavior may have changed from normal due to the Hawthorn Effect

Introduction

Quality improvement (QI) in health care 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 is 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 health care 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 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 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 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. Is the facility or the time of delivery associated with the accuracy of medical record data in participating maternity hospitals?
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 data entered into 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-hour 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 alpha of 0.05 and a power of 0.8.

Data Collection

Performance of the following 17 tasks was recorded from observations then checked against the patient medical record:

- Active management of third stage of labor (AMTSL): administration of a uterotonic, controlled cord contraction and uterine massage (performance of all three elements of AMTSL for the case)
- Uterotonic administration done 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 minutes after delivery
- Monitoring of woman's pulse rate at 30 minutes after delivery
- Monitoring of woman's blood pressure at 30 minutes after delivery
- Monitoring of uterine contraction at 30 minutes after delivery
- Monitoring of woman's pulse rate at 60 minutes after delivery
- Monitoring of woman's blood pressure at 60 minutes after delivery
- Monitoring of uterine contraction at 60 minutes after delivery
- Inspection for laceration
- Newborn eye care
- Apgar score at five minutes after delivery

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

- Woman's diagnosis of post-partum hemorrhage (blood loss >500 ml)
- Neonatal asphyxia

- Neonatal death within one hour of delivery
- Stillbirth
- Maternal death within six hours of delivery

The hospital and the work-shift 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 University Research Co., LLC. Data collectors observing deliveries were all female medical doctors specialized in obstetrics and gynecology 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 anonymized for analysis. Delivering mothers were informed of the nature of the study and gave written consent. 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 commencing the study.

Data Analysis

Results were entered into an Excel database with double entry to ensure accuracy. Analyses were conducted using STATA version 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 (AUROC), 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 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 2, 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 1). Below are presented the results for all variables in all hospitals (Table 2) as well as the results from five indicators selected to represent high, medium and low compliance/occurrence

1
2
3 divided by hospitals and work shift (Tables 3 and 4). Full tables including all variables reported by
4 hospitals and work shift are available online but not included here due to their size.
5

6
7 **Overall:** There was high compliance with the three elements of AMTSL, measurement of blood loss and
8 uterine contraction at 30 minutes, cord care, drying and wrapping newborns and Apgar scores. There
9 was low compliance with taking the mothers' vital signs following delivery, especially one hour after
10 delivery. In many cases of compliance with quality of care indicators, specificity was lower than
11 sensitivity, while in reporting adverse outcomes of stillbirths, neonatal death, asphyxia and post-partum
12 hemorrhage, specificity was higher than sensitivity. Of the 16 variables in the medical charts and birth
13 registries for which there was a statistically significant indication of biased errors, 11 were of the type
14 that made the clinicians' performance or the clinical outcomes seem better than they actually were and
15 five were of the type that made them look worse (71% versus 29%, $p = 0.143$). There were no maternal
16 deaths observed or recorded in the medical records among the women sampled.
17
18
19

20
21 **Hospitals:** All hospitals had high compliance with the three elements of AMTSL and high sensitivity in
22 the medical records. However, specificity was low in all three. Compliance with breastfeeding within the
23 first hour after delivery was variable, and sensitivity was high in the maternity hospital and low in the
24 general and provincial hospitals. Monitoring of uterine contraction one hour after delivery had low
25 compliance in all hospitals, with high sensitivity and low specificity. The proportion of stillbirths was
26 highest in the provincial hospital and lowest (50% lower proportion, $p=0.247$) in the general hospital.
27 Sensitivity and specificity were both relatively high for this indicator. There was variability in compliance
28 with neonatal eye care with the maternity hospital having high sensitivity, low specificity and the
29 general and provincial hospital having moderate sensitivity and specificity. The AUROC for all indicators
30 was less than 0.6 except for stillbirths, which was above 0.92 in all hospitals (Table 3).
31
32
33

34
35 **Work shift:** Compliance, sensitivity and specificity varied less among the three shifts than among the
36 three hospitals. The greatest variation in compliance was in uterine contraction at one hour after
37 delivery while the greatest variation in AUROC was in neonatal eye care. Errors analyzed by workshift
38 did not appear to follow a pattern of errors of self-enhancement of performance (Table 4).
39

40
41 **Post-partum hemorrhage (PPH):** Although the study was not powered to determine whether there
42 were differences in the way women diagnosed with PPH were treated, we included a separate analysis
43 of those nine cases. There were slightly higher proportions of women with PPH who had their uterine
44 contraction, blood loss and vital signs measured but those proportions were still low (Table 5).
45

46 Discussion

47
48 There have been substantial investments in improving the quality of care with the goal of achieving
49 better maternal and neonatal outcomes in several hospitals throughout Afghanistan in the previous
50 several years, including the three hospitals participating in this study (8-10). Compliance with the quality
51 standards measured by the indicators was generally high, particularly for AMTSL and several elements of
52 essential newborn care. These indicators are often used to monitor the processes of maternal and
53 neonatal care, and it is therefore expected that compliance should be reasonably high.
54
55
56
57
58
59
60

1
2
3 About 1.5% of women were diagnosed with PPH. While no published data of the prevalence of PPH in
4 Afghanistan could be found, it is better than the PPH prevalence of 2.6% found in Mali, which also has a
5 high maternal mortality rate similar to that in Afghanistan (11, 12). There were no maternal mortalities
6 among the 600 cases observed in this study. If the maternal mortality ratio of 327/100,000, reported
7 from the Afghanistan Mortality Survey of 2010 (13) was observed in this hospital, about two deaths
8 would have been predicted. However, the maternal mortality ratio is likely to be lower in these hospitals
9 than the country as a whole because of the access emergency obstetric care that is unavailable to many
10 women in the rest of Afghanistan (14). The four infant deaths observed were lower than the 42
11 expected if the national neonatal mortality ratio was seen in the 600 deliveries observed (13). However,
12 only the immediate post-partum period was observed rather than the 28 days as per the definition of
13 neonatal mortality and again, a lower occurrence of death was expected in this setting compared to the
14 country as a whole.

15
16 Sensitivity was high for indicators of compliance with standards of care with the exception of
17 breastfeeding in first hour after delivery, cord care and drying and wrapping of the newborn. Failing to
18 record these accurately when they were actually done is not likely to have a major impact on the safety
19 of the care provided but does lead to underestimation of the level of quality for newborn care.
20 Specificity was lower than 10% in all compliance indicators except the three neonatal care indicators
21 listed above. This showed that clinicians recorded having done a task that observers reported they did
22 not do, which in this case makes the quality of care appear better than it truly is. The clinical implication
23 on patient safety of such errors may not be of great consequence. For example, if the medical record
24 indicates that a specific uterotonic was administered when in reality it was not and that woman is later
25 diagnosed with PPH, an additional dose of a uterotonic may be administered. This may or may not
26 change the clinical outcome for that patient. For the indicators of asphyxia, PPH and laceration, for
27 which sensitivity was low, failing to accurately identify cases may have a detrimental effect on clinical
28 decision making, potentially leading to an increase in the risk of adverse outcomes for resuscitated
29 neonates or mothers because they miss follow-up observation and care indicated by these diagnoses.

