Nigeria
Quality Improvement and Clinical Governance Initiative (QI-CGI)
Piloting Quality Improvement Packages in Primary Care Centers
Impact Evaluation Concept Note

January 2014

Contents

Acknowledgements

This impact evaluation is a collaboration with the National Primary Healthcare Development Agency, the Saving One Million Lives Initiative, Johns Hopkins University, the World Bank's Africa Region, and the Development Impact Evaluation Unit (DECIE) of the World Bank. The team gratefully acknowledges the invaluable support of Dr. Muhammad Ali Pate, former Honorable Minister of State for Health, Dr. Ado J.G, Muhammad, Executive Director of the National Primary Health Care Development Agency, Marie Francoise Marie-Nelly (World Bank Country Director for Nigeria), and Dr Kelechi Ohiri, Program Coordinator, Saving One Million Lives Initiative.

We are also grateful to our peer reviewers, and the individuals listed in the section on staffing below. This impact evaluation is made possible through generous funding from the Bill and Melinda Gates Foundation.

1. Background

While maternal, neonatal, and child health (MNCH) outcomes are improving in Nigeria, the rate of improvement is not sufficient to meet the MDGs related to child and maternal health. Nigeria's under-5 mortality rate, estimated to be 124 deaths per 1000 live births in 2012 is one of the highest in the World. In fact UNICEF (2012) ranked Nigeria as the country with the 12th highest under 5 mortality rate in the world. The Nigerian Federal Ministry of Health (FMOH) is addressing these challenges by introducing important reforms and is committed to learning which of these are working and worth scaling up. Evidence on the causal impact of past and ongoing quality improvement programs is, however, lacking, and so the scope for using previous experience to reliably guide future policy and program design is limited. Assessing and improving the quality of health care delivery in developing countries has been recognized as a priority by the WHO and other health agencies (WHO 2006; Institute of Medicine, 2001).

In this context, **the Nigerian Federal Ministry of Health proposes to experimentally evaluate variants of a health care management consulting intervention to enhance the quality of health care, especially maternal, newborn, and child health care.** The consulting program studied in this IE aims at improving service quality and patient safety at primary healthcare centers (PHCs) by relaying information to providers and through mentoring and tutelage. Service quality and patient safety are systemic healthcare challenges throughout Nigeria and the broader region. The goal is to impact provider knowledge and behavior, improving effort, and ultimately health outcomes, patient safety and patient satisfaction. This program is implemented jointly by the Ministry and the National Primary Healthcare Development Agency (NPHCDA).

The main argument for focusing on quality is that potentially large improvements in outcomes can be achieved from changes in practices without increasing the amount of resources employed. Some evidence suggests that improvements in high-quality care can be provided even in contexts with more limited resources (Walker et al. 1988; Supratikto et al. 2002) and that improving clinical practices and quality of care can lead to better health faster than other factors associated with health outcomes such as economic growth, education, and technological change (Peabody et al., 2006).

Several on-going projects in both the public and private sector address specific elements of quality improvement. In October 2011, the FMOH convened a two-day consultation with both government and non-government stakeholders as an initial step towards the development of a clinical governance and quality improvement strategy. This led to the establishment of the Task Team on Clinical Governance and Quality Improvement which was charged with developing a framework and strategy for planning, governing and improving the quality of health care delivery in the country.

The motivation for supporting quality improvement in identified primary healthcare centres is multidimensional. The Subsidy Reinvestment and Empowerment Program, Maternal and Child Health component (SURE-P MCH) aims to upgrade primary healthcare facilities and increase usage of MCH services through the conditional cash transfer (CCT) scheme. The SURE-P MCH CCT pilot is likely to increase demand for services in the target facilities. Pre-empting this induced demand, the Quality Improvement and Clinical Governance Initiative seeks to ensure that the quality of care provided at such centres is sufficient to benefit patients and avoid harm. More importantly, quality improvement of primary healthcare centres (PHCs) is part of a comprehensive national quality strategy across primary, secondary, and tertiary care facilities. This involves three strategies:

1) defining minimal care standards,
2) supporting facilities by creating a path toward accreditation,[1] and
3) developing long-term domestic capacity for continuous quality improvement.

The SURE-P MCH program, under the leadership of the (NPHCDA) is developing the instruments and know how to deliver facilities' baseline assessments, quality improvement plans and monitoring and supervision. The Saving One Million Lives initiative is providing innovative technical assistance and program oversight to the SURE-P MCH to ensure evidence-based improvements in quality of care. It is actively engaged in the development of program indicators, training, implementation, and evaluation of program elements.

The healthcare consulting intervention constitutes the core of this proposed impact evaluation. The IE will test the effectiveness of the intervention and compare it to a pure control group (no intervention) and a scaled down version of the intervention to (i) establish a causal link between the

[1] See the accreditation guidelines: http://www.safe-care.org/index.php?page=accreditation

program and relevant outcome indicators, and (ii) understand whether a lower cost alternative could be a viable option for an eventual scale up.

The FMOH and NPHCDA in collaboration with the World Bank Group and the Bill and Melinda Gates Foundation, has also initiated a quality performance benchmarking exercise, which involves the assessment and benchmarking of quality of service delivery indicators across primary, secondary and tertiary facilities nationwide, using the Service Delivery Indicator (SDI) tool to enable comparability within country as well as across countries. The SDI instrument will serve as one of the baselines for this project.

2. Summary

The Nigerian Government has prioritized improving the quality of healthcare delivery throughout its care facilities. There are multiple facets to implementing successful quality improvement processes, including providing a transparent system with quantifiable outcome measures and ensuring workforce engagement for healthcare providers.

The SURE-P MCH is contracting a healthcare management consulting firm to provide support to facilities to meet international health care standards. This will include the following activities in 48 PHCs in 6 states as a pilot:

- Conduct Baseline Assessments and Gap Analyses in four key areas – health care organization management, patient care, specialized services and ancillary services (details are provided in Section 3);
- Introduce "Quality Improvement Plans" for each PHC;
- Monitor and provide feedback and support to the PHCs toward implementation of the plans and with the goal of building local capacity in Quality Facilitation and the implementation of the Quality Improvement Plans.

The impact evaluation will feature two treatment arms:

- Treatment A will consist of the "full package" of consulting services, including the initial assessment, action plans, and continuous feedback and support.
- Treatment B is "information only": The consulting firm will conduct the assessment and initial feedback on these indicators will be presented to the PHC workers. Treatment B will not, however, provide hands-on tutelage throughout the quality improvement process.

Treatment A seeks to test the effectiveness of the full consulting program whereas Treatment B measures a lower cost/lower intensity intervention. This comparison will identify whether the main barriers to adopting quality improvement plans are information constraints (PHC staff don't know

what to improve) or implementation constraints (PHC staff know what to improve but don't know how to improve). Dependent on access to relevant data sources, cost-effectiveness analysis will be performed in order to compare outcomes relative to their costs.

The effectiveness will be measured by means of a randomized controlled trial (RCT). The RCT will involve a total of 80 PHCs, located in 20 hospital catchment areas in 6 states. 24 PHCs will be randomly assigned to Treatment A, 24 to Treatment B, and 32 will form the Control Group. The randomization will ensure that the PHCs in the different treatment arms are comparable on average.[2] Stratification on the nearest-hospital clusters will ensure balance on this geographic dimension, which is important because we anticipate that both outcomes and implementation will be affected by local conditions (details on the randomization are presented in Section 5). The only difference between treatment groups is thus the randomly assigned treatment status. This method allows establishing causal effects between the intervention and the outcomes.

Outcomes will be assessed by making use of the following measurement techniques:
(1) Indicators of the PHCs' progress towards the quality standards as defined by the Ministry of Health
(2) Additional measurements of quality of care containing a yet to be defined selection these methods:

a) surveys: facility survey, patient exit interviews, potentially household level data
b) clinical knowledge assessments (e.g. vignettes)
c) direct behavior observations
d) standardized patients

The 48 facilities within this study will serve as pilot cases to evaluate the effectiveness and scalability of this Quality Improvement Program. The IE is highly policy-relevant since it will directly inform the nationwide scale-up decision process.

The Service Delivery Indicator (SDI) tool was used as one of the *baseline surveys* for this IE. The SDI surveys began in June and ended in August of 2013 and covered all SURE-P facilities that are part of the sample. Other data sources include the data that is collected by the firm as well as especially designed follow up data collection instruments that will be employed for this IE.

3. Intervention

As part of the broader NPHCDA plan to partner with private sector and development organizations committed to quality improvement, the NPHCDA/SURE-P MCH has requested a consulting firm (PharmAccess-SafeCare) to support the design and implementation of a Quality Improvement Program for primary healthcare facilities in Nigeria. This Quality Improvement initiative is designed to occur in phases.

