


Facilities are substantially more influential than care providers in the quality of delivery care received: a variance decomposition and clustering analysis in Kenya, Malawi and India

Sarah Helfinstein,¹ Mokshada Jain,¹ Banadakoppa Manjappa Ramesh,^{2,3} James Blanchard,⁴ Hannah Kemp,¹ Vikas Gothalwal,^{2,3} Vasanthakumar Namasivayam,^{2,3} Pankaj Kumar,⁵ Sema K Sgaier ^{1,6,7}

To cite: Helfinstein S, Jain M, Ramesh BM, *et al.* Facilities are substantially more influential than care providers in the quality of delivery care received: a variance decomposition and clustering analysis in Kenya, Malawi and India. *BMJ Global Health* 2020;**5**:e002437. doi:10.1136/bmjgh-2020-002437

Handling editor Soumitra S Bhuyan

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjgh-2020-002437>).

Received 26 February 2020
Revised 2 July 2020
Accepted 8 July 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Sema K Sgaier;
semasgaier@surgofoundation.org

ABSTRACT

Introduction Improving the quality of care during childbirth is essential for reducing neonatal and maternal mortality. One barrier to improving quality of care is understanding the appropriate level to target interventions. We examine quality of care data during labour and delivery from multiple countries to assess whether quality varies primarily from nurse to nurse within the same facility, or primarily between facilities.

Methods To assess the relative contributions of nurses and facilities to variance in quality of care, we performed a variance decomposition analysis using a linear mixed effect model on two data sources: (1) the number of vital signs assessed for women in labour from a study of nurse practices in Uttar Pradesh, India; (2) broad-scale indices of respectful and competent care generated from Service Provision Assessments in Kenya and Malawi. We used unsupervised clustering, a data mining technique that groups objects together based on similar characteristics, to identify groups of facilities that displayed distinct patterns of vital signs assessment behaviour.

Results We found 3–10 times more variance in quality of care was explained by the facility where a patient received care than by the nurse who provided it. The unsupervised clustering analysis revealed groups of facilities with highly distinct *patterns* of vital signs assessment, even when overall rates of vital signs assessments were similar (eg, some facilities consistently test fetal heart rate, but not other vitals, others only blood pressure).

Conclusion Facilities within a region can vary substantially in the quality of care they provide to women in labour, but within a facility, nurses tend to provide similar care. This holds true both for care that can be influenced by equipment availability and technical training (eg, vital signs assessment), as well as cultural aspects (eg, respectful care).

INTRODUCTION

Improving quality of care in low-income and middle-income countries (LMICs) has

Key questions

What is already known?

- Globally, the quality of care that patients receive when giving birth is inadequate.
- Half of all maternal deaths and a million newborn deaths each year are due to poor quality healthcare.
- Some specific characteristics of both facilities and healthcare providers have been associated with improved quality of care, but it is unknown whether more of the total variance in delivery care is due to the aggregated impact of all characteristics of the facility, or all characteristics of the care provider.

What are the new findings?

- The quality of care a patient receives is influenced substantially more by the facility in which she receives care than by the provider who provides the care.
- This is true even for aspects of care that don't require specialised training or equipment.

What do the new findings imply?

- Interventions that focus primarily on shifting the behaviour of individual providers may be less effective in improving quality of care than interventions that focus on the culture and characteristics of the facility as a whole.
- The finding that facility effects were comparably large for aspects of care that do not require specialised training or equipment suggests that facility cultural norms may play an important role in shaping the behaviour of care providers.

become a major focus of the global health community to achieve the United Nations' Sustainable Development Goal around good health and well-being.¹ It is estimated that five million deaths in LMICs in 2016 were due to the receipt of poor-quality healthcare.²

While it is unknown precisely how much aid is spent to improve quality of care in LMICs, more than \$5 billion in aid provided to LMICs in 2018 were for the purpose of ‘health systems strengthening,’ which includes efforts to improve quality of care, as well as access, coverage and efficiency.³

Despite these efforts, progress has been mixed, with overall quality of primary and maternal care being quite low in many LMICs.^{4–7} One challenge in improving quality of care has been understanding the most effective targets for intervention. In particular, interventions aimed at changing provider behaviour dominate efforts to improve quality of care; however, these are having modest impacts.^{4 8–10} A recent call to action⁴ has advocated for shifting resources away from ‘micro-level’ interventions focused at changing behaviours at the level of the provider or the facility in favour of an increased focus on ‘meso-’ or ‘macro-’ level interventions at the level of the district or the country. These arguments are predicated on the notion that the impact an individual provider or facility has on quality of care is slight compared to the impact of local or national policies. This is an empirical question that can be tested by observing the variability in quality of care received between districts versus between facilities in the same district, and between facilities versus between nurses in the same facility. If management at the district level is the primary driver of quality of care, one would expect large differences between the care received in one district versus other, with relatively little variability within each district, whereas if individual nurses were the primary driver, we would expect wide variability within each facility depending on the nurse seen, whereas facilities, on average, would perform relatively similarly. Understanding the levels at which most of the variability lies has profound implications for the types of interventions that are most likely to be impactful in shifting quality of care.

