**American Evaluation Association**

**Roundtable:** Clinical observations to assess quality of care: challenges and solutions in data collection and synthesis from a maternal health evaluation in Madagascar

**Authors:**

Eva Bazant, Sr. Monitoring, Evaluation and Research Advisor, Jhpiego, Baltimore, MD**;** Vandana Tripathi, Consultant, Johns Hopkins School of Public Health

**Location/Date:** Roundtable 931 to be held in 209 A on Saturday, Oct 27, 1:00 PM to 2:30 PM

|  |  |  |
| --- | --- | --- |
| observer.jpg |  |  |

**I. Context**

* Evaluating the quality of medical care is a growing priority across public health programs.
* In global health, programs increasingly aim to increase the quality of health services, not simply coverage [1] – e.g., knowing that a pregnant woman received an antenatal care visit provides no information about the interventions delivered during that visit.

Why Clinical Observation?

* Evaluations of quality of care (QoC) often rely on record reviews and other retrospective methods; or on patient and provider interviews regarding performance and satisfaction.
* However, medical records and health management information systems (HMIS) may be incomplete, especially in developing country/low-resource settings – e.g., ~33% of the world’s births and 67% of deaths are not captured by civil registration systems [2].
* Medical records, registers, and HMIS are not generally designed to provide high-quality data on quality of care.
* Record reviews are valuable but have also been found to have low inter-observer reliability [3-4].
* Interview-based data suffer from recall and desirability bias (on the part of providers) and courtesy bias (on the part of patients) [5-6]. Therefore, clinical observations are an essential way to assess the quality of actual health care.
* Clinical observation may represent the “best balance of sensitivity and specificity in QoC assessment” [7].

**II. MCHIP Case Study – Madagascar QoC Assessment**

Background: A seven-country study evaluated the active management of the third stage of labor using clinical observations in 2008[8]. Building on this, Jhpiego, as part of the USAID-supported Maternal and Child Health Integrated Program (MCHIP), conducted a more comprehensive assessment of routine and emergency maternal and newborn health care in eight sub-Saharan African countries [9]. These QoC assessments combined observations with facility inventories, provider interviews, and record reviews. A QoC Assessment was conducted in Madagascar in 2010. The full report of findings is available at: <http://www.mchip.net/node/820>.

Objectives:

* Guide program planning for improving quality of facility-based maternal and newborn care services in each country and at the global level
* Develop rapid and practical data collection tools and composite indicators
* Provide baseline estimates and, when repeated, endline estimates where the survey is part of an evaluation of program interventions

Methods Overview (Madagascar):

* Cross-sectional national assessment
* Sample: All facilities with >2 births per day; n=36 facilities met this criterion out of 294 with maternity services.
* Data sources:
	+ Facility inventory re PPH: Done in n=36 facilities
	+ Interview with providers on knowledge and experience: n=139
	+ Observations of labor & delivery clients: n=347
* Descriptive statistical analysis

Data Collection: Training and Processes

* 1 Supervisor-observer and 2 observers in a 3-member team; all are doctors or qualified midwives
* 20 observers trained over 8 days
* Training included:
	+ Updated in routine maternal and newborn health care skills and emergency obstetric and neonatal care skills to be observed (2 days)
	+ Study instruments and study design (6 days)
	+ Training and Practice with Windows Mobile Smart Phones
	+ Informed consent and to intervene in life-threatening situations
* Pre-test of tools using phones (2 days) at several health facilities
	+ Inter-rater reliability (percent agreement with trainer) assessed during training
* Data collected in October-November 2010
* Data collection during 3.5 day visits to each health facility
* Deliveries observed over 18 or 24 hours
* All deliveries that occurred during visit observed (or as many as possible at busy hospitals)
* Multiple observer teams deployed at very busy hospitals
* ANC cases consecutively observed
* Each health worker participated in 1 or 2 data collection efforts – observation, and/or interview/knowledge test

**III. Challenges and Solutions**

Challenges:

* Ethical issues
	+ Ethical approvals take time and are more involved for observations due to interactions with human subjects
	+ What if observers see sub-optimal care? Dangerous care?

