Title: Evaluating Concordance between Government Administrative Data and Externally-Collected Data among High-Volume Government Health Facilities in Uttar Pradesh, India

Lead and corresponding author:

Beth S. Phillips, MPH, Senior Research Analyst, UCSF Institute of Global Health Sciences, 550 16th Street, San Francisco CA, 94158, 415.476.5616, beth.phillips@ucsf.edu

Co-authors:

Shreya Singhal, MPS, Statistician, Community Empowerment Lab, Lucknow, India, shreya.singhal@celworld.org

Shambhavi Mishra, PhD, Associate Professor, Department of Statistics, University of Lucknow, Lucknow, Uttar Pradesh, India, shambhavimishra.lko@gmail.com

Fnu Kajal, M.D.,M.S, India Administrative Officer (IASUP-08), National Health Mission, Lucknow, India, kajal3079@gmail.com

Sun Yu Cotter, MPH, Program Manager, University of California, San Francisco, Institute for Global Health Sciences, sun.cotter@ucsf.edu

May Sudhinaraset, PhD, Professor, University of California, Los Angeles, Jonathan and Karin Fielding School of Public Health / University of California, San Francisco, Institute for Global Health Sciences, msudhinaraset@ucla.edu

Abstract

Opportunities to validate government reports through external audits are rare, notably in India. A recently completed cross-sectional maternal health study by our team in Uttar Pradesh, India's most populous state, involved review and use of government administrative data and externally collected data on health service indicators. This paper presents comparative analyses of the two datasets to test for concordance among the same maternal health quality indicators. The study findings indicate concordance between most indicators across government self-reported and externally collected data. When stratified by facility level or service type, results suggest significant over-reporting in the government administrative data on indicators that are incentivized. This is consistent across all levels of care; however, the most significant disparities appear at higher-level facilities, namely District Hospitals. This study has a number of important programmatic and policy implications. Government administrative data has the potential to be highly critical in informing large-scale quality improvements.

Background

Globally, government health administrative data is a common source for deciding, funding, and evaluating national health programs. Open data platforms, news sources, and social media generally are making government or public sector data more accessible and consumable to the general public [1–4]. Meanwhile those who access government data, such as researchers, politicians, consumers, and lay community members, are increasingly skeptical of these data sources [5,6]. While the quality of government data is often questionable, opportunities to validate government reports through external audits are also rare and typically constrained by timing, political instability and/or budgets [5,7]. This paper examines the results of a cross-sectional maternal health study by our team in Uttar Pradesh (UP), India's most populous state, which involved review and comparisons of government facility health data sources at select study sites.

Data sources typically defined as 'government collected' are those that are made or ordered by a government or government-controlled bodies and reported by government employees or workers, such as hospital administrative data, health department performance data or government health ministry findings. These sources enable public health researchers to explain overall health trends, develop standards for care, measure impact of health policies or programs, to establish health performance benchmarks, or implement evidence-based policies [8–10]. Despite this ubiquitous utility and quantity of government data, limited research examines the reliability of this data. Previous research that examines the quality of government data are few. These studies have focused on examining biases in administrative data [11,12] or to ascertain whether household survey data or government-reported data are more reliable for health systems

strengthening [10,13,14]. For example, a recent study of a large clinical trial in the United States, Women Health Initiative (WHI), investigated the concordance between outcomes included in Medicare claims and those identified in the WHI protocol for cardiovascular events requiring hospitalizations [11]. While the authors posit limitations related to administrative barriers that may impact reporting, they argue that the general agreement found between Medicare claims and WHI data demonstrate an important step towards building evidence-based medicine within existing resources.

In other parts of the world, studies have compared household or community-level survey results to government administrative data and found more questionable concordance results [6,14–16]. A recent study from East Africa examines under-reporting and over-reporting in immunization coverage in Kenya by comparing household vaccination survey results to government administrative data [16]. The authors found that government data tended to misrepresent service use when tied to pay-for-performance initiatives. Specifically, the authors found that government data under reported immunization coverage compared to household surveys. Consequently, donor funds went towards purchasing more vaccines than necessary, which ultimately wasted valuable resources [16].