There is relatively little empirical research examining variability in quality of care at different levels within the infrastructure hierarchy, and what has been done has focused on relatively specific metrics. For instance, a study conducted in Colombia in 2015¹¹ found that, for women covered by a particular insurance provider, 13% of the variance in whether a delivery was performed vaginally or by caesarean section was attributable to the specific hospital in which the delivery was performed, and another 7% was due to the region where the hospital was located. A recent meta-analysis of interventions to improve healthcare provider performance found that interventions focused primarily on training the providers had small to moderate effects⁸; however, it’s unclear if this is because the specific interventions examined were sub-optimal, and other interventions could be more effective, or rather because provider performance is shaped in large part by characteristics of the facility or the broader policy environment that interventions focused on provider training cannot influence.

Other studies have examined the amount of variance attributable to specific characteristics of the facility or provider. For example, a recent manuscript¹² using some of the same datasets used in the present analyses examined the per cent of variance in quality of care explained by two characteristics of the providers and two characteristics of the facility, as well as specific characteristics of the patient and the region. However, these measured characteristics only explained between 6% and 33% of the total variance in quality of care. This approach did not examine how much of the remaining variance was due to unmeasured characteristics of the facility, the provider or that specific event.

In the present manuscript, we aim to address this gap in the literature by using two independent, large-scale datasets to explore how variability in quality of care in the context of labour and delivery is distributed across care providers, facilities and, where possible, districts and countries. One, a large-scale, recently collected dataset from Uttar Pradesh, India examining collection of vitals, as well as other metrics from the initial clinical assessment for women arriving at facilities to deliver a child, has the advantage of a large sample size and a nested structure of observation with many observations per facility and per nurse, but does not have many broader measures of patient quality of care beyond the assessment of vital signs. To determine whether the findings seen in Uttar Pradesh were similar to those seen in other cultural contexts and using more holistic measures of quality of care, we used the Service Provision Assessment (SPA) datasets, which included highly detailed observations of labour and delivery visits in facilities in Kenya and Malawi, including initial assessments, the labour and delivery themselves, and care during the immediate post-partum period. These datasets allowed us to examine the distribution of variability in quality of care using a more holistic measure, and to separate technical aspects of quality of care that might require access to specialised equipment or training, from respect-based aspects of quality of care, which don’t have such requirements. Finally, based on the finding that a large per cent of the variance in vital signs assessment in all datasets was attributed to characteristics of the facility, we used unsupervised clustering to identify distinct patterns of vital signs assessment in different facilities to see if facilities differ primarily in the amount of testing that is performed, or if they also systematically differ in the types of testing they perform.

METHODS

The study used data from two independent sources to examine the role of nurses and facilities in quality of care. The first was a newly collected large scale data set examining labour and delivery practices in Uttar Pradesh, India, where we examined differences in vital signs assessment practices as a metric of quality of care. We wanted to see if the findings from this dataset held in other parts of the world using broader metrics of quality of care. To this

end, we obtained data on labour and delivery practices from SPA, surveys of health facilities conducted by the Demographic and Health Surveys (DHS) programme.¹³

Study sample: Uttar Pradesh

Data were collected between December 2017 and March 2018 in Uttar Pradesh, India, as part of a larger study examining the quality of care provided to pregnant women at the time of labour and delivery. One hundred and fifteen Community Health Centers (CHCs) and 129 Primary Health Centers (PHCs) were included in the study. Each facility was observed continuously for either 3 or 7 days by teams of 3–4 trained enumerators with nursing degrees. All CHCs and PHCs in Uttar Pradesh were split into quartiles based on the average number of deliveries performed, and a quarter of the facilities in the study sample were randomly selected from the facilities within each quartile. CHCs and PHCs that typically performed fewer than 0.6 deliveries per day were ineligible for participation in the study, and two observed facilities were excluded from further analysis because no women in labour arrived at the facility during the observation period. Across the remaining 242 facilities, 788 nurses were observed examining 4162 women in labour.

Consent to observe the facility was obtained from the facility's Medical Officer in Charge and from the Chief Medical Officer of the relevant district. Every effort was made to observe the course of intake and examination of every woman in labour who arrived at the facility. Data were collected on a wide variety of characteristics about the facilities, nurses, patients and the interaction between the nurses and patients, but the present study focuses specifically on vital signs assessment as a proxy for quality of care as an indicator of whether nurses are collecting the information needed to make informed decisions about patient care. Specifically, enumerators observed the identity of the nurse examining each patient, and whether nurses measured each of the following five vitals during their examination of the patient: blood pressure, fetal heart rate, temperature, pulse and respiratory rate.

Study sample: Kenya and Malawi

In this analysis, we used data about labour and delivery quality of care from SPA surveys in two African countries, Kenya and Malawi. The data were extracted from modules of the survey where trained enumerators, mostly health workers, observe patient labour and delivery visits and answer a standardised set of questions about the care received during the visit. The surveys are nationally representative samples of a nation's health facilities, and within surveyed facilities, patients are selected for observation by systematic random sampling. In Kenya, 83% of deliveries observed took place in hospitals, 11% in health centres, 4% in maternity facilities and 2% in dispensaries. In Malawi, 48% of the deliveries observed were in hospitals, 51% were in health centres and 1% were in clinics. While the surveyors collect observations from all health workers providing delivery care, we included data only from care