30
31 Specificity and the proportion of cases correctly classified were generally higher in records taken from
32 registers compared to those taken from medical charts while sensitivity was lower in the registers. This is
33 possibly because the registers are generally used to record low-frequency events and clinicians may
34 think it more important to capture the occurrence of those events than their absence.

35
36 Analysis of whether or not women with PPH were treated differently was included because, given the
37 high volume of deliveries attended for the small number of staff, we thought that clinicians may be
38 rationing their time taking vital signs only of those women whose vital signs were very important in the
39 overall management of their condition. While a higher proportion of women with PPH were observed to
40 have their vital signs checked at 30 and 60 minutes, it was far from being complete monitoring in all
41 indicators for these cases. These lower levels of monitoring could have detrimental consequences to
42 clinical care and outcomes.

43
44 While some hospitals in urban centers in Afghanistan are overstaffed, there tends to be too few female
45 staff overall, and given that maternal services are almost exclusively provided by women, maternity

1
2
3 facilities are generally understaffed (15). Maternity hospitals are also reported to have infection control
4 problems and chronic shortages of material resources (16).

5
6 It is expected that the accuracy of medical records may not be as high as desired, and it has been noted
7 by other authors that their quality is poor (4). However, several organizations conducting quality
8 improvement work in this setting rely to a great extent on medical records to monitor the progress of
9 improvement in care processes and outcomes (8-10, 17). These records have also been used for
10 surveillance of maternal health care and outcomes (4, 18). While the efficiency of reviewing medical
11 records for monitoring and evaluating improvement interventions and for surveillance is very attractive
12 to implementers, this should be weighed against the poor accuracy of this resource and may lead to
13 suboptimal policy to the detriment of patients and the health system.

14
15 In situations where resources allow it, procedures to establish the validity of the medical records should
16 be implemented. Those working to improve the quality of care who rely heavily on medical records
17 should stress to frontline clinicians the importance of accurately recording clinical activities in patient
18 charts. Providing training or clinical record keeping, allowing adequate time and staff support and
19 fostering an atmosphere of not assigning punishment or blame for errors in clinical practice may lead to
20 more accurate medical records (19, 20). Given the importance of the accuracy of medical records to the
21 success of improvement efforts, implementers should use the same approaches to addressing record
22 keeping as they do for improving the processes and outcomes of clinical care. Those involved with
23 surveillance based on medical records should take into account the inaccuracies found in this study
24 when interpreting their own results.

31 32 33 **Limitations**

34
35 Compliance was mostly high for the quality measures and occurrence was low for the adverse outcomes
36 such as stillbirths. While this is a positive result for the clinicians, it does not make for an optimal study
37 of the quality of the medical records of clinical processes and outcomes. For example, cord traction
38 conducted to standards following delivery of the neonate occurred in 597 of the 600 observed deliveries
39 (99.5%). This left only three opportunities out of 600 deliveries for clinicians to accurately record not
40 conducting cord contraction to compliance with standards. Missing any or all of these few opportunities
41 gives a low or zero specificity and therefore an AUROC close to 0.5. This was the case for several
42 indicators even though their proportions correctly classified were high. However, with indicators where
43 the compliance or occurrence was not at the extremes, such as the 49.3% compliance with immediate
44 breastfeeding and the 80.5% compliance with neonatal eye care, the results for the AUROC were still
45 not very high and not greatly different to the results obtained for the other indicators. The proportions
46 correctly classified for those indicators were correspondingly low. A larger sample size would have
47 lessened this effect.

48
49 The Hawthorne Effect, defined as the change in the behavior being observed due to the known presence
50 of the observer (21, 22), may have improved compliance with quality of care indicators. Clinician
51 participants were initially aware of the observer because they were required to sign the informed
52 consent form. However, they did not know that the accuracy of the medical records would also be

1
2
3 checked. Also, the delivery rooms where observations took place are large open areas and clinicians are
4 used to operating where many people observe their activities. We do not consider it likely that the
5 Hawthorne Effect had a significant influence over the accuracy of the medical records.
6
7

8 Observations from the researchers were considered the “gold standard”. These were three medical
9 doctors with extensive experience in maternal and neonatal clinical care, and they received training on
10 how to conduct their observations, including a trial of observing deliveries. However, there was no check
11 in this study of intra-tester or inter-tester reliability of these three observers. If observers did make
12 errors, there is no reason these would have biased the results for the accuracy of the records one way or
13 the other.
14
15

16 **Conclusion**

17
18 Compliance was high in some indicators of maternal and neonatal health quality of care but low for
19 others. The accuracy of medical records in capturing clinical activities and outcomes was generally poor.
20 Because the success of activities to improve the quality of care in these settings is heavily reliant on
21 collecting accurate data on processes and outcomes of care, substantial attention needs to be paid to
22 improving medical record accuracy.
23
24
25
26

27 **Data Sharing**

28
29 Additional data can be obtained by e-mailing the corresponding author.
30

31 **Acknowledgements**

32 This study was supported by the American people through the United States Agency for International
33 Development (USAID) and its Health Care Improvement Project (HCI). HCI is managed by University
34 Research Co., LLC (URC) under the terms of Contract Number GHN-I-03-07-00003-00. We thank the
35 directors of the participating hospitals for their generous cooperation. We also thank Stacie Gobin for
36 her assistance with analysis.
37
38

39 **Funding**

40
41 Funded by US Agency for International Development.
42

43 **Contributorship**

44 EB conceived and designed the study and performed most of the analysis with input from AI and IS. AI
45 and IS coordinated all data collection and entry. EB wrote the manuscript with substantial input from AI
46 and IS.
47

48 **Competing Interests**

49
50 None
51
52
53
54
55
56
57
58
59
60

Table 1: Number of observations by work shift and hospital

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

For peer review only

Table 2: 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	◆
Immed. 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	*
Heart rate (HR) at 30 min	1.5	100	4.4	5.8	0.52	100	<0.001	*
Blood pressure (BP) at 30 min	3.0	96.8	2.8	17.7	0.50	99.4	<0.001	*
Uterine contr. 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 contr. 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								
Post-partum hemorrhage	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 lower performance.

*: Statistically significant, supporting hypothesis of errors showing higher performance.

Table 3: Results from all shifts by hospital

	Hosp	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	
Immed. breast feeding	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 contr. @ 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 lower performance.

*: Statistically significant, supporting hypothesis of errors showing higher performance.

Table 4: Results for all hospitals by work shift

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
Immed breast feeding	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 cont @ 60 min	Morning	19	95	4	22	0.50	99	<.001 ◆
	Evening	9	100	3	12	0.51	100	<.001 ◆
	Night	10	100	5	14	0.52	100	<.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 lower performance.

*: Statistically significant supporting hypothesis of errors showing higher performance.

Table 5: Comparison of vital signs, blood loss and uterine contraction monitoring between post-partum hemorrhage cases and controls

Post-partum hemorrhage	At 30 minutes post-partum				At 60 minutes post-partum			
	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

HR: Heart rate, BP: blood pressure, UC: uterine contraction, BL: blood loss.