India offers a unique context to examine the validity of government administrative data because the Government of India (GoI) routinely uses health facility data to inform national benchmarks and service provision guidelines. Of the few studies from India that examine government health data and its validity, all point to a paucity of comprehensive data sources on routinely collected and verifiable health data [14,15]. Morton, et. al. (2016) examine claims data from Rashtriya Swasthya Bima Yojana (RSBY), India's biggest government-sponsored health insurance

scheme, in one district of Orissa state, to determine healthcare cost-effectiveness and government data quality. They conclude, however, that their analyses offer limited applicability for guiding health system strengthening in the state and informing other healthcare settings due to lack of robust benchmarks for clinical quality in India. As a result, the reliability and generalizability of the district's government insurance claim data remains questionable [15]. Similar to another health planning readiness study in peri-urban Kenya [16], a recent readiness assessment to implement Health Technology Assessment across India also finds differences between administrative data and household survey data [14]. Their findings suggest that government reports typically over or under report certain statistics that are tied to performance measures or to financing schemes.

These misrepresentations of health data often produce unintended consequences. Commonly termed "perverse incentives," this concept is cited among experts in health systems and health care policy to explain over reporting of certain procedures or protocols that have higher reimbursable monetary value to the facility and/or the actual health provider [6]. In other words, providers are paid more, or their health facility receives more reimbursements from health insurance schemes, such as government sponsored insurance, when more patients get these 'incentivized' services. These incentives, however, may inadvertently or perversely jeopardize patient well-being and negatively affect health outcomes [6]. Recent reports [17] document high rates of cesarean section (C-section) which is the most common major surgical intervention in the world [18]. C-sections can be life-saving at 10% across a population, and are typically reimbursed at higher amounts than vaginal deliveries [19]. When C-sections are unwarranted, these operations can also lead to increased chance of obstetric and neonatal complications [19]. Sandefur and Glassman (2014) argue that such incentives often "perversely" impact patient well-

being, jeopardize government and funder mutual trust, and limit public expenditure efficiency by exaggerating progress on development or public health indicators [6]. Nearly 15 years ago, the GoI instituted *Janani Suraksha Yojana* (JSY) [20]. This program incentivizes women as well as health providers and facilities with cash transfers to have specific maternal health services done at a government health facility. The 2017 GoI's Record of Proceedings (UP RoP) yearly budget outlines the specific amounts patients and providers receive for institutional vaginal births, C-section births, sterilization, Intrauterine Contraceptive Device (IUCD) and Postpartum IUCD (PPIUCD) insertions [21]. The line item amounts for C-sections and sterilizations are incentivized at nearly three times the amount for vaginal births and less invasive family planning services, such as birth control pills or Depo-Provera. Recent evidence suggests that these incentivized maternal health and family planning services under JSY may inadvertently undermine the Government's aim to lower maternal mortality and morbidities and increase long-acting contraception use in India [22].

It is therefore critical to examine the quality of government self-reported health facility data. This paper reveals key consistencies and highlights striking discrepancies between government administrative health data and externally collected data on maternal health clinical quality by study researchers at the same public health facilities in Uttar Pradesh, India. We hypothesize that government health facilities will over-report on incentivized Maternal Child Health (MCH) procedures, such as institutional deliveries, number of C-sections, sterilizations, IUCD/PPIUCD insertions, and staff numbers while underreport on other maternal health outcome indicators, such as number of essential obstetric medicines and post-natal beds, as compared to externally-collected health facility data.

Methods

Study Background

This ancillary research study is part of the multi-country Strengthening Person-Centered Accessibility, Respect and Quality (SPARQ) project [23,24]. This study, SPARQ Quality-Plus (Q+), aims to understand the drivers of clinical quality and person-centered care quality in high-volume, public maternity facilities in Uttar Pradesh (UP). To create a representative geographic sample of 40 health facilities (including tertiary hospitals, maternity clinics and community health centers), our team first developed a facility assessment survey to collect data on facility-self-reported clinical quality. These facility-self-reported data were used to select a representative sample of 40 health facilities in UP from among all government health facilities reporting 200 deliveries or more every month (n=246, defined as "high volume"). At these 40 study sites, our researchers assessed facility health data via external audits, as well as conducted patient and provider surveys on labor and delivery outcomes, clinical management, and person-centered care (results forthcoming in other study manuscripts).