- Phase 1: Pilot quality improvement in a set of 48 PHCs.

[2] See Annex 2 for the t-tests that were conducted to check the balance of treatment groups.

- Phase 2: Scale up phase, conditional on the outcomes from Phase 1.

In Phase 1, the healthcare consulting firm will assess quality standards for 48 PHCs in 6 Nigerian states (Treatment A & B). In addition, it will develop detailed action plans for a subset of 24 of these PHCs (Treatment A only). The healthcare management consulting company will also assist PHCs in their quality improvement path. The impact of this program will be measured through predetermined outcome indicators in key areas of interest, as defined by the SURE-P MCH and Saving One Million Lives team (in accordance with the evaluation team).

The program started with the identification of primary health centres which feed a secondary care facility (hub and spoke model) to introduce and prove the concept of service quality and outcome improvement through the health consulting firm's package. Phase 1 was designed in alignment with the Service Delivery Indicator (SDI) tool mentioned earlier. The SDI tool will form the baseline data collection of this proposed impact evaluation.

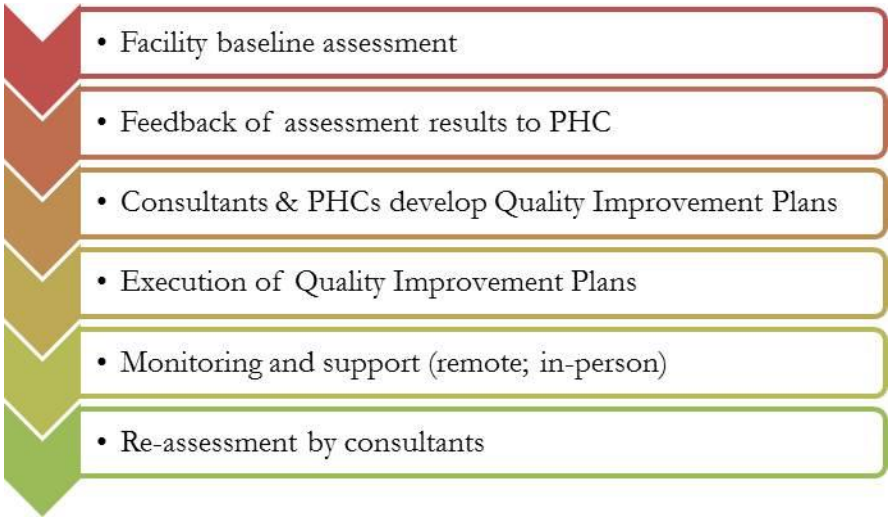
The PHC facilities participating in the evaluation (either as treatment or control) will be the full set of 80 SURE-P primary care facilities (32 control, 48 treatment) in the six initial SDI implementation states (Anambra, Bauchi, Cross River, Ekiti, Kebbi, and Niger). Pending results, this will lead to a scale up strategy and operational plan for 2013-2015, which includes the further institutionalization of Quality Improvement Programs at the FMOH (Phase 2). (Statistical power is discussed later in this note.)

The key goal of the intervention is to help PHC facilities build local capacity in "Quality Facilitation" and for the implementation of "Quality Improvement Plans". To this overall aim, the following activities will be implemented in three stages - "Assessment", "Improvement Planning", and "Monitoring and Support" (see also Figure 1):

- To start out, the healthcare consulting firm will conduct a general training with point persons from the PHCs (1 each). These point people will be the midwives or any other senior staff assigned as the leads for quality improvement in the PHCs. This training will last 2 days in a central location. The attendees will be schooled in standard best practices that apply to all clinics. This training does not yet assess the specific needs of individual PHCs.
- After the training, the healthcare consultants will visit the PHC facilities along with 2 data collectors hired by the SURE-P MCH to conduct Baseline Assessments and Gap Analyses. As part of the assessment, the consultants will assign scores to the PHC on specific indicators in thirteen key "service elements" in four "key areas" (see Table 1 for details). The assessment phase of each facility will last 3 days;
- The results of the assessment will be communicated to the point person at the PHC who distributes them among the staff. The consultants and the PHC will develop a Quality Improvement Plan (for the PHCs in the Treatment A group).
- As the PHC executes the Quality Improvement Plan, the healthcare consulting firm will provide continuous monitoring and support, both remotely and with periodic in-person visits

to the facilities. The "monitoring and support" phase will last 9 months. Specific check-points are 40% and 80% through the improvement cycle. At these points in time, the healthcare consulting firm will collect data on the indicators that were specified as points of improvement in the individual improvement plans.

Figure 1: Structure of the Quality Enhancement Intervention

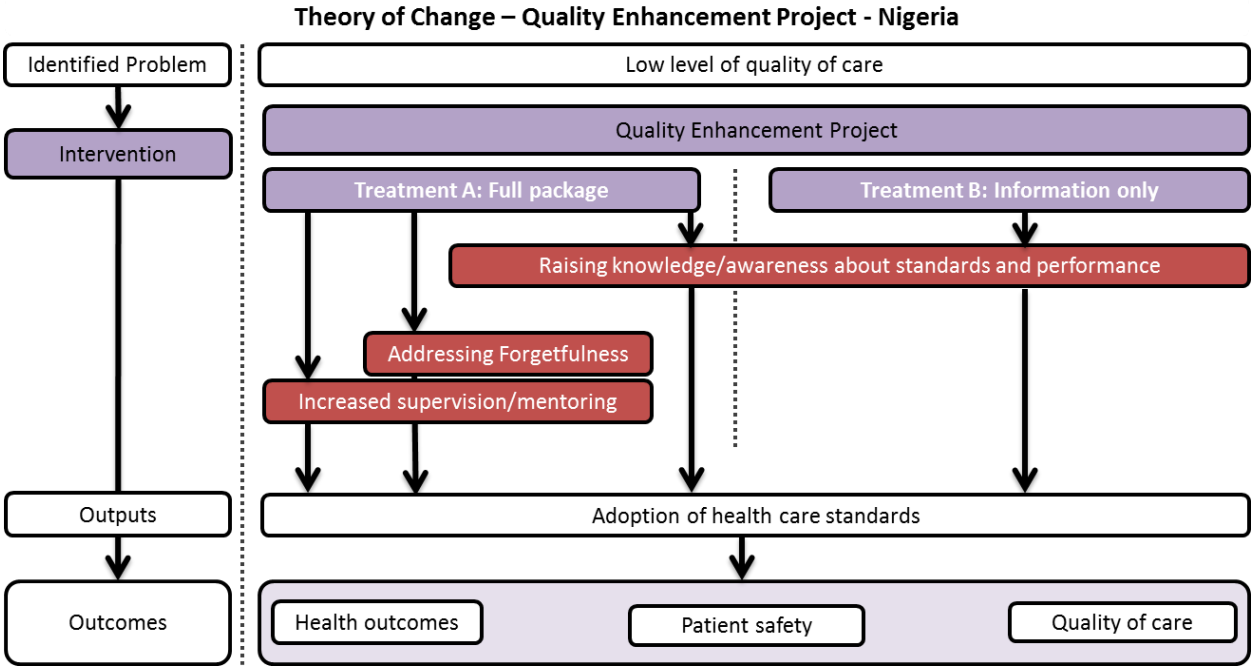


Our hypothesis is that the quality enhancement intervention will affect facility improvement through three possible channels: **information, support/coaching, and incentives.**

1. First, we expect to see an effect through the **informational content** of the initial quality standards assessment and feedback to the clinics. Lack of information on what are the professionally recognized best practices might indeed be a primary reason why such practices are not commonly adopted. (Both Treatments A and B will receive information from the healthcare consulting firm.)
2. Second, the PHC staff might lack the skills to implement the practices. If that is the case, the healthcare consulting firm's **support in designing action plans** to make improvements and **coaching in the implementation phase** might be crucial for improvement in practices to actually occur. (Only Treatment A will receive this implementation support.)
3. Finally, because the quality improvement process will be presented to PHC staff as a step toward accreditation of the facilities, this will **introduce incentives** for managers of health clinics to improve in order to reach the next step on the certification ladder. This may affect both Treatments A and B. While we do not have a separate treatment arm to address this, we can take advantage of non-experimental variation (more and less competition from private clinics, for example) to explore this effect.

More details on the proposed quality improvement approach are provided in Annex 1.

The below figure outlines the theory of change, i.e., the process by which this intervention is expected to improve the quality of care and – subsequently – health outcomes at the relevant centers.