providers categorised as nurses and midwives to remain consistent with the data in our Uttar Pradesh sample. In both Kenya and Malawi, greater than 95% of labour and deliveries observed were conducted by care providers categorised as nurses/midwives. Unlike the Uttar Pradesh sample, where all observed deliveries were conducted by nurses, in Kenya and Malawi the sample includes both midwives and nurses (eg, 10% of the overall nurse provider group in Kenya are midwives). Midwives have different training experiences than nurses. We used data from Kenya and Malawi as these were the only countries that used the SPA tools to do systematic observations of routine delivery care. Kenya data was collected in 2010, and Malawi data were collected in 2013/2014. After excluding cases where a patient was seen by a practitioner other than a nurse/midwife or cases where half or fewer of the 46 quality of care items were completed for a patient (11% of the Kenya dataset; 1% of the Malawi dataset), the Kenya SPA dataset consisted of 552 women in labour seen by 312 nurses in 162 facilities, and the Malawi SPA dataset consisted of 452 women in labour seen by 278 nurses in 214 facilities.

Quality of care indices: Kenya and Malawi

To develop an overall metric of quality of care using the SPA data, we used the 46 items directly related to care quality that were assessed in both Kenya and Malawi (see online supplementary appendix p. 1 for full list), similar to an approach used by Macarayan and colleagues to assess the quality of primary care in LMICs.¹⁴ The WHO¹⁵ has envisioned quality of care for pregnant women as comprising 'provision of care,' reflecting evidence-based practices for routine care and management of complications, and 'experience of care,' reflecting respectful care practices such as effective communication, treating patients with respect and dignity, and offering emotional support. These distinct elements of quality of care could be differently related to the facility versus the nurse given that many aspects of provision of care require specialised training and equipment, while experience of care is largely related to the manner in which medical practitioners interact with the patient. Since items assessing both elements of quality of care were included in the SPA assessment, we categorised all items into either a 'Respectful Care Index' (reflecting the experience of care) or a 'Competent Care Index' (reflecting the provision of care), and examined each of these as separate outcome variables, to determine whether the relative impact of facilities versus nurses differed for these two aspects of care. Finally, to directly replicate the findings from Uttar Pradesh, we calculated the number of vitals tested, using the four vitals where data was collected in the SPA dataset (blood pressure, fetal heart rate, pulse and temperature). SPA data were obtained with approval from the DHS Programme, who authorised their use for the proposed analyses.

Statistical analysis

In order to determine the variation associated with the facility and nurse, as well as with district and country where

Table 1 Mixed effect regressions performed for variance decomposition analyses

Sample	Outcome variable	Fixed effects	Random effects
Uttar Pradesh, India	Number of vitals tested	None	District, Surveyor, Facility, Nurse
Kenya and Malawi	Number of vitals tested	Country	Facility, Nurse
Kenya and Malawi	Quality of care score	Country	Facility, Nurse
Kenya and Malawi	Respectful care score	Country	Facility, Nurse
Kenya and Malawi	Competent care score	Country	Facility, Nurse

appropriate, we used variance decomposition techniques based on hierarchical linear models.¹⁶ Specifically, we used a linear mixed effect regression implemented with the lmer package in R.¹⁷ To estimate the total variance explained by each model and the proportion of variance explained by each fixed and random effect, we used the approach described by Nakagawa *et al*,¹⁸ with the proportion of variance explained by each random factor calculated as $\sigma_l^2 / \left(\sigma_f^2 + \sum_{l=1}^u \sigma_l^2 + \sigma_e^2 \right)$, where σ_f^2 is the variance in the fixed effects, σ_l^2 is the variance in the l th random factor of u random factors, and σ_e^2 is the squared residual SD. Similarly, the proportion of variance explained by the fixed effects is $\sigma_f^2 / \left(\sigma_f^2 + \sum_{l=1}^u \sigma_l^2 + \sigma_e^2 \right)$, and the proportion of error/event-level variance is $\sigma_e^2 / \left(\sigma_f^2 + \sum_{l=1}^u \sigma_l^2 + \sigma_e^2 \right)$.

Altogether, five variance decompositions were performed (see table 1 for details). In the Uttar Pradesh sample, we examined the outcome of the number of vitals tested for a particular patient, considering the district and facility where the patient was examined and the nurse examining the patient as possible sources of outcome variance. The survey team observing the patient was also included as a possible nuisance source of variance to account for any systematic biases across surveyors in their coding. In the Kenya and Malawi sample, we examined four different quality of care metrics, detailed above. In all cases, we considered the country and the facility where the labour and delivery took place and the nurse delivering care as possible sources of variance in the outcome metric. 95% CIs for the per cent of variance explained were calculated using 1000 Monte Carlo simulations.

We then performed k-means clustering on the facilities from each sample to determine the patterns in vital signs assessment seen across facilities in each sample.¹⁹ Facilities with only one patient were excluded from the clustering analysis, yielding a sample of 239 facilities for Uttar Pradesh and 140 facilities for the combined Kenya and Malawi dataset. Facilities were clustered based on the percentage of patients for whom they performed each of the recorded vitals, yielding five clustering variables for the Uttar Pradesh dataset and four clustering variables for the Kenya/Malawi dataset. The six-cluster solution

was chosen for both datasets for ease of interpretability, moderate number of clusters and a relatively high silhouette width in both datasets.

Patient and public involvement

Patients or the public were not involved in the design, conduct, reporting or dissemination of our research.