References

1. Franco L, Marquez L, Ethier K, Balsara Z, Isenhour W. Results of collaborative improvement: effects on health outcomes and compliance with evidence-based standards in 27 applications in 12 countries. Bethesda, MD. 2009.
2. Kelley E, Kelley AG, Simpara CH, Sidibe O, Makinen M. The impact of self-assessment on provider performance in Mali. *Int J Health Plann Manage*. 2003 Jan-Mar;18(1):41-8.
3. Vos L, Duckers ML, Wagner C, van Merode GG. 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, Betran AP, Harris-Requejo J, Hakimi F, et al. Cesarean delivery surveillance system at a maternity hospital in Kabul, Afghanistan. *Int J Gynaecol Obstet*. 2009 Jan;104(1):14-7.
5. Lain SJ, Hadfield RM, Raynes-Greenow CH, Ford JB, Mealing NM, Algert CS, et al. Quality of data in perinatal population health databases: a systematic review. *Med Care*. 2012 Apr;50(4):e7-20.
6. Hermida J, Broughton EI, Miller Franco L. Validity of self-assessment in a quality improvement collaborative in Ecuador. *Int J Qual Health Care*. 2011 Dec;23(6):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(6):489-98.
8. USAID Health Care Improvement Project. USAID HCI Afghanistan Newsletter, 2nd edition. 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 May;91(5):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, Alkozai AS, Cox J, Gardiner J, et al. Difficulties leaving home: a cross-sectional study of delays in seeking emergency obstetric care in Herat, Afghanistan. *Soc Sci Med*. 2011 Oct;73(7):1003-13.
15. Broun D, Debionne E, Ghane S, Ickx P, Ickx-Laumonier L, Lanlo P, 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 Jul-Aug;50(4):e31-5.
17. Kim YM, Tappis H, Zainullah P, Ansari N, Evans C, Bartlett L, 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.
18. Dott MM, Orakail N, Ebadi H, Hernandez F, MacFarlane K, Riley PL, et al. Implementing a facility-based maternal and perinatal health care surveillance system in Afghanistan. *J Midwifery Womens Health*. 2005 Jul-Aug;50(4):296-300.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
19. Clayton HB, Sappenfield WM, Gulitz E, Mahan CS, Petersen DJ, Stanley KM, 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 Jun 20.
20. Schnelle JF, Osterweil D, Simmons SF. Improving the quality of nursing home care and medical-record accuracy with direct observational technologies. *Gerontologist.* 2005 Oct;45(5):576-82.
21. De Amici D, Klersy C, Ramajoli F, Brustia L, Politi P. Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. *Control Clin Trials.* 2000 Apr;21(2):103-14.
22. Kohli E, Ptak J, Smith R, Taylor E, Talbot EA, Kirkland KB. Variability in the Hawthorne effect with regard to hand hygiene performance in high- and low-performing inpatient care units. *Infect Control Hosp Epidemiol.* 2009 Mar;30(3):222-5.

1
2
3
4 Dear Editors,

5 We have responded to the following questions concerning conflicts of interest relevant to this
6 manuscript.
7

8 1. Have you in the past five years accepted the following from an organisation that may in any way gain
9 or lose financially from the results of your study or the conclusions of your review, editorial, or letter:
10

- 11 • NO Reimbursement for attending a symposium?
- 12 • NO A fee for speaking?
- 13 • NO A fee for organising education?
- 14 • NO Funds for research?
- 15 • NO Funds for a member of staff?
- 16 • NO Fees for consulting?
- 17
- 18
- 19
- 20
- 21

22 2. Have you in the past five years been employed by an organisation that may in any way gain or lose
23 financially from the results of your study or the conclusions of your review, editorial, or letter?

24 NO
25

26 3. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from
27 the results of your study or the conclusions of your review, editorial, or letter?

28 NO
29

30 4. Have you acted as an expert witness on the subject of your study, review, editorial, or letter?

31 NO
32

33 5. Do you have any other competing financial interests? If so, please specify.

34 NO
35

36 Sincerely,

37 Edward Broughton
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Contributorship Statement

Manuscript title: How Accurate Are Medical Record Data in Afghanistan's Maternal Health Facilities? An observational validity study.

Authors: Edward I Broughton, MPH, PhD
Abdul Naser Ikram, MD, MPH
Ihsanullah Sahak, MD

All three authors listed above fulfilled all three of the ICMJE guidelines for authorship which are:

- 1) substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data
- 2) drafting the article or revising it critically for important intellectual content
- 3) final approval of the version to be published.

EB and AI conceived the idea for the study. EB drafted the protocol with input from AI and IS. AI organized submission to the Afghanistan IRB. EB submitted to the US IRB. AI and IS were responsible for organizing data collection and data entry and cleaning. EB analyzed the data. EB, AI and IS drafted the paper with EB taking the lead.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	Separate
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	3
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	3
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants	4
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4
Bias	9	Describe any efforts to address potential sources of bias	5&9
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4-5
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	5
		(b) Describe any methods used to examine subgroups and interactions	5
		(c) Explain how missing data were addressed	-
		(d) If applicable, describe analytical methods taking account of sampling strategy	5
		(e) Describe any sensitivity analyses	-
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5-6
		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	-
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	6
Outcome data	15*	Report numbers of outcome events or summary measures	6
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6
		(b) Report category boundaries when continuous variables were categorized	6
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	6
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	6
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	8-9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8
Generalisability	21	Discuss the generalisability (external validity) of the study results	8
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	9

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.



**How Accurate Are Medical Record Data in Afghanistan's
Maternal Health Facilities? An observational Validity Study**

Journal:	<i>BMJ Open</i>
Manuscript ID:	bmjopen-2013-002554.R1
Article Type:	Research
Date Submitted by the Author:	07-Mar-2013
Complete List of Authors:	Broughton, Edward; University Research Co., LLC, Research and Evaluation; Johns Hopkins School of Public Health, International Health Ikram, Abdul; University Research Co., LLC, Sahak, Ihsanullah; University Research Co., LLC,
Primary Subject Heading:	Obstetrics and gynaecology
Secondary Subject Heading:	Health informatics, Epidemiology, Reproductive medicine, Health services research, Research methods
Keywords:	Maternal medicine < OBSTETRICS, Fetal medicine < OBSTETRICS, AUDIT, EPIDEMIOLOGY, Organisation of health services < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

1
2
3
4 **How Accurate Are Medical Record Data in Afghanistan's Maternal Health**
5 **Facilities? An observational Validity Study**
6
7
8
9

10 Corresponding author address: Edward I Broughton, MPH, PhD
11 University Research Co., LLC
12 7200 Wisconsin Ave Suite 600
13 Bethesda MD 20814
14
15
16 Ph: +1-301-941-8533
17 Fax: +1-301-941-8427
18
19

20
21
22 Co-authors Abdul Naser Ikram, MD, MPH
23 University Research Co., LLC (Afghanistan)
24 Research and Evaluation Department
25 Kabul, Afghanistan
26
27
28 Ihsanullah Sahak, MD
29 University Research Co., LLC (Afghanistan)
30 Research and Evaluation Department
31 Kabul, Afghanistan
32
33
34

35
36 Word Count: 3816
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 work-shifts.

Design/Setting

This observational cross-sectional 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. We determined sensitivity, specificity, areas under the receiver operator characteristics curves (AUROC), proportions correctly classified and the tendency to make performance seem better than it actually was.