4. Evaluation questions

Assessing and improving the quality of health care delivery in developing countries has been recognized as a priority by the WHO and other health agencies (WHO 2006; Institute of Medicine, 2001). The main argument for focusing on quality is that potentially large improvements in outcomes can be achieved from changes in practices without increasing the amount of resources employed. This applies to wealthy economies, where a wide dispersion in health outcomes remains after controlling for access or spending (Chandra et al. 2013; Skinner 2011), but is especially crucial for low- and middle-income countries. In particular, there is some evidence that suggests that improvements in high-quality care can be provided even in contexts with more limited resources (Walker et al. 1988; Supratikto et al. 2002) and that improving clinical practices and quality of care can lead to better health faster than other factors associated with health outcomes such as economic growth, education, and technological change (Peabody et al., 2006).

Several quality improvement policies have been studied, including legal mandates, accreditation and administrative regulations, professional oversight, national and local guidelines, information sharing, and incentive provision, with mixed results (Peabody et al., 2006). This IE focuses on the role of health care management consulting. Some "change interventions" have been analyzed previously, but they mainly focused on specific processes, individual practitioners, or specific diseases. For instance, Berwick (2004) reports on a successful intervention in Peru aimed at improving tuberculosis care by adopting standard practices such as treatment planning, systematic drug supply management, and

maintenance of registries. Chakraborti et al. (2000) studied the effect of information, feedback and monitoring on private practitioners' case-management skills for treating sick children in rural India, finding large positive effects of the interventions on a number of standard procedures. Tan (1999) describes a program aimed at encouraging adherence to protocols to improve anesthesia safety in Malaysia. Even though these studies report positive results of interventions aimed at improving organizational and individual performance in adopting standards, the assignment of the intervention was often not randomized, which makes the interpretation of the results problematic. Moreover, the interventions typically had multiple components, but the design of the studies did not allow for the effects of the various components to be separately assessed.

This IE will evaluate the effectiveness of a comprehensive health care management intervention in primary care facilities, including organization management, patient care, specialized services and ancillary services. In contrast with the existing literature described above, the randomized-control nature of this evaluation allows us to make causal statements, and the specific design allows us to disentangle the effects of different components of the intervention. They key outcome is the improvement of quality of care in primary health care facilities. The evaluation will focus specifically on maternal, newborn and child health outcomes, analyze service utilization at the population level, and assess quality and quantity of care indicators, measured through administrative data, vignettes, standardized patients, and surveys. This IE will yield the first quantitative evidence on this kind of health care management intervention in a low- or middle-income context.

A World Bank pilot is currently underway in Kenya**,** studying multiple models of health inspections to improve patient safety and the quality of care in Kenyan private and public health facilities. One treatment arm of the Kenya pilot will be very similar to the proposed study in Nigeria. This will provide RCT evidence in at least two settings and significantly add to our cumulative knowledge base.[3] Furthermore, at least key aspects of data collection (e.g., standardized patients) will be coordinated to ensure that comparable data are collected.

Specifically, this IE seeks to answer the following evaluation questions:

<u>Primary research questions:</u>

1. What is the impact of a **full Quality Improvement package** (baseline assessment, gap analysis, quality improvement plan, and mentoring, coaching and implementation support) on (a) Adoption of standards, (b) Quality of care, (c) Health outcomes, and (d) Patient satisfaction?

2. What is the impact of providing facilities with **baseline assessment information only** on (a) Adoption of standards, (b) Quality of care, (c) Health outcomes, and (d) Patient satisfaction? How does it compare to the impact of the full Quality Improvement package?

[3] https://www.wbginvestmentclimate.org/results/upload/Kenya-IEconcept-Jan2013.pdf

This project will further illuminate several related research areas: For example, if certain practices are conducive to better outcomes, why aren't they adopted? Is it because facilities managers and staff are unaware of the beneficial effects of those practices, or is it the case that they are aware but lack the knowledge, effort or ability necessary to implement them (the **"know-do" gap)?** What is the **relative cost effectiveness** of the interventions? We are unaware of cost-effectiveness analyses of quality improvement plans of health care facilities in developing country contexts.

The data collected could potentially provide insights into the question as to whether competition from private health care facilities affect the incentives of public PHC facilities to improve quality standards. (Since treatment PHCs were chosen randomly, the variation in how many private centers are in the vicinity is exogenous.) This could occur because staff (medical and non-medical) employed at public facilities might have "career concerns" and thus be interested in learning and applying the new skills learned from the healthcare consulting firm to obtain a higher-paying job in a private facility. On the other hand, if private facilities have already attracted the best nurses, midwives, etc., then lack of adequate human capital at public PHCs might reduce the effectiveness of the intervention. This is clearly not the principal research question, but it could play a role in explaining the outcomes of the primary research questions.

Outcome indicators

The main categories of outcome indicators for this study are the following:
(a) Adoption of standards,
(b) Quality of care,
(c) Health outcomes, and
(d) Patient satisfaction.

These categories were prioritized by the FMOH and the staff of the Saving One Million Lives in the Nigerian Government. Detailed data collection instruments are currently in developed.

Some specific outcome indicators, as well as some intermediate indicators, which are plausibly linked to improved health outcomes, will be tracked. These include, among others, the following:

- Access (waiting times, ante-natal care coverage, routine immunization coverage
- Patient rating of satisfaction
- Clinical effectiveness (obstetric complication fatality rate, neonatal complication fatality rate, persons receiving ACTs as malaria treatment, number of health facilities that have guidelines for maternal health, family planning, immunization, pneumonia, malaria)
- Efficiency (Days of stock-outs of essential/obstetric medicines, length of patients' stays)
- Patient safety (Neonatal sepsis and neonatal tetanus rates)
- Infrastructure (Days without power, days without access to clean water)
- Staff experience (Percentage of health facilities receiving supervision using national monitoring checklists)

5. Evaluation design: internal validity

Identification strategy

In its first pilot phase, the intervention will be limited to 48 PHCs in 6 states. These PHCs will all come from the universe of SURE-P facilities. In addition, 32 SURE-P PHCs will form the control group for this study.

In order to gather accurate causal estimates of the intervention, this impact evaluation will employ a randomized controlled trial (RCT) design. PHCs will be randomly assigned to either the control group, or one of the two treatment groups, Treatment A or Treatment B.

The **control group** will receive no treatment. They are, however, part of the SDI exercise which also serves as a baseline data source for the study.

PHCs in **Treatment A – "full package" –** will take part in the SDI survey, and will benefit from the full intervention. This includes the following components:

i. A baseline assessment
ii. Comprehensive feedback on the baseline assessment
iii. Detailed quality improvement plans and regular ongoing support provided by the healthcare consulting firm. This includes a gap analysis, quality improvement plan, and mentoring, coaching and implementation support.

The units in the **Treatment B – "information only" –** group will receive components (i) and (ii) but not component (iii). This is testing the effect of providing "information only", i.e., the value of providing the baseline assessment and only the initial feedback.

Progress will be assessed comparing the progress of relevant outcome indicators relative to the progress of the control group.

Treatment A (24 PHCs)	Treatment B (24 PHCs)	Control (32 PHCs)
• Baseline quality assessment • Comprehensive feedback • Detailed plan • Ongoing support (Basic data collection – SDI tool)	• Baseline quality assessment • Comprehensive feedback (Basic data collection – SDI tool)	(Basic data collection – SDI tool)

Data collection methods and instruments

As data sources, the IE will use a combination of PHC administrative data, facility level survey data, the tools developed by the healthcare consulting firm, the SDI and SURE-P surveys, as well as additional instruments to assess the quality of care.

SDI survey

The main baseline data collection will be carried out by the SDI team in Nigeria from June until August of 2013. The SDI tool in Nigeria encompasses 5 parts:

- **Facility questionnaire:** General facility information, infrastructure, availability of equipment, materials, drugs, and supplies
- **Staff roster:** Part A: List of all health workers by cadre type; Part B: Administered to 10 randomly selected health workers to measure absenteeism
- **Clinical knowledge assessment:** Clinical knowledge using 5 medical vignettes + 2 vignettes for maternal & newborn complications
- **Public expenditure module:** Collects receipts and spending (monetary and in-kind) by health facilities
- **Exit module:** User satisfaction, socio-demographic characteristics & payments

The impact evaluation team will have access to the data of all these five modules. Especially valuable is the **facility questionnaire** which provides data on the following indicators which serve as proxies for the quality of care:

- General facility information
- Infrastructure
- Availability of equipment, materials, drugs, and supplies

For the facility questionnaire, the person in charge of the PHC at the time of the visit will function as the respondent. Due to the length and complexity of these surveys, the SDI tool will require a full day at the facility to be completed.

Data gathered by the healthcare consulting firm

The firm will conduct multiple rounds of extensive data collection as part of their work program. This includes a 2-3 day-long baseline assessment and check-ins at the 40% and 80% mark of the assessment cycle. The baseline assessment will be carried out for the 48 PHCs that are part of treatment groups A and B, but not for the control group. The less extensive check-in data collections are being done to track indicators (that were marked as points for improvement) for the Treatment A group.