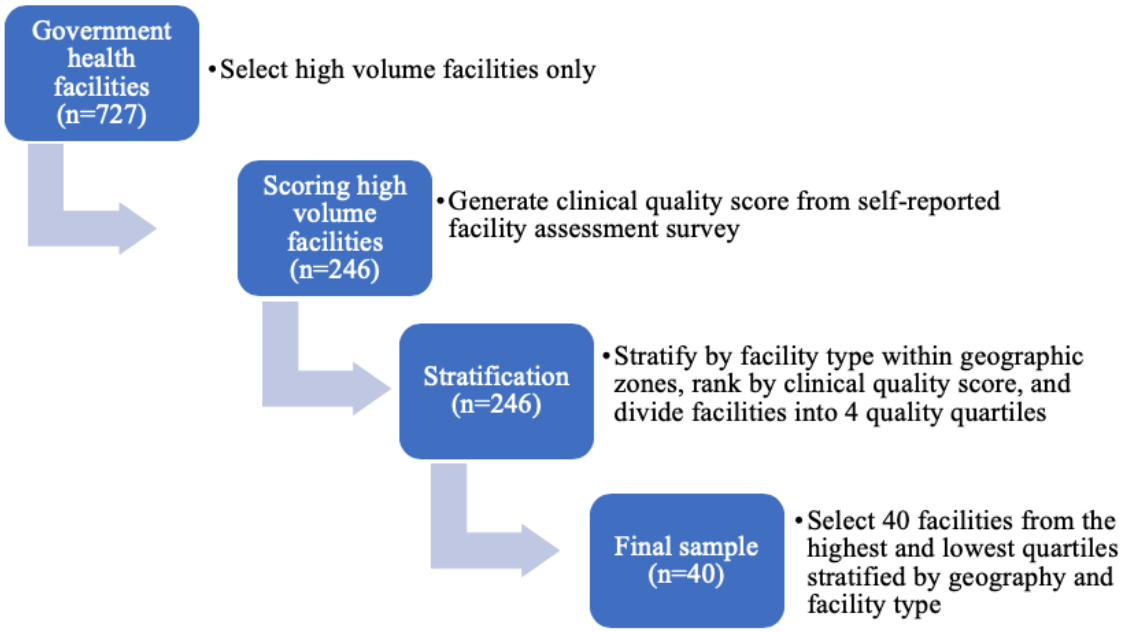
Surveys

We created a Q+ facility assessment survey that combines the specific structural and process quality indicators used and validated by previous researchers to measure maternal health quality [25–27]. This questionnaire focused on key maternal health quality indicators, including facility demographics, available procedures and services, types of patients, medicine and vaccine inventories, sanitation and hygiene, medical equipment, rooms and beds, and human resources. Questions asked for specific numbers or the availability of items/services currently at the facility. If permission was granted by the interviewee, enumerators looked at facility logs or physical

storage units to confirm the availability of these supplies and recorded expiration dates of medication and vaccines. For questions related to number of patients for specific services, such as deliveries, sterilizations, or IUCD/PPIUCD insertions, interviewees were asked to report the total patient number for the previous three months. For example, for surveys administered in February 2017, enumerators asked for the total number of patients who had IUCD/PPIUCDs inserted in January 2017, December 2016, and November 2016. This facility assessment survey was translated from English to Hindi and piloted at a government hospital in Lucknow, India in December 2016. From the pilot, we made minor updates to translations, re-structured certain questions, and re-organized the flow of the survey to reduce redundancies.