Tools

* + Observation checklists adapted from approved service delivery guidelines are lengthy and may lack of reliability, not yet validated
	+ Data collection instruments can require interpretation – e.g., distinguishing between “No,” “Not applicable,” and blank responses
* Observer fatigue, particularly for multi-faceted “episodes of care” such as childbirth
* Analytical challenges:
	+ How to create summary scores? Which items to include and exclude?
	+ How to set a threshold for acceptable levels of quality
* Smart Phones: despite training and IT consultant, some observers reported that some data was not captured by the phones even when charged

Potential Solutions:

* Anticipating the need for ethical approvals and analysis; Institutional Review Board may consider the evaluation Exempt under category 5 of Federal Regulation 45 CFR 46
* Training observers well to promote validity and conducting exercises to promote inter-rater reliability
* Using structured methods to highlight key tasks
* Streamlining and pre-testing observation checklists
* Observer may intervene if dangerous care but data may be lost
* Triangulating data from several sources in attempt to tell coherent story
* Smart Phones: training and job aids for smart phone use, daily communication with IT manager

**IV. Discussion Questions - Roundtable:**

1. What **experience** do participants have in using direct clinical observation to assess quality of care in either training or research contexts? If so, please describe briefly:
* Location, Objective(s), Overview of methods, Perceptions of success/failure, Challenges encountered, Solutions identified
* What References did you find useful?
* What publications document the experience?
1. How have participants promoted and measured validity and **inter-rater reliability** among multiple observers using a long observation checklist – what are useful techniques, measurement methods, activities?
2. To what extent do participants agree with notion that **clinical observation provides a gold standard** relative to other quality of care assessment methods?
	* How did participants synthesize/**triangulate**/make sense of findings from several data sources, such as medical chart reviews, interviews AND observations of care?
3. Do participants have **solutions** for the challenges outlined above or other challenges discussed today?
	* Ethical issues, Inadequate tools, Observer fatigue, Analytical challenges

Questions – regarding the Madagascar Clinical Observation Experience:

* Can one observer truly capture all health interventions administered by a team?
	+ Example of “Golden minute” – Active management of third stage of childbirth and essential newborn care actions should both happen within one minute of newborn birth.
* Observations of PPH and PE/E cases (complications that rarely arose) were sometimes incomplete.
	+ Was this due to the difficulty of recording data on a large number of items (long checklist, items in a certain order) in a rapidly changing emergency clinical situation?
	+ Or was it due to observer fatigue?

**Appendix 1: Variability Affecting Clinical Observation Results – and How to Reduce**

Several types of variability can affect the accuracy and reliability of results. These include: a) variability in an individual subject, 2) *intra*-observer variability, 3) variation due to the method of measurement/ instrument/tool, 4) *inter*-observer variation, and 5) data entry errors. A description of these types of variability and their how they may present themselves during clinical observation is below.