Although the data do not encompass data for all 80 facilities, the data will inform the IE. The independently commissioned data collection (see below) will gather a selected set of indicators in order to make data comparable.

A data sharing agreement between the healthcare consulting firm, the SURE-P MCH, Saving One Million Lives initiative and the World Bank is currently being discussed. Any data obtained from the firm and/or the SURE-P MCH will only be used for research purposes and not shared with third parties. Any publication will only contain aggregate information on PHCs, and no identifying information on specific PHCs or individuals involved in the study.

Additional (independent) data collection

This IE will make use of at least one round of independently designed data collection. Multiple options are currently under discussion.

As a follow up, we will use a facility survey to record, for example, the absence of drugs or availability of medical equipment as proxies for quality. These measures are necessary but insufficient indicators of quality. They do not consistently predict the accuracy of the advice given by midwives and physicians.[4] In addition to the follow up survey, another preferred option is to visit all the PHCs of the study with *high frequency* in order to a collect a few indicators relevant to quality. This could be in the form of a 5 to 10 question survey administered on a bi-weekly basis to record incidents, absenteeism, drug availability etc. Due to the limited number of PHCs that are part of this study (80), an increase in frequency of data collected would exert positive effects on the power of the study (McKenzie 2012). Patient exit interviews will help assess the quality of care given, and provide subjective measures of patient safety and well-being.

In addition to, we propose to use a combination of state of the art measurement techniques to better assess quality of care. These include the above mentioned (i) facility surveys and patient exit interviews, (ii) vignettes, (iii) direct observations and (iv) standardized patients (SPs). Vignettes are hypothetical cases presented to the doctor, who is invited to proceed as he would with a normal patient. Providers know that they are being tested; this therefore qualifies as a test of their knowledge. Direct observation provides information on what the doctor or midwife actually does. This allows comparing what doctors knew, and what actions they performed. Thirdly, we will employ the SP method, which requires the most effort but delivers the most reliable results. Members of the local community are trained as actors to present a pre-selected case to multiple doctors. After each interaction they are debriefed with a detailed questionnaire. We record time spent per patient, adherence to checklists of appropriate behavior, correct diagnosis and accuracy of prescribed treatment. The standardized patient method is regarded as the "gold standard" in measuring quality of medical care (Das and Hammer, 2014), but it has not been applied in Africa, so this – together with a planned World Bank pilot in Kenya – will add significant value to health research on the continent.

All collected data would be triangulated with the indicators from the healthcare consulting firm. This will allow us to cross-check how well those indicators capture actual quality. The additional data collection allows assessing indicators that are of specific interest to the SURE-P MCH and Saving One Million Lives Initiative.

[4] While lack of inputs can be a problem and should be measured, in most academic studies the correlation between structural inputs and practice-quality is found to be fairly low (Das and Hammer 2014).

The independent data collection will be coordinated by the Saving One Million Lives initiative and the World Bank. The details are currently being worked out. It is important to time this data collection well, in order to reduce "survey fatigue" with respondents at the PHCs. Please see Annex 3 for an overview of possible follow-up data collection options.

Sharing SURE-P data

The related SURE-P impact evaluation is conducting multiple specially designed surveys. These include very detailed household, midwife, and facility surveys. The SURE-P data collection baseline took place in September 2013, a first follow up will be conducted in September 2014. The SURE-P team agreed to share data on the 80 PHCs that form the universe of this study so that synergies can be exploited. This allows us to obtain a richer dataset than relying on the SDI survey alone. The IE will compare these baseline data sources to the quality enhancement indicators that are part of the implementation plan and available administrative data.

Other data sources

- This IE would also make use of randomized case-pulls and other patient-level information (logbooks) as data sources for follow up survey rounds: This will depend on the quality of such records, which will be evaluated in the coming months.
- Regular administrative monitoring data from the PHCs will inform the analysis.

Sampling strategies & power calculations

The sampling frame for this impact evaluation consists of all 80 PHCs in the 6 states that are being covered by the project.

Randomization

Randomization of PHCs into Treatment A, Treatment B, and control followed these steps:

1. We assigned a random number to each of the 80 PHCs in our population.[5]
2. These numbers were ranked in ascending order.
3. We ranked these numbers within each cluster (of 4 PHCs around a referral hospital).
4. The PHC with the highest random number in each was assigned to Treatment A, the second highest number was assigned to Treatment B, and the third highest number was assigned to the control group. This created groups of 20 for each treatment arm.
5. Lastly, the 20 PHCs with the fourth highest numbers were ranked again. Then, the 4 highest numbers were allocated to Treatment A, numbers 5-8 went to Treatment B, and the rest was assigned to the control group. This resulted in the following group sizes:
    - Treatment A: 24
    - Treatment B: 24
    - Control group: 32

[5] As the seed number we chose the starting day of the Uyo Health impact evaluation workshop: 7052013

Baseline balance

Baseline balance checks (t-tests) were conducted to ensure that the randomization yielded the expected results. The results show that the randomization was successful to balance the groups on average. Across 12 indicators, there are no differences significant at the 95% level between Treatment A and Treatment B, and there is only one significant difference between Treatment A and Control and one between Treatment B and Control. That level of imbalance on some variables is expected by chance. The results are listed in Annex 2 of this document.

Power calculations

Reliable power calculations are difficult at this stage due to numerous obstacles:

- The expected effect size and intertemporal correlations between measurements are unknown
- Data on potential outcome variables is limited
- The exact frequency and scope of follow-up data collection efforts is yet to be finalized.

As an example of quality of care, we use the number of children provided with a BCG (tuberculosis) vaccination at birth. These data were made available through regular PHC monitoring data. These data were cleaned of outliers (top and bottom deciles). The following table illustrates initial results of the power calculations. (Although some outcomes will be measures at the individual level and some at the clinic level, the variation is only at the clinic level.)

With the current design we would need to achieve an effect size of approximately 31% with one and two rounds of follow up data collection; in the figure below, green indicates accepted levels of statistical power. Due to the lack of data, these numbers must be understood as an estimated guess.

We expect that the study will be sufficiently powered since an effect size of 30% or more is feasible to achieve given the nature of the intervention and the relatively low levels of quality that provide the base for the improvement process.

That said, the team is taking steps to increase the power through multiple smaller follow-up data collection exercises and – depending on the quality of the logs – using health center logbooks to construct additional pre-baseline observations.

Power Calculations

Version 1: **1 follow-up data collection**	**Effect size in %**			
	11,1	**22,2**	**31,1**	**33,3**
Statistical Power (Ancova Estimation)	0.25	0.617	0.859	0.898
Statistical Power (Estimated Change)	0.227	0.558	0.806	0.851
Notes: Stata code used was sampsi 90 x, sd(60) n(26) r(2) onesided pre(1) post(1) r1(0.7)				
Version 1: **2 follow-up data collections**	**Effect size in %**			
	11,1	**22,2**	**31,1**	**33,3**
Statistical Power (Ancova Estimation)	0.313	0.748	0.944	0.966

Statistical Power (Estimated Change)	0.271	0.664	0.895	0.928
Notes: Stata code used was sampsi 90 x, sd(60) n(26) r(2) onesided pre(1) post(2) r1(0.7)				

Risks and threats to validity

All studies that include intensive survey work have to cope with the so-called Hawthorne effect (Landsberger, 1958). As much as possible, we plan to have identical non-intervention related data collection procedures at all facilities, including the control group: All groups share the same baseline survey, the SDI and SURE-P instruments. This should ensure a balanced (and hopefully minimal) Hawthorne effect across all groups for data collection efforts that are not part of the actual intervention. The intervention itself is, of course, centered around information for results.

We expect **attrition** to be minimal or non-existent. The only way by which PHCs could drop out of the sample is due to closing of the entire facility, which is uncommon.

Spillovers may occur if perceptions of the quality enhancement program, and information about it, spread through word of mouth. In fact, this itself would be a positive sign for the intervention, but would lead to an underestimate in our evaluation if the control group is positively affected by the treatment groups. However, the quality enhancement interventions and action plans are specifically tailored for each participating clinic. Thus, although information spillovers might exist, practical utility of this information is not expected to alter the results of the study. The specific engagement of the healthcare consulting firm is not particularly visible for outsiders. The consultants always speak to the assigned point person and visit facilities only for a couple of days each year. No "physical" changes are introduced by the firm; all changes in appearance are due to the actions and management practices of personnel. Another form of spillover could be crowding in from the private sector, which the team can gauge by asking patients where they were most recently treated for their last illness.