Main Results

Six hundred observations across the three shifts and three hospitals showed high compliance with active management of the third stage of labor, measuring blood loss and uterine contraction at 30 minutes, 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% versus 29%, $P = 0.143$). Compliance, sensitivity and specificity varied less among the three shifts than among hospitals.

Conclusions

Medical records 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 is heavily reliant on collecting accurate data on processes and outcomes of care, substantial improvement is needed in medical record accuracy.

Word count: 274/300

Running title: Medical record accuracy in Afghanistan

Article summary

Article focus

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

Key Messages

- Medical records 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 very high or low compliance score, decreasing the usefulness of some sensitivity or specificity measures
- Clinician behavior may have changed from normal due to the Hawthorn Effect

Introduction

Quality improvement (QI) in health care 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 is 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 health care 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 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 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 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.
 - a. Does the accuracy of medical record data vary between facilities?
 - b. Does the accuracy of medical records vary among the three work shift 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

1
2
3 deliveries taking place in participating facilities and recording their observations on a written checklist. They
4 checked their observations against data entered into corresponding patient medical records and facility
5 registers 24 or more hours after the observed delivery to ensure adequate time for the records to be
6 completed by the attending clinician.
7
8

9 ***Sampling***

10 The sampling frame was any vaginal delivery that took place in one of the three maternity facilities on the
11 days in which the observations took place. Three observers were assigned to each of the operational
12 delivery rooms for the three shifts in a 24-hour period of consecutive days until an adequate sample size
13 was achieved. The same three observers were used in each of the three hospitals. Deliveries were excluded
14 if they occurred outside the delivery rooms (those occurring in other rooms in the hospital or before arrival
15 at the facility) and deliveries that progressed to caesarian sections. The sample size calculation was based
16 on a level of agreement between observations and the patient medical record / registers of 50% and the
17 ability to detect a 15% difference between agreement in the referral hospitals and the general hospitals
18 with an alpha of 0.05 and a power of 0.8. This yielded a minimum sample size of 186 in each facility and
19 from each workshift. We aimed for approximately 200 in each group.
20
21
22
23
24

25 ***Data Collection***

26 Performance of the following 17 tasks was recorded from observations then checked against the patient
27 medical record. These tasks were chosen because they are all considered standard practice and part of
28 the clinical guidelines for vaginal delivery by the Afghanistan MOPH
29

- 30 • Active management of third stage of labor (AMTSL): administration of a uterotonic, controlled
31 cord contraction and uterine massage (performance of all three elements of AMTSL for the
32 case)
- 33 • Uterotonic administration done in the first minute following delivery
- 34 • Controlled traction of the umbilical cord
- 35 • Uterine massage following delivery
- 36 • Drying and wrapping of the newborn
- 37 • Umbilical cord care
- 38 • Breastfeeding within the first hour following delivery
- 39 • Measuring maternal blood loss in 30 minutes after delivery
- 40 • Monitoring of woman's pulse rate at 30 minutes after delivery
- 41 • Monitoring of woman's blood pressure at 30 minutes after delivery
- 42 • Monitoring of uterine contraction at 30 minutes after delivery
- 43 • Monitoring of woman's pulse rate at 60 minutes after delivery
- 44 • Monitoring of woman's blood pressure at 60 minutes after delivery
- 45 • Monitoring of uterine contraction at 60 minutes after delivery
- 46 • Inspection for laceration
- 47 • Newborn eye care
- 48 • Apgar score at five minutes after delivery
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

- Woman's diagnosis of post-partum hemorrhage (blood loss >500 ml)
- Neonatal asphyxia
- Neonatal death within one hour of delivery
- Stillbirth
- Maternal death within six hours 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 a 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 work-shift 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 University Research Co., LLC. Data collectors observing deliveries were all female medical doctors specialized in obstetrics and gynecology 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 anonymized 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 commencing the study.

Data Analysis

Results were entered into an Excel database with double entry to ensure accuracy. Analyses were conducted using STATA version 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 (AUROC), 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 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 2, 60% of the discordances were when the medical record indicated that AMTSL was completed but the observer reported that it was not

1
2
3 actually done. This proportion is not significantly different to the 50% expected if the errors occurred at
4 random as determined by Fisher's Exact test ($p=0.138$)
5
6
7

8 9 Results

10
11 A total of 600 observations were completed with close to equal distribution across the three shifts and
12 three hospitals (Table 1). Below are presented the results for all variables in all hospitals (Table 2) as well
13 as the results from five indicators selected to represent high, medium and low compliance/occurrence
14 divided by hospitals and work shift (Tables 3 and 4). Full tables including all variables reported by
15 hospitals and work shift are available online but not included here due to their size.
16
17

18 **Overall:** There was high compliance with the three elements of AMTSL, measurement of blood loss and
19 uterine contraction at 30 minutes, cord care, drying and wrapping newborns and Apgar scores. There
20 was low compliance with taking the mothers' vital signs following delivery, especially one hour after
21 delivery. In many cases of compliance with quality of care indicators, specificity was lower than
22 sensitivity, while in reporting adverse outcomes of stillbirths, neonatal death, asphyxia and post-partum
23 hemorrhage, specificity was higher than sensitivity. Of the 16 variables in the medical charts and birth
24 registries for which there was a statistically significant indication of biased errors, 11 were of the type
25 that made the clinicians' performance or the clinical outcomes seem better than they actually were and
26 five were of the type that made them look worse (71% versus 29%, $p = 0.143$). There were no maternal
27 deaths observed or recorded in the medical records among the women sampled.
28
29

30
31 **Hospitals:** All hospitals had high compliance with the three elements of AMTSL and high sensitivity in
32 the medical records. However, specificity was low in all three. Compliance with breastfeeding within the
33 first hour after delivery was variable, and sensitivity was high in the maternity hospital and low in the
34 general and provincial hospitals. Monitoring of uterine contraction one hour after delivery had low
35 compliance in all hospitals, with high sensitivity and low specificity. The proportion of stillbirths was
36 highest in the provincial hospital and lowest (50% lower proportion, $p=0.247$) in the general hospital.
37 Sensitivity and specificity were both relatively high for this indicator. There was variability in compliance
38 with neonatal eye care with the maternity hospital having high sensitivity, low specificity and the
39 general and provincial hospital having moderate sensitivity and specificity. The AUROC for all indicators
40 was less than 0.6 except for stillbirths, which was above 0.92 in all hospitals (Table 3).
41
42

43
44 **Work shift:** Compliance, sensitivity and specificity varied less among the three shifts than among the
45 three hospitals. The greatest variation in compliance was in uterine contraction at one hour after
46 delivery while the greatest variation in AUROC was in neonatal eye care. Errors analyzed by workshift
47 did not appear to follow a pattern of errors of self-enhancement of performance (Table 4).
48
49

50
51 **Post-partum hemorrhage (PPH):** Although the study was not powered to determine whether there
52 were differences in the way women diagnosed with PPH were treated, we included a separate analysis
53 of those nine cases. There were slightly higher proportions of women with PPH who had their uterine
54 contraction, blood loss and vital signs measured but those proportions were still low (Table 5).
55
56
57

Discussion

There have been substantial investments in improving the quality of care with the goal of achieving better maternal 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 mortality 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 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 post-partum period was observed rather than the 28 days as per the definition of neonatal mortality and again, a lower occurrence of death was expected in this setting compared to the country as a whole.