|  |  |  |
| --- | --- | --- |
| **Type of variability** | **Description** | **Examples** |
| 1. variability in an individual subject | Over the course or a day or week, a subject may vary their practice/performance or a specific intervention | * The blood pressure reading of a patient may vary over the course of the day.
* Among health providers being observed performing a service, the “Hawthorne effect” may affect performance.[[1]](#footnote-1)
 |
| *2. intra*-observer variability | The same observer may observe differently over the course of a day, or from one facility to the next.  | * An observer may not be familiar with a long observation checklist with 100+ items. S/he may start out observing with little familiarity with the tool and after doing many observations, s/he is more familiar with the tool.
* An observer is fatigued after many hours of observations and is less observant; s/he does not “see” everything that is occurring.
* Daydreaming or personal biases may affect observation.
 |
| 3. variation due to the method of measurement/ instrument/tool | The observation checklist or the medium (such as paper vs. hand-(held electronic device) changes.  | * An observer may start out using paper checklists and then is asked to switch to recording data on a hand-held device.
* Observer is not familiar with how to scroll from beginning to end or switch page-views on hand-held device…
* A tool with a poor translation from one language to another may not accurately convey elements to be observed.
* An observer is asked to use a 20-item newborn resuscitation checklist and later is asked to switch to a more streamlined 15-item checklist.
 |
| *4. inter*-observer variation | Across different observers, observation is carried out differently | * Observers have different interpretations of key terms. One observer may think that some items on the checklist are less important to observe, compared to other observers.
* Some observers may be less comfortable with observation or with the content of the checklist than others.
* Some observers may be recording items as “n/a” while others may be leave items blank/missing
 |
| 5. data entry errors | Variation in how results are recorded and entered to electronic database  | * written responses are illegible
* responses are not complete (some cells left blank without noting why)
* Written data not carefully entered to a database, typos
 |

It is important to remember: a program or study manager can guard against these sources of variability in the planning phase. Some ideas are provided below.

**Table 2. Ways to reduce variability in accuracy and reliability of observation data**

|  |
| --- |
| **Actions** |
| *1. variability in an individual subject** Observe multiple cases in a health facility over several days if possible, not just one observation.
* Reassure the provider that the result of observation will not affect his/her job (if that is true) and that s/he can act as if the observer was not there
* Observer can be as unobtrusive as possible, not causing worry among providers being observed
* Ensure that the facility in-charge has given approval for observer to be there
 |
| 2. intra-observer variability* Ensure in the training workshop and pre-testing, that all observers know the content and skip patterns of the tools as much as possible; practice makes perfect; make sure tools have only essential items to begin with (best to validate the tools prior to giving them to observers)
* Come up with strategies for observers to be as observant as possible throughout the day, and ensure shifts are not too long, and that observers take short breaks when needed; discuss with observers what latent biases may affect observation
 |
| *3. variation due to the method of measurement/ instrument/tool** Ensure that the data collection instrument has been vetted by content experts and M&E experts. Consider skip patterns (for example, items that apply only to HIV positive clients). Validate the tool prior to using with many observers.
* Ensure that the data collection tools/checklists and medium does not change, and if it has to chance (for example, a hand-held device is lost) that this is noted for the record;
* Ensure accurate translation and review of the translation
 |
| 4. inter-observer variation* Ensure in the training workshop that the terms are explained and that there is a shared understanding of terms. Discuss how comfortable each observer is with observation in general and with the tools; adjust training workshop and pre-test as appropriate.
* Carry out an inter-observer reliability exercise and provide extra training (calibration, as needed).
* Absolutely clarify when “n/a” can be recorded.
 |
| *5. Data recording and data entry errors** Ensure that a supervisor reviews each completed checklist (the same day or next day) for completeness and any things written down or recorded incorrectly.
* Have a written plan for how the data will flow, from the observer to the supervisor to the data entry clerk. Have tips and guidelines for data entry. Cross-check 10% of records against the paper forms (if appropriate). Have data consistency and plausibility checks built into the hand-held device software programs, or the electronic database.
 |

**Appendix 2. References for Background Section**

1. van den Broek NR, Graham WJ. Quality of care for maternal and newborn health: The neglected agenda. *BJOG*. 2009 Oct;116 Suppl 1:18-21.

2. Pirkle CM, Dumont A, Zunzunegui MV. Criterion-based clinical audit to assess quality of obstetrical care in low- and middle-income countries: A systematic review. *Int J Qual Health Care*. 2011 Aug;23(4):456-63.

3. Partnership for Maternal, Child, and Neonatal Health. *Knowledge Summary 17: Civil Registration and Vital Statistics*. 2012. Available at: http://portal.pmnch.org/knowledge-summaries/ks17.

4. Hofer TP, Bernstein SJ, DeMonner S, Hayward RA. Discussion between reviewers does not improve reliability of peer review of hospital quality. *Med Care*. 2000;38(2):pp. 152-161.