A theoretical possibility – although not often observed in RCTs – is the **John Henry Effect,** i.e., the control group exerting more effort than the treatment group. We would attempt to make records of these by comparing outcomes of control facilities just before and after the roll-out of the program and by including relevant items in the facility surveys to capture this effect.

Ethical issues – IRB plans

We will seek to obtain ethical clearance from Nigeria's National Health Research Ethics Committee as well as well as from the Ethical Review Board of Johns Hopkins University. This will follow the path of the SURE-P impact evaluation which obtained clearance from Nigeria's Ethics Committee as well as from the University College of London earlier this year.

6. Evaluation design: external validity

This IE is directly relevant for future policies of the NPHCDA. If successful, the quality enhancement intervention will likely be scaled up to more Nigerian states and PHCs. A second pilot round, including more PHCs and potentially expanding this IE, is currently under discussion.

The IE attempts to improve external validity by working across 6 different states which represent diverse regions and geographical zones of Nigeria, with differing disease burdens and socio-economic circumstances. This encompassing snapshot will permit some degree of inference regarding the country as a whole, which is crucial for the scale-up decision.

Additionally, this IE includes a costing exercise. This means that not only the effects of the different treatment arms will be examined but also that these effects can be weighed against their respective costs. Only if contrasted with potential costs, IEs can credibly inform policy-decisions. This analysis will consider both budget and actual expenditure.

The cost analysis would include the following:

(i) describing the cost of the results (achieved and expected) of the project, including a detailed cost description of the quality enhancement intervention;
(ii) comparing the per-facility and per-patient cost effectiveness of the different treatment arms of the project;
(iii) exploring alternative options which would produce better cost effectiveness;
(iv) sensitivity analysis, including alternative discount rates and assumptions concerning key cost drivers over time; and
(v) assessing financial and opportunity costs to beneficiary households and to their communities.

Data for costing will be derived from the project's budget (and facility expenditures on quality improvement plans), planning documents as well as from the IE data collection efforts including the SDI surveys and the health consulting firm's assessment tools.

7. Policy relevance & impact

Saving lives by improving quality and demand of health care services is a comprehensive endeavor. Improving the supply side conditions and easing access to care are both equally important. This complete cycle was represented at the recent DIME health IE workshop in Uyo, Akwa Ibom State in May 2013.[6] (DIME is the World Bank's Development Impact Evaluation Group.)

The proposed impact evaluation is highly relevant for Nigeria's health service governance and beyond. The impact evaluation was requested by the SURE-P MCH and was designed jointly by its staff, a

[6] Find the initial design note from the workshop here.

team from the Saving One Million Lives initiative and the external research team (see "Evaluation Team" section). The goal is to inform upcoming policy designs for the scale up of the quality improvement packages to be tested.

The primary audience for the results is therefore the NPHCDA/ SURE-P MCH. However, this IE will have high relevance for other national and international stakeholders. Evidence on quality improvement plans for facilities and accompanying mentoring is still inconclusive and relies only sparingly on rigorous causal inference. Along with the other tested interventions to boost health care performance that are ongoing in Nigeria as part of the Saving One Million Lives Initiative, this evaluation will add to the evidence base on improving quality of care.

The impact evaluation will generate a baseline report, a final policy report, a more technical note, as well as one or more full research paper(s) which constitutes the end of the formal research process.

We seek to disseminate our findings to a wide audience in and outside Nigeria. The results will be presented first to key stakeholders in the Nigerian government. The results will then be disseminated using World Bank distribution channels, including the DIME Newsletter (with a reach of 20,000+), the DIME Seminar series to present findings, and its Facebook page IEKnow.[7] The research paper(s) will be presented in one or more conferences and published in a peer-reviewed journal, and the results will be highlighted in a World Bank blog, all in order to reach the widest audience possible. In general, we will seek to present this work at related Bank and non-Bank events, such as seminars and panel discussions. This IE will also benefit from taking part in the "IE Community of Practice" (CoP), an initiative led by DIME to bring together all health project teams working on IEs in Nigeria. The CoP members will meet several times annually to share ideas and results, exploit synergies, benefit from technical training, and discuss progress and challenges.

The main beneficiaries of this study are the Nigerian Government, the health clinics, and patients, as well as a broader group of impact evaluation, health care and service delivery practitioners. We will inform the NPHCDA/ SURE-P about the results as soon as they emerge to help them build better systems for the PHCs and the patients they seek to serve.

8. Workplan and Timeline

Deliverables and preliminary timeline:

- Baseline data collection – SDI / SURE-P: July/November 2013
- Rollout of implementation/facility assessments: October 2013
- Completed baseline report February/March 2014
- Information on baseline assessment & quality improvement plans February/March 2014

[7] www.facebook.com/ieknowpage

- Conclusion of first round(s) of follow-up data collection — December 2014
- Report of preliminary outcomes — February 2015
- Conclusion of additional rounds of data collection — September 2015
- Dissemination products (Working papers/policy papers) — January 2016
- Potential scale-up, phase II of impact evaluation — April 2016

9. Budget

The implementation of the quality enhancement intervention is fully financed by the Nigerian Federal Ministry of Health. This budget reflects the costs of the proposed impact evaluation activities.

The total budget for the impact evaluation is estimated at $890,364.

Baseline data were already collected through the SDI and SURE-P baseline surveys. Agreements were made to share data with this evaluation at no additional costs, exploiting synergies of multiple ongoing impact evaluations in Nigeria.

In order to additionally support this IE and to leverage the funding from the Gates Foundation, we are actively soliciting other potential external funding sources. In addition, World Bank operating budget will be used to support the program by supporting the time of World Bank staff working on the project.

The scope and analytical depth of this IE, especially in regard to the accuracy of data collection methods that will be employed, will ultimately depend on the amount of funding that can be made available for this activity. Please see below a detailed table of the proposed budget.

Activities / deliverables	FY14	FY15	FY16	Total FY13-FY16
1. Follow-up data collection rounds	70,000	225,000	160,500	455,500
Instruments preparation / pilot	10,000	0	0	10,000
Follow-up data collection (round 1 FY 15, round 2 FY 16)	60,000	225,000	160,500	445,500
Facility survey, patient exit interviews	0	40,000	40,000	80,000
Census of alternative health care options	0	35,000	0	35,000
Standardized patients/Vignettes/Observations	0	40,000	40,000	80,000
Vignettes	0	40,000	40,000	80,000
Direct observations	0	40,000	40,000	80,000
Training, preparation, piloting of protocols, manuals, cases	60,000	30,000	0	90,000
Cost data	0	0	500	500
2. Data documentation	1,000	1,000	1,000	3,000

3. Staff and Consultants	92,000	104,500	79,500	276,000
Project TTL	2,000	2,000	2,000	6,000
Principal Investigator	9,000	9,000	9,000	27,000
Co-Principal Investigator	9,000	9,000	9,000	27,000
Field coordinator	37,500	50,000	25,000	112,500
Research assistance	4,500	4,500	4,500	13,500
Other, specify: ETC / Program coordinator	30,000	30,000	30,000	90,000
4. Travel	28,400	28,400	28,400	85,200
International	**15,000**	**15,000**	**15,000**	45,000
Local	5,000	5,000	5,000	15,000
Hotel + Per diem	8,400	8,400	8,400	25,200
5. Dissemination	2,700	15,700	20,700	39,100
Workshop, meeting	2,000	15,000	20,000	37,000
Other, specify (Seminar, printing)	700	700	700	2,100
6. Other expenditures	21,188	5,188	5,188	31,564
Capacity building / training	**3,000**	**3,000**	**3,000**	9,000
Institutional Review Board/Ethics Committee	**1,000**	**0**	**0**	1,000
Onine server license & maintenance (SurveyCto)	**1,188**	**1,188**	**1,188**	3,564
Other, specify: Android Tablet PC/Phones, chargers	16,000	1,000	1,000	18,000
Total (USD)	**215,288**	**379,788**	**295,288**	**890,364**

10. Evaluation Team

The overall evaluation team consists of the implementation team and the research team with, at times, cross-cutting responsibilities.

Implementation & Internal Research Team

Nnenna Mba-Oduwusi is a consultant with the Saving One Million Lives Initiative. She functions in the capacity of the Quality Improvement Advisor, providing overall leadership and strategic support to the team. Nnenna is a physician with a Master's degree in Public Health from Johns Hopkins University. She has over a decade's experience as a consultant for several government structures and development partners. She worked briefly as a lecturer with the Institute of Primary Health at the College of Medicine, University of Lagos. She was formerly the Senior Reproductive Health Adviser with FHI-360 and Country Manager Insurance with the Pharm Access Foundation.