RESULTS

Rates of vital signs assessment

In Uttar Pradesh, vitals were tested relatively infrequently in our sample, ranging from 16.5% of women in labour having their blood pressure tested to 0.7% of women in labour having their respiratory rate tested (see figure 1 and table 2). These results indicate that many patients were receiving no vital signs assessment at all, and virtually none were receiving all the recommended testing.

Nurses in both Kenya and Malawi showed substantially higher rates of testing during the initial examination of patients than in Uttar Pradesh across all vitals measured. In both countries, measurement of fetal heart rate was nearly universal (Malawi: 96.0%; Kenya: 97.4%), and other vitals were tested frequently, but inconsistently, ranging from 42.5% for temperature in Kenya to 73.7% for blood pressure in Kenya.

Quality of care performance

Overall quality of care, as well as the technical and respectful care subcomponents of quality of care were generally similar in Kenya and Malawi, but with a great deal of variability from patient to patient within each country. Both Kenya and Malawi had an overall Quality of Care score of 0.69, indicating the mean patient received 69% of the indicated practices. In Malawi, both technical and respectful care score had the same mean (both means: 0.69), while in Kenya, technical scores averaged 0.72 and respectful care scores averaged 0.56. Scores on all metrics followed a normal distribution, with half of patients receiving a QoC score between 0.61 and 0.78, and all scores ranging from 0.30 to 1.

Decomposition of variance

Based on the variance decomposition of vitals sign testing in Uttar Pradesh, we found that 0.0% of the variance in testing behaviour was explained by the district (95% CI 0.0% to 5.1%), 5.8% (95% CI 0.4% to 11.8%) of the variance was explained by the survey team recording

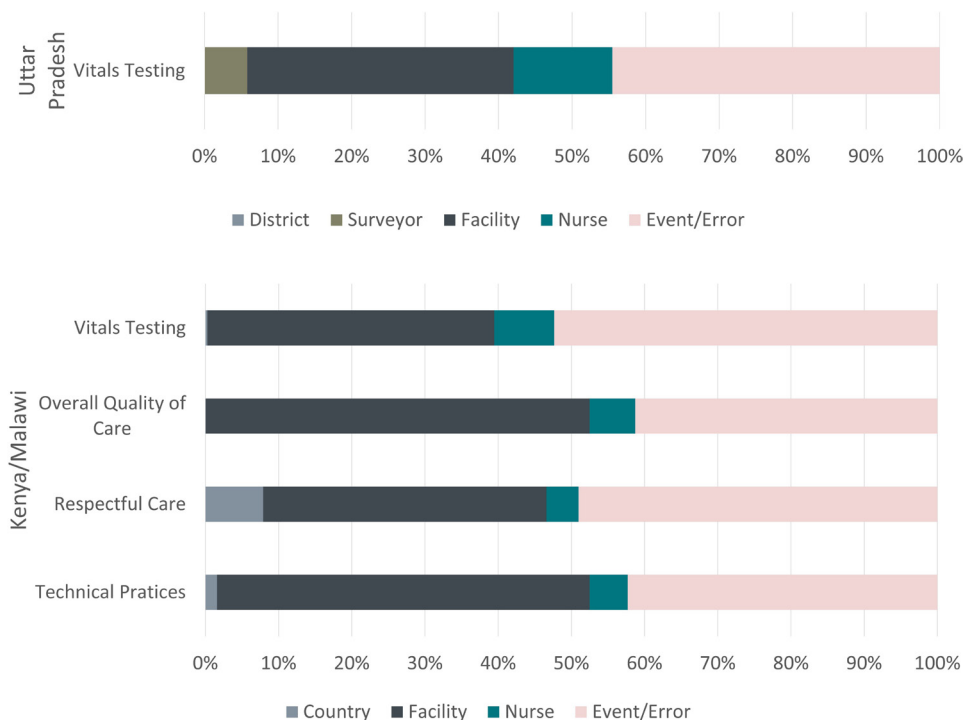


Figure 1 Variance decomposition of quality of care metrics for Uttar Pradesh, India, and Kenya and Malawi. across all measures of quality of care in both locations, the facility and the circumstances of the particular patient explain large percentages of the variance in care seen, while the specific nurse seen within a facility explains relatively little variance.

the results (a form of error variance suggesting some systematic bias in the measurements of the surveyors), 38.5% (29.7%–43.5%) of the variance was explained by the facility, 14.2% (11.4%–18.2%) of the variance was explained by the nurse and 47.3% (42.6%–52.3%) of the variance was explained at the event level or unexplained (ie, the residual variance, which includes differences in testing behaviour due to the circumstances of that particular patient’s examination, as well as measurement error; see [figure 1](#) and [table 1](#)). These results indicate that once variability between facilities is taken into account, little can be explained by individual differences between nurses. In addition, the district within Uttar Pradesh in which an examination is performed accounted for none of the variability in vital signs testing. In short, it appears that the specific facility at which a patient in labour receives care is a critical determinant of the number of vital tests that she receives.