Sensitivity was high for indicators of compliance with standards of care with the exception of breastfeeding in first hour after delivery, cord care and drying and wrapping of the newborn. Failing to record these accurately when they were actually done is not likely to have a major impact on the safety of the care provided but does lead to 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 done 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 a 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 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 to those taken 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 minutes, 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 breast feeding 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, it is unclear the reason for this large variation.

While some hospitals in urban centers in Afghanistan are overstaffed, there tends to be too few female staff overall, and 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 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 was 85.1% and 72.4% respectively compared to the gold standard of observation of the surgery. Another study found a discrepancy of around 30% in identifying patients at risk for under-nutrition between observations done by researchers and records of the evaluation in the patients' medical charts (18). We found no benchmark study using observations of deliveries to test 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 the all hospitals in Afghanistan and the performance in the participating 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 organizations conducting quality improvement 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 surveillance of maternal health care 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 or clinical record keeping, allowing adequate time and staff support and

1
2
3 fostering an atmosphere of not assigning punishment or blame for errors in clinical practice may lead to
4 more accurate medical records (21, 22). Given the importance of the accuracy of medical records to the
5 success of improvement efforts, implementers should use the same approaches to addressing record
6 keeping as they do for improving the processes and outcomes of clinical care. Those involved with
7 surveillance based on medical records should take into account the inaccuracies found in this study
8 when interpreting their own results.
9
10

11 12 13 **Limitations**

14
15 Compliance was mostly high for the quality measures and occurrence was low for the adverse outcomes
16 such as stillbirths. While this is a positive result for the clinicians, it does not make for an optimal study
17 of the quality of the medical records of clinical processes and outcomes. For example, cord traction
18 conducted to standards following delivery of the neonate occurred in 597 of the 600 observed deliveries
19 (99.5%). This left only three opportunities out of 600 deliveries for clinicians to accurately record not
20 conducting cord contraction to compliance with standards. Missing any or all of these few opportunities
21 gives a low or zero specificity and therefore an AUROC close to 0.5. This was the case for several
22 indicators even though their proportions correctly classified were high. However, with indicators where
23 the compliance or occurrence was not at the extremes, such as the 49.3% compliance with immediate
24 breastfeeding and the 80.5% compliance with neonatal eye care, the results for the AUROC were still
25 not very high and not greatly different to the results obtained for the other indicators. The proportions
26 correctly classified for those indicators were correspondingly low. A larger sample size would have
27 lessened this effect.
28
29

30
31 The Hawthorne Effect, defined as the change in the behavior being observed due to the known presence
32 of the observer (23, 24), may have improved compliance with quality of care indicators. Clinician
33 participants were initially aware of the observer because they were required to sign the informed
34 consent form. However, they did not know that the accuracy of the medical records would also be
35 checked. Also, the delivery rooms where observations took place are large open areas and clinicians are
36 used to operating where many people observe their activities. We do not consider it likely that the
37 Hawthorne Effect had a significant influence over the accuracy of the medical records.
38
39

40
41 We did not distinguish between data that were incorrectly reported and data that were missing from
42 the chart. The reason was because regardless of whether clinical information is missing or incorrectly
43 recorded, the patients' care may be compromised and the medical record cannot be trusted as a
44 reflection of reality. Had we considered only the accuracy of the non-missing data in the medical
45 records, they would have appeared to be of better quality for clinical care than they actually were.
46 However, it could be argued that we should have considered missing and erroneously recorded data
47 separately.
48
49

50
51 Observations from the researchers were considered the "gold standard". These were three medical
52 doctors with extensive experience in maternal and neonatal clinical care, and they received training on
53 how to conduct their observations, including a trial of observing deliveries. However, there was no check
54 in this study of intra-tester or inter-tester reliability of these three observers. If observers did make
55
56
57
58
59
60

1
2
3 errors, there is no reason these would have biased the results for the accuracy of the records one way or
4 the other.
5

6 7 **Conclusion**

8 Compliance was high in some indicators of maternal and neonatal health quality of care but low for
9 others. The accuracy of medical records in capturing clinical activities and outcomes was generally poor.
10 Because the success of activities to improve the quality of care in these settings is heavily reliant on
11 collecting accurate data on processes and outcomes of care, substantial attention needs to be paid to
12 improving medical record accuracy.
13
14

15 16 17 **Data Sharing**

18 Additional data can be obtained by e-mailing the corresponding author.
19
20

21 22 23 **Acknowledgements**

24 This study was supported by the American people through the United States Agency for International
25 Development (USAID) and its Health Care Improvement Project (HCI). HCI is managed by University
26 Research Co., LLC (URC) under the terms of Contract Number GHN-I-03-07-00003-00. We thank the
27 directors of the participating hospitals for their generous cooperation. We also thank Stacie Gobin for
28 her assistance with analysis.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1: Number of observations by work shift and hospital

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

For peer review only

Table 2: 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	◆
Immed. 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	*
Heart rate (HR) at 30 min	1.5	100	4.4	5.8	0.52	100	<0.001	*
Blood pressure (BP) at 30 min	3.0	96.8	2.8	17.7	0.50	99.4	<0.001	*
Uterine contr. 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 contr. 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								
Post-partum hemorrhage	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 lower performance.

*: Statistically significant, supporting hypothesis of errors showing higher performance.

Table 3: Results from all shifts by hospital

	Hosp	Compliance/ Occurrence	Sensitivity	Specificity	Correctly classified	AUROC	Self- enhancement	P- value	
		(%)	(%)	(%)	(%)		errors		
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	
Immed. breast feeding	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 contr. @ 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 lower performance.

*: Statistically significant, supporting hypothesis of errors showing higher performance.

M: Maternity Hospital; G: General Hospital; P: Provincial Hospital

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Table 4: Results for all hospitals by work shift

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
Immed breast feeding	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 cont @ 60 min	Morning	19	95	4	22	0.50	99	<.001 ◆
	Evening	9	100	3	12	0.51	100	<.001 ◆
	Night	10	100	5	14	0.52	100	<.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 lower performance.
 * : Statistically significant supporting hypothesis of errors showing higher performance.

Table 5: Comparison of vital signs, blood loss and uterine contraction monitoring between post-partum hemorrhage cases and controls

Post-partum hemorrhage	At 30 minutes post-partum				At 60 minutes post-partum			
	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

HR: Heart rate, BP: blood pressure, UC: uterine contraction, BL: blood loss.