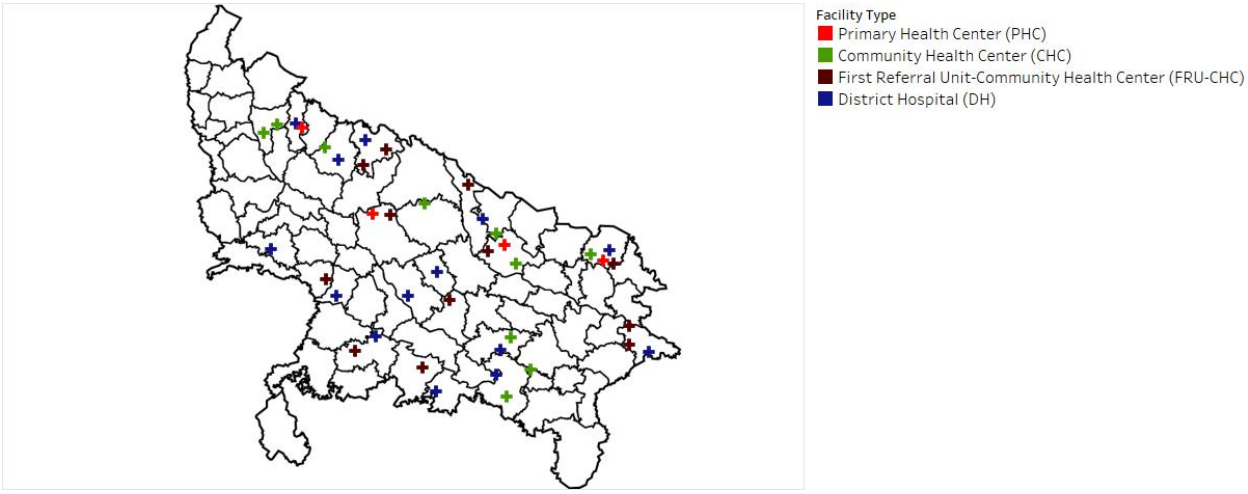
Between January 2017-March 2017, government enumerators, contracted by the National Health Mission (NHM) to conduct research activities, conducted the Q+ facility assessment survey at the ten health facilities with highest patient loads in each of UP's 75 districts (N=750). They interviewed key government facility staff such as Management Officers, Chief Medical Officers, and/or Pharmacist Assistants. The government hired enumerators also conducted direct observations and review of facility records as question instructions indicated. Figure I shows that nearly all facilities (n=727) completed the facility surveys and also outlines the procedures for the 40 Q+ sites selected.

Figure I: Q+ Study Site Selection Process



Following survey completion, our team created clinical quality scores by facility, including structural and process indicators. Structural quality (62 items) included indicators such as number of doctors and nurses and availability of unexpired medicines and equipment. Process quality (15 items) included indicators such as availability of emergency obstetric care services, monitoring labor with partograph, and active management of third stage of labor. All responses were coded as binary, with 1indicating that item was "present" and 0 indicating that the item was "missing." Following analyses conducted by Nesbitt, et al [28] on the Service Provision Assessment (SPA) data, a nationally-representative data source on quality of care in over 30 countries [29, 30], clinical quality indicators were summed to produce a clinical quality index (continuous variable). Using these scores, we ranked facilities according to their clinical quality score. Then, we stratified all by geography and by facility-type to purposefully select a generally representative spread of 40 high-volume government health facilities in UP, as displayed in the study map (Figure II).

Figure II: Map of Q+ Study Sites by Facility Type



During the site selection process and data review, the study team consulted with the local state health ministry and decided to validate these government-collected facility results through an external audit using the same Q+ facility assessment survey.[1] This tool was administered in all 40 Q+ study facilities by locally hired, trained, non-government-employed data collectors, during August 2017-October 2017. The same interviewing and data collection techniques used by the government enumerators were employed, with training conducted by the joint study team. Ethical clearance for this research was provided by the ethics review boards of the authors' respective institutions.