To assess whether our finding on the role of facilities versus nurses in vital signs assessment held up in a different geographic context, we performed the equivalent variance decomposition on the number of vitals assessed for each patient in the Kenya/Malawi dataset. Despite the overall much higher level of vital signs assessment in Kenya and Malawi relative to Uttar Pradesh, the breakdown of variance in testing was quite similar, with 0.3% (95% CI: 0.0% to 1.9%) of variance attributed to the country level, 39.2% (95% CI: 30.3% to 48.1%) of variance attributed to the facility level, 8.2% (95% CI: 0.4% to 16.4%) of variance attributed to the nurse level and 52.4% (44.9%–60.4%) of variance attributed to the event level or unexplained. Note that while little variance was attributed to the country, this is due to the fact that the two countries that were considered in the analysis, Kenya and Malawi, happened to have very similar rates of vital signs assessment (as well as similar rates on other

Table 2 Frequency of vital signs assessment by country and vital

Vital	Uttar Pradesh, India	Kenya	Malawi
Fetal heart rate	11.3% (n=4162)	97.4% (n=425)	96.0% (n=445)
Blood pressure	16.5% (n=4162)	73.7% (n=422)	60.7% (n=445)
Pulse	5.2% (n=4162)	58.8% (n=422)	50.8% (n=445)
Temperature	2.0% (n=4162)	42.5% (n=421)	49.2% (n=445)
Respiratory rate	0.7% (n=4162)	N/A	N/A

Measurement of respiratory rate was assessed in Uttar Pradesh, but not in Kenya or Malawi.
N/A, Not applicable.

metrics of Quality of Care), but substantial variability within each country. Other evidence has clearly shown that there are often large differences between countries in the quality of care provided,¹⁴ suggesting that an analysis with a larger and more diverse group of countries would likely show a great deal more variance attributed to the country.

Next, we explored whether this same pattern held on a more broad-based metric of quality of care, which included 46 items across the initial examination of the woman, labour, delivery and immediate post-partum care. We again found that the facility at which the patient received care explained vastly more variance than the nurse who provided the care, with 52.4% (44.7%–59.9%) of variance attributed to the facility, 6.3% (0.1%–12.7%) of the variance attributed to the nurse, 0.0% (0.0%–1.2%) of variance attributed to the country and 41.2% (34.8%–48.0%) of the variance attributed to the event level or unexplained.

Finally, to better understand the characteristics of the facility that could be responsible for these differences between facilities, we split the overall Quality of Care score into a Technical Score, which included health practices that could require specialised equipment or training, and a Respectful Care Score, which included items that predominantly wouldn't require specialised training or equipment, but could be influenced by the cultural norms within a facility, such as respectfully greeting the patient, and informing her about the results of examinations. We found that for both types of care quality, a large amount of variability explained by the facility and very little explained by the specific nurse seen, although Respectful Care Practices did show more variability between the two countries, and more variance attributed to the event level, than Competent Care (Respectful Care: 38.7% Facility, 4.4% Nurse, 7.9% Country, 49.0% Event/Unexplained; Competent Care: 50.9% Facility, 5.2% Nurse, 1.6% Country, 42.3% Event/Unexplained). A summary of the variance decompositions for all measures is presented in table 3 and figure 1.

Facilities vitals segmentation

Given that the variance decomposition in both Uttar Pradesh and Kenya/Malawi revealed that a large

proportion of the variance in vital signs assessment was due to the facility where the patient was seen, we performed a clustering analysis to determine the different patterns of vital signs assessment seen in facilities in each dataset. In both datasets, we found that facilities varied considerably in how much vital signs assessment they performed, but also that they displayed quite distinct patterns in the particular types of tests which were performed. The majority of facilities in the Uttar Pradesh sample (figure 2) were 'non-testers,' but the remaining 35% of facilities could be split into five categories, including 'occasional BP testers' that tested for little else, or 'Fetal HR only' testers which very reliably tested for fetal heart rate, but nothing else. Particularly interesting was the emergence of clusters, such as the 'Fetal HR Only' and 'BP and Pulse' Testers, who on average had similar rates of vital signs assessment, but virtually no overlap in the type of testing performed. While testing rates were overall much higher in the Kenya/Malawi dataset, with a fifth of the facilities reliably testing every vital measured and virtually all facilities testing fetal heart rate at a minimum, distinct testing patterns still emerged (figure 3), with certain facilities focusing exclusively on fetal heart rate and blood pressure, for example, while others focused exclusively on fetal heart rate and temperature. These findings suggest that facilities differ dramatically from one another not only in the overall level of quality of care, but in the precise types of care that are emphasised and practised.

DISCUSSION

To our knowledge, this is the first study to directly examine and quantify the relative holistic contributions of the facility versus the care provider to the quality of care that a patient receives in a labour and delivery setting. We found that across very different geographical contexts and multiple metrics of quality of care, facilities vary substantially from one another in the quality of care they administer, while individual care providers within the same facility do not differ substantially from one another. This is true both for aspects of quality of care, such as measurement of vital signs, that may be

Table 3 Variance decomposition of Quality of Care metrics for Uttar Pradesh, India, and Kenya and Malawi

Outcome variable	Country/district level	Facility level	Nurse level	Event level (residual variance)
Vital Signs Assessment: Uttar Pradesh, India	0.0% (0.0%–5.1%)	38.5% (29.7%–43.5%)	14.2% (11.4%–18.2%)	47.3% (42.6%–52.3%)
Vital Signs Assessment: KE and MW	0.3% (0.0%–1.9%)	39.2% (30.3%–48.1%)	8.2% (0.4%–16.4%)	52.4% (44.9%–60.4%)
Quality of Care: KE and MW	0.0% (0.0%–1.2%)	52.4% (44.7%–59.9%)	6.3% (0.1%–12.7%)	41.2% (34.8%–48.0%)
Respectful Care: KE and MW	7.9% (0.0%–19.0%)	38.7% (31.2%–49.0%)	4.4% (0.0%–12.0%)	49.0% (41.6%–59.6%)
Competent Care: KE and MW	1.6% (0.0%–5.0%)	50.9% (43.5%–58.8%)	5.2% (0.0%–11.7%)	42.3% (36.1%–49.8%)

For Uttar Pradesh, India, percentages are of remaining variance after excluding variance due to surveyor. 95% CIs around estimates derived from Monte Carlo simulations are in parentheses. KE, Keniya; MW, Malawi.