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. 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 Jan-Mar;18(1):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, Meriardi M, Guidotti RJ, et al. Cesarean delivery surveillance system at a maternity hospital in Kabul, Afghanistan. *Int J Gynaecol Obstet*. 2009 Jan;104(1):14-7.
5. Lain SJ, Hadfield RM, Raynes-Greenow CH, et al. Quality of data in perinatal population health databases: a systematic review. *Med Care*. 2012 Apr;50(4):e7-20.
6. Hermida J, Broughton EI, Miller Franco L. Validity of self-assessment in a quality improvement collaborative in Ecuador. *Int J Qual Health Care*. 2011 Dec;23(6):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(6):489-98.
8. USAID Health Care Improvement Project. USAID HCI Afghanistan Newsletter, 2nd edition. 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. Tegute I, Maiga AW, Leppert PC. Maternal and neonatal outcomes of grand multiparas over two decades in Mali. *Acta Obstet Gynecol Scand*. 2012 May;91(5):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 Oct;73(7):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 Jul-Aug;50(4):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 Feb;100(3):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 May-Jun;3(3):140-5.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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 Jul-Aug;50(4):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 Jun 20.
 22. Schnelle JF, Osterweil D, Simmons SF. Improving the quality of nursing home care and medical-record accuracy with direct observational technologies. *Gerontologist*. 2005 Oct;45(5):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 Apr;21(2):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 Mar;30(3):222-5.

1
2
3
4 **How Accurate Are Medical Record Data in Afghanistan's Maternal Health**
5 **Facilities? An observational Validity Study**
6
7
8
9

10
11 Corresponding author address: Edward I Broughton, MPH, PhD
12 University Research Co., LLC
13 7200 Wisconsin Ave Suite 600
14 Bethesda MD 20814
15
16 Ph: +1-301-941-8533
17 Fax: +1-301-941-8427
18
19

20
21
22 Co-authors Abdul Naser Ikram, MD, MPH
23 University Research Co., LLC (Afghanistan)
24 Research and Evaluation Department
25 Kabul, Afghanistan
26
27
28 Ihsanullah Sahak, MD
29 University Research Co., LLC (Afghanistan)
30 Research and Evaluation Department
31 Kabul, Afghanistan
32
33

34
35
36 Word Count: **3816**
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 work-shifts.

Design/Setting

This observational cross-sectional 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. We determined sensitivity, specificity, areas under the receiver operator characteristics curves (AUROC), proportions correctly classified and the tendency to make performance seem better than it actually was.

Main Results

Six hundred observations across the three shifts and three hospitals showed high compliance with active management of the third stage of labor, measuring blood loss and uterine contraction at 30 minutes, 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% versus 29%, $P = 0.143$). Compliance, sensitivity and specificity varied less among the three shifts than among hospitals.

Conclusions

Medical records 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 is heavily reliant on collecting accurate data on processes and outcomes of care, substantial improvement is needed in medical record accuracy.

Word count: 274/300

Running title: Medical record accuracy in Afghanistan

Article summary

Article focus

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

Key Messages

- Medical records 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 very high or low compliance score, decreasing the usefulness of some sensitivity or specificity measures
- Clinician behavior may have changed from normal due to the Hawthorn Effect

Introduction

Quality improvement (QI) in health care 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 is 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 health care 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 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 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 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. a. Does the accuracy of medical record data vary between facilities?
b. Does the accuracy of medical records vary among the three work shift 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

1
2
3 deliveries taking place in participating facilities and recording their observations on a written checklist. They
4 checked their observations against data entered into corresponding patient medical records and facility
5 registers 24 or more hours after the observed delivery to ensure adequate time for the records to be
6 completed by the attending clinician.
7
8

9 **Sampling**

10 The sampling frame was any vaginal delivery that took place in one of the three maternity facilities on the
11 days in which the observations took place. Three observers were assigned to each of the operational
12 delivery rooms for the three shifts in a 24-hour period of consecutive days until an adequate sample size
13 was achieved. The same three observers were used in each of the three hospitals. Deliveries were excluded
14 if they occurred outside the delivery rooms (those occurring in other rooms in the hospital or before arrival
15 at the facility) and deliveries that progressed to caesarian sections. The sample size calculation was based
16 on a level of agreement between observations and the patient medical record / registers of 50% and the
17 ability to detect a 15% difference between agreement in the referral hospitals and the general hospitals
18 with an alpha of 0.05 and a power of 0.8. This yielded a minimum sample size of 186 in each facility and
19 from each workshift. We aimed for approximately 200 in each group.
20
21
22
23
24

25 **Data Collection**

26 Performance of the following 17 tasks was recorded from observations then checked against the patient
27 medical record. These tasks were chosen because they are all considered standard practice and part of
28 the clinical guidelines for vaginal delivery by the Afghanistan MOPH
29

- 30 • Active management of third stage of labor (AMTSL): administration of a uterotonic, controlled
31 cord contraction and uterine massage (performance of all three elements of AMTSL for the
32 case)
- 33 • Uterotonic administration done in the first minute following delivery
- 34 • Controlled traction of the umbilical cord
- 35 • Uterine massage following delivery
- 36 • Drying and wrapping of the newborn
- 37 • Umbilical cord care
- 38 • Breastfeeding within the first hour following delivery
- 39 • Measuring maternal blood loss in 30 minutes after delivery
- 40 • Monitoring of woman's pulse rate at 30 minutes after delivery
- 41 • Monitoring of woman's blood pressure at 30 minutes after delivery
- 42 • Monitoring of uterine contraction at 30 minutes after delivery
- 43 • Monitoring of woman's pulse rate at 60 minutes after delivery
- 44 • Monitoring of woman's blood pressure at 60 minutes after delivery
- 45 • Monitoring of uterine contraction at 60 minutes after delivery
- 46 • Inspection for laceration
- 47 • Newborn eye care
- 48 • Apgar score at five minutes after delivery
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

1
2
3 The following data were recorded during observations then checked against the birth register:
4

- 5 • Woman's diagnosis of post-partum hemorrhage (blood loss >500 ml)
- 6 • Neonatal asphyxia
- 7 • Neonatal death within one hour of delivery
- 8 • Stillbirth
- 9 • Maternal death within six hours of delivery

10
11
12 Medical records were considered correctly classified only if they were completed and agreed with
13 what the research assistant observed. For example, if the observer saw that a uterotonic was
14 administered in the first minute following delivery but it was reported as not administered or the
15 information on uterotonic administration was completely missing from the chart, then this was
16 considered incorrectly classified.
17
18

19 The hospital and the work-shift in which the delivery occurred were also recorded.
20

21 **Ethical considerations:**

22
23 The study was approved by the institutional review boards of both the Afghan MoPH and University
24 Research Co., LLC. Data collectors observing deliveries were all female medical doctors specialized in
25 obstetrics and gynecology and dressed in scrubs as appropriate for the delivery room. As the settings
26 for the study were teaching hospitals, there are often personnel observing patient care without being
27 part of that care. Also, the nature of the delivery rooms in all three facilities allowed unobtrusive
28 observations with no interference to clinical care. Data were anonymized for analysis. Delivering
29 mothers were informed verbally of the nature of the study and gave written consent or thumb-print for
30 those who were illiterate. Participant health workers who were observed also signed a written consent
31 form before participating in the study. If the observers saw any practice dangerous to the delivering
32 mother or the neonate, they informed the clinician delivering the care. Permission from hospital
33 administrators and MoPH officials was obtained prior to commencing the study.
34
35
36
37