Ezinne Eze-Ajoku (M.B.B.S) is a Consultant with the Saving One Million Lives Initiative. Here she serves as the program officer for quality improvement. She plays a lead role in the, conceptualization and management of Quality Improvement initiatives, including structuring and coordinating of

training and capacity building for health workers. In her role as the program officer, she coordinates the process for the development and finalization of the National quality of care strategy. She also manages the relationships with multinational and private sector organizations, customizing their best practice approaches to the Nigerian local context

Adeyemi Adeyinka, DDS, MPH, Dr.PH (Performance Program Advisor, Saving One Million Lives). Provides technical support and advisement across programs at Saving One Million Lives. He had worked as Research Associate/Assistant Professor at the Department of Pediatric Dentistry, University of Maryland Dental School and the School of Public Health, Morgan State University, Baltimore, Maryland. He engages in development of methods design for concepts that will facilitate data management components of programs. He is a Healthcare Data Analyst that also has competencies in analyzing data dictionaries and standards documentation to identify issues impacting data quality. He has worked with the WHO/AFRO as its Program Director for Oral Health for Africa in developing system reporting, performance specifications and user manuals.

Chinny Offor is a performance management and private sector advisor with the Saving One Million Lives Initiative. Her primary functions include supporting healthcare businesses to improve their strategy and operations, to expand models of access to capital for the private health sector, and to create an enabling environment for private sector health firms through regulatory and fiscal policy improvements. In her role with Saving One Million Lives, she also works to increase the capacity of providers through physician networks, quality improvement, and knowledge/learning. Chinny is a doctoral candidate at the Harvard Business School, where her research centers on innovate models of healthcare delivery and private entrepreneurship in emerging markets

Oyebanji Filani is a trained physician and health economist with experience in health systems strengthening, health economics and financing. He's a programme officer on health financing with the Saving One Million Lives Initiative. He serves as the government lead on the Resource Tracking study of health care resources in two states in Nigeria. He also works very closely with the Results for Development team, in revalidating the initial Saving One Million Lives cost model in order to create scenarios for a Fiscal Space analysis. As a Senior Project officer with the Quality team, Filani helps with the development of the quality indicator and metrics for the quality scorecard and assessment. He also provides technical assistance in the planning and implementation phase of the quality improvement and clinical governance programme.

Olayiwola Olatawura is a medical doctor currently working as a performance management consultant on the Saving One Million Lives' Initiative. He obtained his degree in Medicine from the University of Southampton, UK in 2007. He has worked in various internal medicine specialties for a number of NHS Hospitals Trusts in Hampshire and Wiltshire. He has also, during the course of his clinical experience, been actively involved in clinical governance within the NHS, particularly in the areas of clinical effectiveness, audit and patient risk management. As a performance management consultant he currently monitors how thirty-six states and the country as a whole is performing using key

performance indicators in areas including maternal and child health, malaria, routine immunization, nutrition, essential medicines and prevention of mother-to-child transmission of HIV.

External Research team

David Evans is a senior economist in the Chief Economist Office of the Africa Region. After completing his PhD in economics at Harvard University and a period as a researcher at the RAND Corporation, he joined the World Bank as Co-Coordinator of the Africa Program of Education Impact Evaluations before joining the LAC region. He has participated in the design, set up, and rigorous evaluation of education projects in the Gambia, Kenya, Sierra Leone, Tanzania, Brazil, and El Salvador. This work has examined a range of education topics, including uniform provision, conditional cash transfers conditioned on children's education, early child education, school grants, school management training, teacher training, and textbook provision.

Mario Macis, PhD (Economics, University of Chicago) is an Assistant Professor of Economics and Management at Johns Hopkins University, Carey Business School. His expertise is in the areas of labor and human resources economics, health economics and experimental economics. He joined the Johns Hopkins Carey Business School in 2010. Prior to joining Carey, he was Assistant Professor of Business Economics and Public Policy at the University of Michigan, Ross School of Business. Mario Macis is also Associate Faculty at the Armstrong Institute for Patient Safety and Quality at the Johns Hopkins University School of Medicine, and Research Fellow at the Institute for the Study of Labor (IZA). Since 2012, he is co-Editor of the Social Science Research Network (SSRN) Labor and Organizations & Markets Abstracts eJournals.

Felipe Alexander Dunsch is a consultant at the Development Research Department's Development Impact Evaluation Unit (DECIE). He is working on health and financial and private sector impact evaluations and also coordinates the outreach activities of DIME.

Oversight and strategic guidance is provided by Dr. Ugo Okoli (SURE-P Project Director), Marie Francoise Marie-Nelly (World Bank Country Director for Nigeria), Arianna Legovini (Manager, DIME/DECIE), Trina Haque (Sector Manager for Health, Nutrition, and Population, West and Central Africa), Benjamin Loevinsohn (Lead Public Health Specialist, Health, Nutrition, and Population, West and Central Africa), Dan Kress (Deputy Director and Chief Economist, Policy Analysis and Financing, Bill & Melinda Gates Foundation), Hong Wang (Senior Program Officer, Bill & Melinda Gates Foundation), and Mara Hansen (Associate Program Officer, Bill & Melinda Gates Foundation).

References

Ashraf, N., Bandiera, O., & Jack, K. (2012). No margin, no mission? A field experiment on incentives for pro-social tasks.

Bateman, C. (2011). Expert trio take on Africa's health care problems: one step at a time. *SAMJ: South African Medical Journal, 101*(5), 290-292.

Berwick, D. 2004. "Lessons from Developing Nations on Improving Health Care."British Medical Journal 328 (7448): 1124–29.

Bruhn, M., & McKenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, *1*(4), 200-232.

Chandra, A., Finkelstein, A., Sacarny, A., & Syverson, C. (2013). *Healthcare Exceptionalism? Productivity and Allocation in the US Healthcare Sector* (No. w19200). National Bureau of Economic Research.

Das, Jishnu and Jeffrey Hammer. 2014. Quality of Primary Care in Low-Income Countries: Facts and Economics. Annual Review of Economics. DOI 10.1146/annurev-economics-080213_041350

Kosfeld, M., & Neckermann, S. (2011). Getting More Work for Nothing? Symbolic Awards and Worker Performance. *American Economic Journal: Microeconomics*, *3*(3), 86-99.

Landsberger, H. A. (1958). Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, *99*(2), 210-221.

Peabody JW, Taguiwalo MM, Robalino DA, et al. Improving the Quality of Care in Developing Countries. In: Jamison DT, Breman JG, Measham AR, et al., editors. Disease Control Priorities in Developing Countries. 2nd edition. Washington (DC): World Bank; 2006. Chapter 70. Available from: http://www.ncbi.nlm.nih.gov/books/NBK11790/

Peabody, J., P. Gertler, and A. Liebowitz. 1998. "The Policy Implications of Better Structure and Process on Birth Outcomes in Jamaica." Health Policy 43 (1): 1–13.

Supratikto, G., Wirth, M. E., Achadi, E., Cohen, S., & Ronsmans, C. (2002). A district-based audit of the causes and circumstances of maternal deaths in South Kalimantan, Indonesia. *Bulletin of the World Health Organization*, *80*(3), 228-235.

Tan, S. K. P. 1999. "Safety with Anaesthesia: The Paradigm of Continuous Improvement."Medical Journal of Malaysia 54 (1): 1–3.

UNICEF, 2012. Under five mortality rankings.

http://www.unicef.org/sowc2012/pdfs/UNDER-FIVE-MORTALITY-RANKINGS.pdf

Walker, G. J., Ashley, D. E., & Hayes, R. J. (1988). The quality of care is related to death rates: hospital inpatient management of infants with acute gastroenteritis in Jamaica. *American Journal of Public Health*, *78*(2), 149-152.

List of acronyms

BCG – Bacillus Calmette–Guérin (Tuberculosis)

CCT – Conditional Cash Transfer

DIME – Development Impact Evaluation (World Bank Unit)

FMOH – Federal Ministry of Health

HMSH – Honourable Minister of State for Health

IE – Impact Evaluation

MCH – Maternal and Child Health

PHC – Primary Healthcare Centers

PSM – Propensity Score Matching

SDI – Service Delivery Indicator

SE – Service element

SURE-P – Subsidy Reinvestment and Empowerment Program

Annex 1: The Quality Enhancement Approach

Formed in 2001 by three partners, PharmAccess, the Council for Health Services Accreditation of South Africa (Cohsasa) and the Joint Commission International (JCI), SafeCare produced internationally acknowledged quality standards specific to resource-constrained public and private healthcare facilities of all kinds.[8] These include tertiary (teaching) hospitals, referral hospitals, district hospitals, primary health centers (as in our case), basic health centers, and health shops/nurse driven clinics. The standards encompass the entire range of clinical services. SafeCare issues certificates for different stepwise achievement levels (1-5), where the level that clinics seek to achieve can be selected by the user/client. This unified set of standards allows benchmarking and comparing facilities. In addition, SafeCare stresses its technical assistant component, with a focus on building local knowledge for the guidance and facilitation of quality improvement measures.