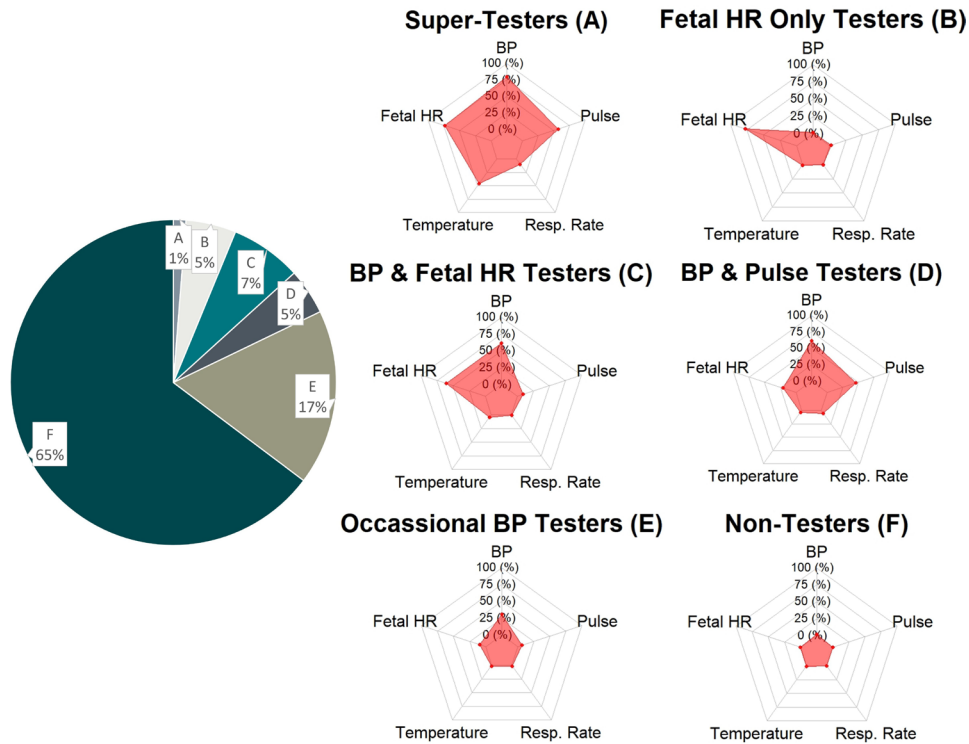


Figure 2 Segmentation of vital signs assessment rates in Uttar Pradesh facilities. Pie chart on left shows the percent of facilities belonging to each segment and the six radar charts on the right show the mean rate of testing of each vital within each facility segment. BP, blood pressure; HR, heart rate.

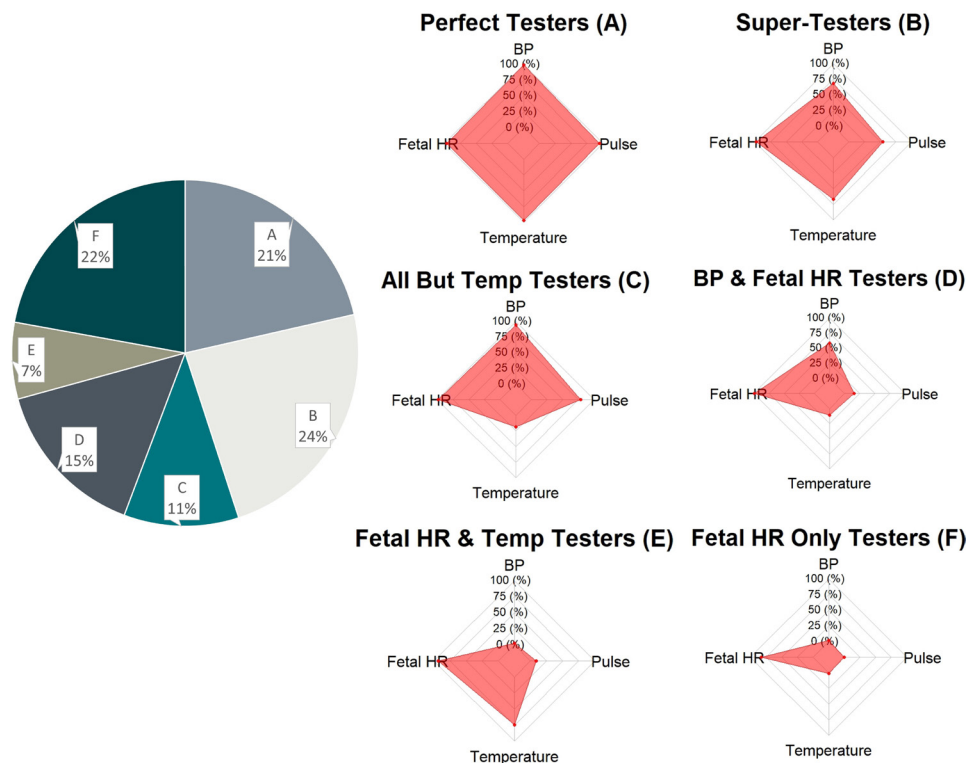


Figure 3 Segmentation of vital signs assessment rates in Kenya/Malawi facilities. Pie chart on left shows the percent of facilities belonging to each segment and the six radar charts on right show the mean rate of testing of each vital within each facility segment. BP, blood pressure; HR, heart rate.