38 **Data Analysis**

39
40 Results were entered into an Excel database with double entry to ensure accuracy. Analyses were
41 conducted using STATA version 11. P-values were calculated for statistical significance. We calculated
42 sensitivity (proportion of cases where performance to standard was accurately reported) and specificity
43 (proportion of cases where non-performance to standard was accurately reported). We also calculated
44 the area under the receiver operator characteristics (AUROC), combining the proportion of true and
45 false positives to give an indication of the usefulness of the medical record, where 1.0 is a perfect
46 indicator while 0.5 is a test no better than a guess. We recorded overall compliance with the indicator
47 and raw agreement between observers and the medical records. Self-enhancement errors are the
48 proportion of discordances between the medical records and observers where the medical record shows
49 a positive result: either compliance with an indicator such as AMTSL or the non-occurrence of an
50 adverse outcome such as PPH. The p-value is for the test of whether or not this proportion is 50% as
51 would be expected if errors occurred at random. For example, in Table 2, 60% of the discordances were
52 when the medical record indicated that AMTSL was completed but the observer reported that it was not
53
54
55
56
57
58
59
60

1
2
3 actually done. This proportion is not significantly different to the 50% expected if the errors occurred at
4 random as determined by Fisher's Exact test ($p=0.138$)
5
6
7

8 9 Results

10
11 A total of 600 observations were completed with close to equal distribution across the three shifts and
12 three hospitals (Table 1). Below are presented the results for all variables in all hospitals (Table 2) as well
13 as the results from five indicators selected to represent high, medium and low compliance/occurrence
14 divided by hospitals and work shift (Tables 3 and 4). Full tables including all variables reported by
15 hospitals and work shift are available online but not included here due to their size.
16
17

18 **Overall:** There was high compliance with the three elements of AMTSL, measurement of blood loss and
19 uterine contraction at 30 minutes, cord care, drying and wrapping newborns and Apgar scores. There
20 was low compliance with taking the mothers' vital signs following delivery, especially one hour after
21 delivery. In many cases of compliance with quality of care indicators, specificity was lower than
22 sensitivity, while in reporting adverse outcomes of stillbirths, neonatal death, asphyxia and post-partum
23 hemorrhage, specificity was higher than sensitivity. Of the 16 variables in the medical charts and birth
24 registries for which there was a statistically significant indication of biased errors, 11 were of the type
25 that made the clinicians' performance or the clinical outcomes seem better than they actually were and
26 five were of the type that made them look worse (71% versus 29%, $p = 0.143$). There were no maternal
27 deaths observed or recorded in the medical records among the women sampled.
28
29

30
31 **Hospitals:** All hospitals had high compliance with the three elements of AMTSL and high sensitivity in
32 the medical records. However, specificity was low in all three. Compliance with breastfeeding within the
33 first hour after delivery was variable, and sensitivity was high in the maternity hospital and low in the
34 general and provincial hospitals. Monitoring of uterine contraction one hour after delivery had low
35 compliance in all hospitals, with high sensitivity and low specificity. The proportion of stillbirths was
36 highest in the provincial hospital and lowest (50% lower proportion, $p=0.247$) in the general hospital.
37 Sensitivity and specificity were both relatively high for this indicator. There was variability in compliance
38 with neonatal eye care with the maternity hospital having high sensitivity, low specificity and the
39 general and provincial hospital having moderate sensitivity and specificity. The AUROC for all indicators
40 was less than 0.6 except for stillbirths, which was above 0.92 in all hospitals (Table 3).
41
42

43
44 **Work shift:** Compliance, sensitivity and specificity varied less among the three shifts than among the
45 three hospitals. The greatest variation in compliance was in uterine contraction at one hour after
46 delivery while the greatest variation in AUROC was in neonatal eye care. Errors analyzed by workshift
47 did not appear to follow a pattern of errors of self-enhancement of performance (Table 4).
48
49

50
51 **Post-partum hemorrhage (PPH):** Although the study was not powered to determine whether there
52 were differences in the way women diagnosed with PPH were treated, we included a separate analysis
53 of those nine cases. There were slightly higher proportions of women with PPH who had their uterine
54 contraction, blood loss and vital signs measured but those proportions were still low (Table 5).
55
56
57

Discussion

There have been substantial investments in improving the quality of care with the goal of achieving better maternal 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 mortality 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 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 post-partum period was observed rather than the 28 days as per the definition of neonatal mortality and again, a lower occurrence of death was expected in this setting compared to the country as a whole.

Sensitivity was high for indicators of compliance with standards of care with the exception of breastfeeding in first hour after delivery, cord care and drying and wrapping of the newborn. Failing to record these accurately when they were actually done is not likely to have a major impact on the safety of the care provided but does lead to 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 done 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 a 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 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 to those taken 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 minutes, 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 breast feeding 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, it is unclear the reason for this large variation.

While some hospitals in urban centers in Afghanistan are overstaffed, there tends to be too few female staff overall, and 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 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 was 85.1% and 72.4% respectively compared to the gold standard of observation of the surgery. Another study found a discrepancy of around 30% in identifying patients at risk for under-nutrition between observations done by researchers and records of the evaluation in the patients' medical charts (18). We found no benchmark study using observations of deliveries to test 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 the all hospitals in Afghanistan and the performance in the participating 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 organizations conducting quality improvement 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 surveillance of maternal health care 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 or clinical record keeping, allowing adequate time and staff support and

1
2
3 fostering an atmosphere of not assigning punishment or blame for errors in clinical practice may lead to
4 more accurate medical records (21, 22). Given the importance of the accuracy of medical records to the
5 success of improvement efforts, implementers should use the same approaches to addressing record
6 keeping as they do for improving the processes and outcomes of clinical care. Those involved with
7 surveillance based on medical records should take into account the inaccuracies found in this study
8 when interpreting their own results.
9
10

11 12 13 **Limitations**

14
15 Compliance was mostly high for the quality measures and occurrence was low for the adverse outcomes
16 such as stillbirths. While this is a positive result for the clinicians, it does not make for an optimal study
17 of the quality of the medical records of clinical processes and outcomes. For example, cord traction
18 conducted to standards following delivery of the neonate occurred in 597 of the 600 observed deliveries
19 (99.5%). This left only three opportunities out of 600 deliveries for clinicians to accurately record not
20 conducting cord contraction to compliance with standards. Missing any or all of these few opportunities
21 gives a low or zero specificity and therefore an AUROC close to 0.5. This was the case for several
22 indicators even though their proportions correctly classified were high. However, with indicators where
23 the compliance or occurrence was not at the extremes, such as the 49.3% compliance with immediate
24 breastfeeding and the 80.5% compliance with neonatal eye care, the results for the AUROC were still
25 not very high and not greatly different to the results obtained for the other indicators. The proportions
26 correctly classified for those indicators were correspondingly low. A larger sample size would have
27 lessened this effect.
28
29

30
31
32
33 The Hawthorne Effect, defined as the change in the behavior being observed due to the known presence
34 of the observer (23, 24), may have improved compliance with quality of care indicators. Clinician
35 participants were initially aware of the observer because they were required to sign the informed
36 consent form. However, they did not know that the accuracy of the medical records would also be
37 checked. Also, the delivery rooms where observations took place are large open areas and clinicians are
38 used to operating where many people observe their activities. We do not consider it likely that the
39 Hawthorne Effect had a significant influence over the accuracy of the medical records.
40
41
42