5. Hardee K, Janowitz B, Stanback J, Villinski MT. What have we learned from studying changes in service guidelines and practices? *International Family Planning Perspectives*. 1998;24(2):pp. 84-90.

6. Glick P. How reliable are surveys of client satisfaction with healthcare services? Evidence from matched facility and household data in Madagascar. *Soc Sci Med*. 2009 Jan;68(2):368-79.

7. Hermida J, Nicholas DD, Blumenfeld SN. Comparative validity of three methods for assessment of the quality of primary health care. *Int J Qual Health Care*. 1999 Oct;11(5):429-33.

8. Stanton C, Armbruster D, Knight R, Ariawan I, Gbangbade S, Getachew A, Portillo JA, Jarquin D, Marin F, Mfinanga S, Vallecillo J, Johnson H, Sintasath D. Use of active management of the third stage of labour in seven developing countries. *Bull World Health Organ*. 2009 Mar;87(3):207-15.

## 9. Maternal and Child Health Integrated Program. *Measuring Maternal and Perinatal Quality of Care during Labor and Delivery.* Available at: http://www.mchip.net/node/189.

**Appendix 3. Additional Resources for Clinical Observation**

Articles & Presentations

* Encandela J, Konopasek L, Rao D. *What You See is What You Get: Conducting Direct Clinical Observations*. 2008. Available at: http://sklad.cumc.columbia.edu/acgme/toolbox/toolbox59/Direct%2520Clinical%2520Observation.ppt&ei=w2x8UIqBKcWzrAe4n4H4Dw&usg=AFQjCNHOtmH469bfi9YIeQg7Qs19sg3LRQ&sig2=BUTeIu-bN2Ca\_Ddyw-mnYg
* Hauer KE, Holmboe ES, Kogan JR. Twelve tips for implementing tools for direct observation of medical trainees’ clinical skills during patient encounters. *Medical Teacher* 2011; 33: 27–33. Available at: http://informahealthcare.com/doi/full/10.3109/0142159X.2010.507710
* Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of

clinical skills of medical trainees: a systematic review. *JAMA*. 2009 Sep

23;302(12):1316-26. Review.

* Jibson MD, Broquet KE, Anzia JM, Beresin EV, Hunt JI, Kaye D, Rao NR, Rostain

AL, Sexson SB, Summers RF. Clinical Skills Verification in General Psychiatry:

Recommendations of the ABPN Task Force on Rater Training. *Acad Psychiatry*. 2012

Sep 1;36(5):363-8.

* Vreeman DJ, McDonald CJ, Huff SM. LOINC® - A Universal Catalog of Individual

Clinical Observations and Uniform Representation of Enumerated Collections. *Int J*

*Funct Inform Personal Med.* 2010;3(4):273-291.

Examples of Clinical Observation Tools

* Structured Clinical Observation (Pediatrics).
Available at: <http://depts.washington.edu/dbpeds/Clinics.html>
* SPA Observation Protocols – Service Provision Assessments. Measure DHS.
Available at: <http://www.measuredhs.com/publications/publication-spaq2-spa-questionnaires.cfm>
* Appendix IV: Assessment of the Quality of Prenatal Care. Health Center Observation Checklist. In: Al-Qutob R, Mawajdeh S, Nawar L, Saidi S, Raad F. *Assessing the Quality of Reproductive Health Services.* 1998. New York, NY: Population Council. Available at: <http://www.popcouncil.org/pdfs/assess.pdf>.
* Mini-CEX. Clinical Observation for Trainees. Direct Observation Assessment Tool. American Board of Internal Medicine. Available at: <http://www.abim.org/program-directors-administrators/assessment-tools/mini-cex.aspx>
1. Hawthorne effect – when a participant being observed changes their behavior simply because s/he being observed or to seem to have better performance than what is normal [↑](#footnote-ref-1)