influenced by access to resources or training, but also for respectful care practices, such as greeting patients respectfully and explaining the results of examinations to them, suggesting a key role for the cultural practices and norms established at the facility. We also found that, at least in the realm of vital signs assessment, facilities vary not only in how frequently testing is performed, but also in the specific types of tests that are administered, with each facility seeming to establish their own ‘standard of care’ that is administered fairly consistently by all care providers within the facility.

Together, these findings suggest that focusing interventions primarily on individual nurses, through targeted training and other similar practices, may be an ineffective strategy for influencing quality of care due to the tendency of nurses to conform to the common practices within their facilities. Instead, interventions that are targeted at shifting practices across the facility as a unit will be key. One essential avenue for future research will be understanding the characteristics of the facility that determine the quality of care received there. Given that facilities play as much of a role in determining respectful care practices as they do in other aspects of quality of care, it’s likely that that norms, not just resources, are key. Developing a deeper understanding of how those norms develop and how they can be influenced will be critical next steps.

While this study focuses primarily on the relative impact of facilities versus care providers on the quality of care received, it has recently been argued that ‘meso-’ or ‘macro-’ level interventions emphasising policy change at the level of the district or the country are likely to be more impactful than interventions focused either on the facility or the care provider.⁴ Our dataset in Uttar Pradesh does not seem to support a large role for district-level policy makers, with essentially no variability in vital signs assessment being explained by the district in which the patient seeks care. However, this is based on only one fairly narrow metric of quality of care and could vary in different locales depending on the role district-level government plays in the administration of medical services in that particular locality. This also does not speak to the impact of state-wide policy decisions. There is ample evidence that quality of care varies substantially from state to state within India^{20–22} and from country to country across the world,^{23–25} leaving open the possibility that policy decisions at these “macro-” levels can indeed have large impacts on quality of care. Similarly, while our Kenya/Malawi dataset shows very little impact of country on quality of care, this is a byproduct of the fact that the two countries included in this sample have very similar mean scores on labour and delivery quality of care. This would clearly not be the case if a larger sample of countries had been available for inclusion, and can be seen by comparing vitals scores for our Kenya/Malawi and our Uttar Pradesh sample, where some of the best-performing facilities in Uttar Pradesh were comparable to some of the worst-performing facilities in Kenya/Malawi. Had

vitals scores from both datasets been combined into a single analysis, country would have explained a very large proportion of the variability in vital signs assessment. We believe that using a similar analytic approach with different datasets can be extremely helpful in identifying where within a system the greatest variability in outcomes lies, which can be useful both for generating hypotheses about plausible mechanisms for those differences in outcomes, as well as where within the system targeted interventions are mostly to prove most fruitful.

One strength of the present study was the ability to examine and replicate findings in two independently collected datasets from different parts of the world, and across multiple different metrics of labour and delivery quality of care. However, each dataset had its own weaknesses. The Uttar Pradesh dataset was very large and collected data from a relatively large number of patients per nurse and nurses per facility, an ideal nested structure for an analysis based on hierarchical linear modeling such as this. However, because enumerators observed the initial examination of the patient, but not the labour and delivery itself, there were relatively few quality of care related metrics to produce a robust measure of quality of care. By contrast, the SPA datasets collected in Kenya and Malawi had much richer measurements of quality of care, but a smaller sample size, with fewer nurses observed per facility, and fewer patients observed per nurse. As a result, the CIs on per cent of variability explained in the SPA datasets are fairly large, although still small enough to show conclusively the substantially larger amount of variability explained by the facility relative to the care provider.

Overall, our findings highlight the critical role that facilities play in creating a culture of high-quality care, and that the behaviour of individual care providers often conforms to the broader facility culture. The critical role of facilities is seen cross-culturally, and for aspects of quality of care that do not require specialised training or equipment, suggesting that the facility has a key role in shaping norms, not just in providing needed equipment and training. Moreover, while our findings do not contradict the hypothesis that interventions focused on shifting policy at the level of the state or country can be highly impactful, they suggest that interventions targeted at individual facilities can potentially be extremely fruitful, and that finding ways to close the gap between better- and worse-performing facilities within the same region can produce large improvements in quality of care received. Gaining a better understanding of how cultural norms around quality of care within facilities are created and how they can be shifted will help us find a way to ensure that women globally can receive the high-quality labour and delivery care that they deserve.