43 **We did not distinguish between data that were incorrectly reported and data that were missing from**
44 **the chart. The reason was because regardless of whether clinical information is missing or incorrectly**
45 **recorded, the patients' care may be compromised and the medical record cannot be trusted as a**
46 **reflection of reality. Had we considered only the accuracy of the non-missing data in the medical**
47 **records, they would have appeared to be of better quality for clinical care than they actually were.**
48 **However, it could be argued that we should have considered missing and erroneously recorded data**
49 **separately.**
50
51

52
53 Observations from the researchers were considered the "gold standard". These were three medical
54 doctors with extensive experience in maternal and neonatal clinical care, and they received training on
55 how to conduct their observations, including a trial of observing deliveries. However, there was no check
56 in this study of intra-tester or inter-tester reliability of these three observers. If observers did make
57
58
59
60

1
2
3 errors, there is no reason these would have biased the results for the accuracy of the records one way or
4 the other.
5

6 7 **Conclusion**

8 Compliance was high in some indicators of maternal and neonatal health quality of care but low for
9 others. The accuracy of medical records in capturing clinical activities and outcomes was generally poor.
10 Because the success of activities to improve the quality of care in these settings is heavily reliant on
11 collecting accurate data on processes and outcomes of care, substantial attention needs to be paid to
12 improving medical record accuracy.
13
14

15 16 17 **Data Sharing**

18 Additional data can be obtained by e-mailing the corresponding author.
19
20

21 22 23 **Acknowledgements**

24 This study was supported by the American people through the United States Agency for International
25 Development (USAID) and its Health Care Improvement Project (HCI). HCI is managed by University
26 Research Co., LLC (URC) under the terms of Contract Number GHN-I-03-07-00003-00. We thank the
27 directors of the participating hospitals for their generous cooperation. We also thank Stacie Gobin for
28 her assistance with analysis.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1: Number of observations by work shift and hospital

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

For peer review only

Table 2: 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	◆
Immed. 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	*
Heart rate (HR) at 30 min	1.5	100	4.4	5.8	0.52	100	<0.001	*
Blood pressure (BP) at 30 min	3.0	96.8	2.8	17.7	0.50	99.4	<0.001	*
Uterine contr. 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 contr. 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								
Post-partum hemorrhage	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 lower performance.

*: Statistically significant, supporting hypothesis of errors showing higher performance.

Table 3: Results from all shifts by hospital

	Hosp	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	
Immed. breast feeding	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 contr. @ 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 lower performance.

*: Statistically significant, supporting hypothesis of errors showing higher performance.

M: Maternity Hospital; G: General Hospital; P: Provincial Hospital

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Table 4: Results for all hospitals by work shift

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
Immed breast feeding	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 cont @ 60 min	Morning	19	95	4	22	0.50	99	<.001 ◆
	Evening	9	100	3	12	0.51	100	<.001 ◆
	Night	10	100	5	14	0.52	100	<.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 lower performance.

*: Statistically significant supporting hypothesis of errors showing higher performance.

For peer review only

Table 5: Comparison of vital signs, blood loss and uterine contraction monitoring between post-partum hemorrhage cases and controls

Post-partum hemorrhage	At 30 minutes post-partum				At 60 minutes post-partum			
	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

HR: Heart rate, BP: blood pressure, UC: uterine contraction, BL: blood loss.

References

1. Franco L, Marquez L, Ethier K, Balsara Z, Isenhower W. Results of collaborative improvement: effects on health outcomes and compliance with evidence-based standards in 27 applications in 12 countries. Bethesda, MD. 2009.
2. Kelley E, Kelley AG, Simpara CH, Sidibe O, Makinen M. The impact of self-assessment on provider performance in Mali. *Int J Health Plann Manage*. 2003 Jan-Mar;18(1):41-8.
3. Vos L, Duckers ML, Wagner C, van Merode GG. Applying the quality improvement collaborative method to process redesign: a multiple case study. *Implement Sci*. 2010;5:19.
4. Kandasamy T, Meriardi M, Guidotti RJ, Betran AP, Harris-Requejo J, Hakimi F, et al. Cesarean delivery surveillance system at a maternity hospital in Kabul, Afghanistan. *Int J Gynaecol Obstet*. 2009 Jan;104(1):14-7.
5. Lain SJ, Hadfield RM, Raynes-Greenow CH, Ford JB, Mealing NM, Algert CS, et al. Quality of data in perinatal population health databases: a systematic review. *Med Care*. 2012 Apr;50(4):e7-20.
6. Hermida J, Broughton EI, Miller Franco L. Validity of self-assessment in a quality improvement collaborative in Ecuador. *Int J Qual Health Care*. 2011 Dec;23(6):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(6):489-98.
8. USAID Health Care Improvement Project. USAID HCI Afghanistan Newsletter, 2nd edition. 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. Tegute I, Maiga AW, Leppert PC. Maternal and neonatal outcomes of grand multiparas over two decades in Mali. *Acta Obstet Gynecol Scand*. 2012 May;91(5):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, Alkozai AS, Cox J, Gardiner J, et al. Difficulties leaving home: a cross-sectional study of delays in seeking emergency obstetric care in Herat, Afghanistan. *Soc Sci Med*. 2011 Oct;73(7):1003-13.
15. Broun D, Debionne E, Ghane S, Ickx P, Ickx-Laumonier L, Lanlo P, 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 Jul-Aug;50(4):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 Feb;100(3):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 May-Jun;3(3):140-5.

19. Kim YM, Tappis H, Zainullah P, Ansari N, Evans C, Bartlett L, 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, Hernandez F, MacFarlane K, Riley PL, et al. Implementing a facility-based maternal and perinatal health care surveillance system in Afghanistan. *J Midwifery Womens Health*. 2005 Jul-Aug;50(4):296-300.
21. Clayton HB, Sappenfield WM, Gulitz E, Mahan CS, Petersen DJ, Stanley KM, 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 Jun 20.
22. Schnelle JF, Osterweil D, Simmons SF. Improving the quality of nursing home care and medical-record accuracy with direct observational technologies. *Gerontologist*. 2005 Oct;45(5):576-82.
23. De Amici D, Klersy C, Ramajoli F, Brustia L, Politi P. Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. *Control Clin Trials*. 2000 Apr;21(2):103-14.
24. Kohli E, Ptak J, Smith R, Taylor E, Talbot EA, Kirkland KB. Variability in the Hawthorne effect with regard to hand hygiene performance in high- and low-performing inpatient care units. *Infect Control Hosp Epidemiol*. 2009 Mar;30(3):222-5.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study’s design with a commonly used term in the title or the abstract	Separate
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	3
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	3
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants	4
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4
Bias	9	Describe any efforts to address potential sources of bias	5&9
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4-5
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	5
		(b) Describe any methods used to examine subgroups and interactions	5
		(c) Explain how missing data were addressed	-
		(d) If applicable, describe analytical methods taking account of sampling strategy	5
		(e) Describe any sensitivity analyses	-
Results			

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5-6
		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	-
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	6
Outcome data	15*	Report numbers of outcome events or summary measures	6
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6
		(b) Report category boundaries when continuous variables were categorized	6
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	6
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	6
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	8-9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8
Generalisability	21	Discuss the generalisability (external validity) of the study results	8
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	9

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.