The "Entry Certification Assessment" takes about 2-3 days per facility. Upon this assignment, a first certificate (may) be awarded. Furthermore, SafeCare develops the "Quality Improvement Plan" for the respective clinic. The plan is accompanied by regular visits and trainings.

After 6 months, A SafeCare "assessor" visits the clinic to assess and report on the progress made. Next, after approximately 12 months, a certification assessment takes place, again for 2-3 days. The SafeCare certification committee may issue a certificate of recognition. If level 5 is not yet reached, the cycle can start anew.

Country-based surveyors (hired by the FMOH) and technical assistant partners are mainly responsible for the activities to be carried out under the SafeCare plan.[9] The FMOH foresees to hire 12 data collectors (enumerators) for this task.

The SafeCare standards are grouped in 13 different "Service Elements" (SEs).[10] SEs 1 to 5, and 12 and 13 are "generic" and apply to all types of facilities. The other SEs are, of course, dependent on the type of clinic that is being assessed.

SafeCare standards – Service elements

Health care organization management	1. Management and leadership
	2. Human resource management
	3. Patient rights and Access to care
	4. Management and information
	5. Risk management

[8] The standards were accredited by International Society for Quality in Health Care (ISQua) in March of 2013. http://tinyurl.com/onaw5zg

[9] This information comes from the SafeCare website: http://www.safe-care.org/index.php?page=what-we-offer.

[10] The document is available here: http://www.safe-care.org/index.php?page=safecare-standards

Care of patients	6. Primary health care services 7. In-patient care
Specialized services	8. Operating theatre and anesthetic services 9. Laboratory services 10. Diagnostic imaging services 11. Medication management
Ancillary services	12. Facility management services 13. Support services

Examples of standards

For each of these 13 Service Elements, SafeCare has developed a list of indicators or standards, including very specific practices. We report a few examples below:

6. Primary health care services 6.3 Assessment of Patients
Criteria:

- 6.3.1.1 There is a system, which includes patient identification, for initiating screening at the point of first contact.
- (…)
- 6.3.1.4 Patients who require early attention are identified (e.g. the very frail or ill, or women in an advanced stage of pregnancy).
- 6.3.1.5 There is a system for "fast tracking" patients requiring early attention.
- 6.3.1.6 Waiting times are monitored as part of the organisation's quality management and improvement programme and kept to the minimum.
- (…)

7. In-patient care
7.1 Management of the service

- 7.1.2.1 There is a regular schedule of ward rounds with medical personnel.
- 7.1.2.2 Information exchanged includes a summary of the care provided.
- 7.1.2.3 Information exchanged includes patient resp onse to treatment.
- (…)

11. Medication management11.1 Management of the service
Criteria:

- 11.1.1.1 A designated individual, who is suitably qualified, has clearly defined responsibilities and accountability for all aspects of the pharmaceutical service.
- 11.1.1.2 Individuals who dispense medications act in accordance with legislation affecting pharmacy practice and current pharmaceutical, medical and nursing guidelines.

- 11.1.1.3 The scope of and limitations to the responsibilities and activities of the personnel who manage medications are clearly defined.
- 11.1.1.4 The name of the responsible pharmacist is clearly displayed.
- 11.1.1.5 The pharmaceutical service is coordinated with other related services within the health facility.
- (…)

The standards are generally very specific and can thus be measured and taught to PHC personnel.

Scoring standards

The following section explains the scoring mechanism. It stems from the SafeCare document **"Standards for Clinics /Health Centers in Resource Restricted Settings in Africa."** (p. 7)[11]

Standards are written expectations of structures, processes or performance expectations and it is assumed that if standards are met, better care can be delivered. The standards, in turn, are defined by objective, measurable elements called criteria. Criteria are given weighted value according to how important the criterion is in relation to medico-legal requirements and the impact on safe patient care. This is the "severity rating" and, for the scoring system linked to this document, criteria are rated from 1 (not very serious) to 4 (very serious). During an evaluation visit (survey), criteria are scored either as compliant (C), meaning that the condition is met and that evidence of compliance is present in a tangible and observable form; partially compliant (PC) if the condition required is not totally met but there is positive progress towards compliance and the deficiency does not seriously compromise the standard; or non-compliant (NC) meaning that there is no observable progress towards complying with the required condition.

Scores allocated for each criterion depend on the severity rating for that criterion and whether it is C, PC or NC. Aggregating and averaging criterion scores calculates the level of compliance with the standard.

While progress towards standards that are fully met will bring recognition, it is only when all criteria and standards are substantially met can the organization be accredited.

Compliant criteria are scored as 100. NC or PC criteria are scored as below.

Severity	Partially Compliant (PC)	Non-compliant (NC)
Mild (1)	75	35
Moderate (2)	65	25
Serious (3)	55	15
Very serious (4)	45	5

[11] The document is available on-line here: http://tinyurl.com/l68of3r (August 13, 2013).

A standard may have a criterion that is marked 'critical'. This is where noncompliance will compromise patient or staff safety, or where there are legal implications. Noncompliance with critical criteria is not compatible with accreditation.

These provide a fair, transparent and consistent approach to the scoring of criteria and standards and making decisions regarding recognition and accreditation.

Accreditation ladder

These are the different steps in the accreditation ladder:

Accreditation steps

Accreditation status	Excellent quality systems in place: health care provider has a proven track record of continuous quality improvement, is in substantial compliance with the SafeCare standards, and meets the decision rules for accreditation.
Step 5	Demonstrates long-term commitment to continuous quality improvement, ready for accreditation programme and self sufficiency of continuous quality improvement. Very limited technical assistance required.
Step 4	Strong quality systems in place, but high-risk areas still in need of attention. Limited technical assistance required.
Step 3	Medium quality strength, acceptable but vulnerable to changing environment. Focus on self-evaluation of quality improvement processes using quality indicators, guidelines and standard operating procedures.
Step 2	Modest quality strength, requiring medium technical assistance. Healthcare quality is still likely to fluctuate. Focus on the securing of quality systems, and processes especially in high risk areas.
Step 1	Very modest quality, with continued need for periodic technical support. Focus on implementation of processes and quality systems and the availability of financial means to ensure availability of proper infrastructure and assets.
Letter of Entry to the Graded Recognition Program:	The organization has shown leadership commitment and a strong desire to provide safe health care and recognizes that significant improvements are needed to reach levels of consistent, efficient, safe quality care for each patient. It still has fluctuating quality healthcare provision due to the unavailability of services at times.

Annex 2 – Baseline balance checks

Treatment A vs Treatment B	*# PHCs*	*Mean A*	*Mean B*	*difference*	*p-value*
No of community health workers	46	3.625	3.727	-0.102	0.911
No of Junior com. health workers	47	2.083	1.739	0.344	0.563
Number of delivery instruments	47	3.875	2.435	1.44	0.255
Number of mattresses	48	10.417	7.417	3	0.231
Number of rooms	48	9.208	8.917	0.292	0.856
Number of delivery beds	48	1.917	2.25	-0.333	0.611
Number of hospital beds	48	12.042	10.5	1.542	0.518
Number of doctors	48	0.083	0.25	-0.167	0.127
Number of nurses	48	0.708	0.5	0.208	0.628
Number of nurse-midwives	48	1.083	0.5	0.583	0.246
Number of midwives	48	0.25	0.333	-0.083	0.791
No of medical record officers	42	0.714	0.286	0.429	0.25
Treatment A vs Control	*# PHCs*	*Mean Control*	*Mean A*	*difference*	*p-value*
No of community health workers	54	2.781	3.727	-0.946	0.097
No of Junior com. health workers	55	1.906	1.739	0.167	0.745
Number of delivery instruments	55	2.844	2.435	0.409	0.708
Number of mattresses	56	8.094	7.417	0.677	0.755
Number of rooms	55	9.258	8.917	0.341	0.825
Number of delivery beds	55	1.645	2.25	-0.605	0.227
Number of hospital beds	55	9.742	10.5	-0.758	0.679
Number of doctors	56	0.063	0.25	-0.188	0.048
Number of nurses	56	0.094	0.5	-0.406	0.081
Number of nurse-midwives	56	0.344	0.5	-0.156	0.583
Number of midwives	56	0.125	0.333	-0.208	0.379
No of medical record officers	47	0.308	0.286	0.022	0.911
Treatment B vs Control	*# PHCs*	*Mean Control*	*Mean B*	*difference*	*p-value*
No of community health workers	56	2.781	3.625	-0.844	0.201
No of Junior com. health workers	56	1.906	2.083	-0.177	0.732
Number of delivery instruments	56	2.844	3.875	-1.031	0.403
Number of mattresses	56	8.094	10.417	-2.323	0.388
Number of rooms	55	9.258	9.208	0.05	0.973
Number of delivery beds	55	1.645	1.917	-0.272	0.537
Number of hospital beds	55	9.742	12.042	-2.3	0.334
Number of doctors	56	0.063	0.083	-0.021	0.77
Number of nurses	56	0.094	0.708	-0.615	0.047
Number of nurse-midwives	56	0.344	1.083	-0.74	0.071
Number of midwives	56	0.125	0.25	-0.125	0.49
No of medical record officers	47	0.308	0.714	-0.407	0.22

Annex 3 – Data collection options – Measuring quality of health care

The Quality Enhancement intervention has 15-20 proposed priority indicators, covering the areas of access, patient experience, clinical effectiveness, efficiency, patient safety, health system infrastructure, and staff experience. Some of these indicators (e.g., obstetric major complication case fatality rate) will be captured through existing systems, whereas others (e.g., patient satisfaction) would require an additional survey.