Author affiliations

¹Surgo Foundation, Washington, DC, USA

²India Health Action Trust, Lucknow, Uttar Pradesh, India

³Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada

⁴Centre for Global Public Health, University of Manitoba, Winnipeg, Manitoba, Canada

⁵National Health Mission, Government of Uttar Pradesh, Lucknow, Uttar Pradesh, India

⁶Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States

⁷Department of Global Health, University of Washington, Seattle, WA, United States

Twitter Sema K Sgaier @SemaSgaier

Contributors SKS conceived the idea for the study. SKS, MJ, BMR and SH designed the study. SH conducted the analysis. SH, JB, HK and BMR, interpreted results. SH drafted manuscript. All authors reviewed the manuscript.

Funding This study was funded by Surgo Foundation, a non-profit organisation. Employees of the organisation (SH, MJ, HK, SKS) were involved in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval This study was approved by the Sigma India IRB (#10042/IRB/D/17–18).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Service Provision Assessment data are publicly available from the Demographic and Health Survey Program (<http://www.dhsprogram.com>). Data collected in Uttar Pradesh are hosted at the Centre for Global Public Health at the University of Manitoba, and the de-identified data along with the questionnaires and data dictionary will be made available on request (contact e-mail: Ramesh.BanadakoppaManjappa@Umanitoba.ca).

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Sema K Sgaier <http://orcid.org/0000-0002-8311-2686>

REFERENCES

- Kruk ME, Kelley E, Syed SB, *et al*. Measuring quality of health-care services: what is known and where are the gaps? *Bull World Health Organ* 2017;95:389–389A.
- Kruk ME, Gage AD, Joseph NT, *et al*. Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. *Lancet* 2018;392:2203–12.
- (IHME), I.f.H.M.a.E. *Financing global health 2018: countries and programs in transition*. Seattle, WA: IHME, 2019.
- Kruk ME, Gage AD, Arsenault C, *et al*. High-Quality health systems in the sustainable development goals era: time for a revolution. *Lancet Glob Health* 2018;6:e1196–252.
- Owili PO, Muga MA, Mendez BR, *et al*. Quality of care in six sub-Saharan Africa countries: a provider-based study on adherence to who's antenatal care guideline. *Int J Qual Health Care* 2019;31:43–8.
- Das J, Hammer J, Leonard K. The quality of medical advice in low-income countries. *J Econ Perspect* 2008;22:93–114.
- Merali HS, Lipsitz S, Hevelone N, *et al*. Audit-identified avoidable factors in maternal and perinatal deaths in low resource settings: a systematic review. *BMC Pregnancy Childbirth* 2014;14:280.
- Rowe AK, Rowe SY, Peters DH, *et al*. Effectiveness of strategies to improve health-care provider practices in low-income and middle-income countries: a systematic review. *Lancet Glob Health* 2018;6:e1163–75.
- Rowe AK, Labadie G, Jackson D, *et al*. Improving health worker performance: an ongoing challenge for meeting the sustainable development goals. *BMJ* 2018;362:k2813.
- Witter S, Fretheim A, Kessy FL, *et al*. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev* 2012;2:CD007899.
- Vecino-Ortiz AI, Bardey D, Castano-Yepes R. Hospital variation in cesarean delivery: a multilevel analysis. *Value Health Reg Issues* 2015;8:116–21.
- Arsenault C, English M, Gathara D, *et al*. Variation in competent and respectful delivery care in Kenya and Malawi: a retrospective analysis of national facility surveys. *Trop Med Int Health* 2020;25:442–53.
- ICF. The DHS program: spa overview, 2017. Available: <http://dhsprogram.com/What-We-Do/Survey-Types/SPA.cfm>
- Macarayan EK, Gage AD, Doubova SV, *et al*. Assessment of quality of primary care with facility surveys: a descriptive analysis in ten low-income and middle-income countries. *Lancet Glob Health* 2018;6:e1176–85.
- Tunçalp Ö, Were WM, MacLennan C, *et al*. Quality of care for pregnant women and newborns—the who vision. *BJOG* 2015;122:1045–9.
- Raudenbush SWB. *Hierarchical linear models: applications and data analysis methods*. Thousand Oaks, CA: Sage, 2002.
- Bates D, Mächler M, Bolker B, *et al*. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 2015;67:1–48.
- Nakagawa S, Schielzeth H, O'Hara RB. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* 2013;4:133–42.
- Engl E, Smittenaar P, Sgaier SK. Identifying population segments for effective intervention design and targeting using unsupervised machine learning: an end-to-end guide. *Gates Open Res* 2019;3:1503.
- Rani M, Bonu S, Harvey S. Differentials in the quality of antenatal care in India. *Int J Qual Health Care* 2008;20:62–71.
- Powell-Jackson T, Mills A, Acharya A. An assessment of the quality of primary health care in India. *Econ Polit Wkly* 2013;48:53–61.
- Peters DHY, Sharma AS, Ramana R, *et al*. *Better health systems for India's poor: findings, analysis, and options*. Washington, D.C: The World Bank, 2002.
- Kruk ME, Chukwuma A, Mbaruku G, *et al*. Variation in quality of primary-care services in Kenya, Malawi, Namibia, Rwanda, Senegal, Uganda and the United Republic of Tanzania. *Bull World Health Organ* 2017;95:408–18.
- Health at a Glance 2017. *Health at a glance*. Paris: Organisation for Economic Co-operation and Development, 2017.
- Huicho L, Scherpier RW, Nkowane AM, *et al*. How much does quality of child care vary between health workers with differing durations of training? an observational multicountry study. *The Lancet* 2008;372:910–6.