This annex highlights complementary measures of quality, most of which could be collected together with a mini-survey carried out by the SURE-P state facilitator or another, specially trained member of the health team.

A high priority for the Quality Enhancement Impact Evaluation (QE-IE) is making sure that true significant impacts are captured. Including various measures of quality minimizes the probability of reporting no significant impact when in fact the program has improved the quality of care in program health centers. Any decisions about data collection would likewise work to minimize the burden of providing data that falls on health centers.

1. Baseline data

The SURE-P baseline and the SDI indicators will both serve as a baseline survey for the QE-IE. While those surveys may not capture every indicator of quality of interest in the QE-IE, establishing balance between the treatment groups using those indicators makes it easier to argue that differences between treatment and comparison groups in other, additional indicators at endline are due to the quality enhancement intervention. In addition, during follow up data collection, logs at all health centers can be scanned in order to compare whatever data are available in those for the baseline (including the simple fact of whether and in how much detail they are maintained at all), as well as log book data over the course of the evaluation.

2. Structural measures

Facility surveys are useful to record, for example, absence of drugs, availability of medical equipment, record keeping or cleanliness of the facilities as proxies for quality. These measures are useful, cheap to collect, and objective (i.e., there are records or there are not). We propose to collect at least one round of facility level data as a follow-up survey. These could easily be captured using health professionals.

However, these should be seen as a necessary but insufficient condition for health center quality: They do not always provide reliable information on the *accuracy* (quality) of the advice given by midwives and physicians. While lack of inputs can be a problem and should be measured, in most academic studies the correlation between structural inputs and practice-quality is found to be fairly low (Das and Hammer 2014).

In addition to available *structure* (i.e. infrastructure, tools, technology) the ultimate goal of improved health outcomes is preceded by three other (necessary) conditions:

(i) knowledge of health care practitioners

(ii) their actual behavior & exerted effort, and
(iii) other confounding factors (e.g. at the population level).

None of these dimensions can be fully captured by facility surveys; to assess these we require additional measurement techniques. For other, potentially confounding factors, we will rely on the large, detailed SURE-P baseline and follow-up surveys. We will make sure that the facility survey at endline does not capture variables twice that are already included in the SURE-P follow up survey. Making sure that only relevant (i.e., missing) indicators are included will reduce the reporting burden of PHCs, mitigating "survey fatigue".

3. Measuring provider knowledge

Provider knowledge is an essential link in the theory of change towards better health outcomes. The SDI instrument captures knowledge through vignettes and direct observation at baseline.

Clinical knowledge assessments – Vignettes
Appropriate knowledge about diseases and the ways to treat those is essential. Knowledge of health care providers can be best measured by means of vignettes. Vignettes are hypothetical cases presented to the doctor, who is invited to proceed as he would with a normal patient. This can be done with pen and paper, a puppet (to simulate treatment of babies) or ideally individuals simulating real conditions. This technique cannot measure the provider's actual daily behavior, since she knows that she is being observed and potentially exerts more *effort* while being watched. Typically a series of vignettes is presented to each provider, including different conditions, of different diagnostic difficulty and severeness. Results are compared with an existing protocol or the judgment of medical experts, and a composite knowledge score is calculated. With some specialized training, this work could also be carried out by health officials in the state, or by an independently contracted expert who works closely with the state health officials.

Direct observation
Another form of measuring quality is through direct observation of doctor/midwife-patient interactions. An interviewer would sit at the PHC for some time, for example a day, to record the behavior of the provider. Measures include time spent per patient, examinations, treatments assigned, prices charged etc. This procedure provides information on what the doctor or midwife actually does during real interactions with patients. This also allows comparing what doctors knew, and what actions they performed. However, the caveat is that there is a high likelihood that the provider changes her/his behavior due to the fact that a third person is recording their behavior ("Hawthorne-effect"). On the other hand, the provider can only changer his or her behavior to the extent that he or she is able, so this represents the upper bound on their behavior.

The IE would repeat these measures in a follow-up data collection instrument only once, and only if it is confirmed that the SDI will not be used again in 2014).

Patient exit surveys can supplement the approach in order to assess whether treatment varies by education or apparent income. They also provide insights into the subjective experience of patients. Here it is important to note that subjective patient satisfaction can be diametrically opposed to the objective assessment of quality of care. For example, a patient might be highly satisfied with the prescription of a medicine that alleviates short-term symptoms but is detrimental to longer term health outcomes ("overprescription"). Patient exit surveys capturing patient satisfaction could be

conducted simultaneously with the proposed higher frequency data collection (outlined below) in order to realize synergies and mitigate survey fatigue.

4. Measuring actual provider behavior

Standardized patients (Dummy patients)
Both vignettes and the direct observation method share the caveat that doctors and midwives are likely changing their behavior because they know that their actions are being recorded for research purposes. The standardized patient method requires the most effort but delivers the most reliable results. Since doctors and midwives do not know that their actions are being recorded, they behave as in regular conditions. Individuals from the same region are trained as actors to present a pre-selected case to multiple doctors. After each interaction they are debriefed with a detailed questionnaire. We would record time spent per patient, adherence to checklists of appropriate behavior, correct diagnosis and accuracy of prescribed treatment. The standardized patient method is regarded as the "gold standard" in measuring quality of medical care (Das and Hammer 2014), but it has not been applied in Africa. In addition to providing information to the FMOH on the effectiveness of this program, this study – together with a planned pilot in Kenya for early 2014 – will add tremendous value to health research in Nigeria and on the continent.
Obviously, this particular instrument could not be collected by state health inspectors. However, having at least one piece of data gathered by an independent agency could serve as a valuable complement to the data gathered through existing personnel, in terms of the credibility of the evaluation.

The standardized patients would arrive unannounced, but, following ethical human subject research protocols, all PHCs will be informed in advance that standardized patients will arrive at some point in time (informed consents). According to current planning, the standardized patient method will be employed once during the lifespan of the IE (approximately 5 cases per PHC).

5. High Frequency Follow Up Data Collection

To gather certain variables, difficult to capture in any other way, a quick questionnaire could be completed using mobile phones or tablet PCs provided for the evaluation. Some variables that would be unlikely to be captured in log books are (a) staff absenteeism, (b) drug stock-outs, (c) visits from SafeCare, (d) major events (e.g., births), etc. This could consist of weekly, bi-weekly, or monthly record taking at the PHC either by the focal point or an enumerator. Questions from the list of priority indicators can also be included in this higher frequency data collection providing longitudinal insights of a selection of important outcome indicators.

By making use of computer assisted data entry using "Open Data Kit" on tablet PCs or phones, the team would be able to supervize the data collection centrally and monitor the status of health clinics in real-time. These indicators could provide a good picture of what is happening over time in the PHCs while helping to increase statistical precision and power, which can pose a problem for the fairly small sample size of 80 (McKenzie 2012).

6. Conclusion

All of these methods promise to contribute to a rich picture of the quality of health care centers covered by the study and move toward maximizing the chances of capturing the impacts of the SafeCare and the complementary information-only interventions. In addition, most of these methods would potentially add to existing knowledge about the quality of care across all of these clinics, opening the window to a deeper image of the elements of quality than have been measured at scale previously, capturing elements of provider knowledge and effort as well as patient satisfaction.

	Elements of Quality of Care Measured			
Data Methods	Structural measures	Provider knowledge	Provider behavior	Provider effort in real situations
Standardized patients	-	✓	✓	✓
Direct observation	-	✓	✓	-
Vignettes	-	✓	-	-
Facility surveys	✓	-	-	-