Analyses of Data

We compared facility assessment survey results from the government administrative data collection and the external Q+ data collection to evaluate concordance between the two datasets. Of the 52 questions collected, we selected those most related to clinical quality, infrastructure, and maternal health outcomes. Procedures, such as deliveries and IUCD/ PPIUCD insertions, are

[1] Along with the original domains included in the Q+ facility assessment survey, daily monitoring indicators on beds, availability of food, ambulance, basic sanitation and hygiene services were also assessed during our team's external audit.

reported by average total number performed over the previous three months and grouped under 'Reproductive health procedures and outcomes.' Those questions related to infrastructure, such as staffing, vaccines, or distance to drinking water, are also averaged across all facilities' totals. We termed these items 'Facility Indicators.' Using simple summary statistics, we compared these individual questions by examining the absolute difference and direction of difference between their means. We used t-tests to see if these differences were statistically significant. Following this analysis, we identified and examined outliers at the facility-level through scatter plots on all indicators. Subsequently, we stratified by facility-type to assess the level of concordance within each level of care. Finally, we categorized services by incentivized versus non-incentivized services. We defined incentivized services following the GoI UP RoP 2017, as described previously [21].

We used Cohen's Kappa scores to cross-validate the t-test results and to assess discordance between the individual level categorical government administrative data and the external audited results. The Cohen Kappa test was applied to each categorical variable and checked for concordance. The level of concordance was categorized into five categories based on the values of Cohen's Kappa which are Poor ($\kappa K<0.0$), Slight ($\kappa K=0.0$-0.20), Fair ($\kappa K=0.21$-0.40), Moderate ($\kappa K=0.41$-0.60), Substantial ($\kappa K=0.61$-0.80), and Almost Perfect ($\kappa K=0.81$-1.00).

All analyses were completed using Stata MP 15.0 [31].

Results

From our comparative analyses of government administrative and the external facility datasets, we present the average frequencies and mean differences between the two datasets using both t-tests and the Cohen's Kappa analyses, explore potential confounding, and assess concordance by facility-level of care and incentivized versus non-incentivized services.

Table I displays the government administrative reported means and external audit means of the main Clinical Quality Indicators for maternal health. These average frequencies are summarized across all facility types and organized by Reproductive Health Services and Outcomes or Facility Indicators respectively. Overall, a majority of the indicators appear to match and several differences between the mean frequencies reported by government vs external audit are statistically significant. Specifically, C-section (459.24) and sterilization clients (243.71) mean differences are statistically significant at $p<0.001$ and $p<0.05$ respectively. While facility indicators generally show less dramatic differences on average, government administrative results yield significantly more essential drugs for mothers and babies than our external audit results ($p<0.0001$). Similarly, the government administrative data reports 25 more beds in post-natal ward ($p<0.001$) than the audited data. Staff, especially clinical, are also reported at a higher load ($n=25$) compared to the external audit ($n=17$, $p<0.05$).

Table I: Clinical Quality Indicators*

Indicator Name	Government Administrative Mean Frequency (SD)	External Collected Mean Frequency (SD)	p-value
Reproductive Health Procedures and Outcomes			
Deliveries	973.55 (478.01)	914.03 (461.44)	0.06
C-sections	535.47 (575.79)	76.24 (155.66)	0.00
IUCD/PPIUCD^ insertions	394.75 (905.71)	268.24 (220.08)	0.37

Female Sterilization	287.32 (596.14)	43.61 (65.99)	0.01
Facility Indicators			
Basic medical equipment	7.88 (0.33)	7.93 (0.35)	0.32
Essential Drugs for mothers and infants	23.6 (6.74)	12.88 (4.46)	0.00
Vaccines	4.68 (2.27)	5.53 (1.09)	0.02
Functioning Toilets	9.18 (7.59)	9.18 (3.07)	0.46
Beds in Post Natal Ward	47.26 (42.78)	21.56 (14.76)	0.00
Estimated minutes to walk to safe drinking water	3.65 (6.06)	2.46 (1.91)	0.25
Clinical Staff	25.25 (23.56)	17.1 (8.43)	0.02
Non-Clinical Staff	7.18 (5.50)	6.38 (3.67)	0.36
*Reported over previous 3 months: Government collected: January-March 2017; External collected: May 2017-July 2017 ^ Intrauterine Contraceptive Device (IUCD) and Post-Partum IUCD (PPIUCD) insertions			

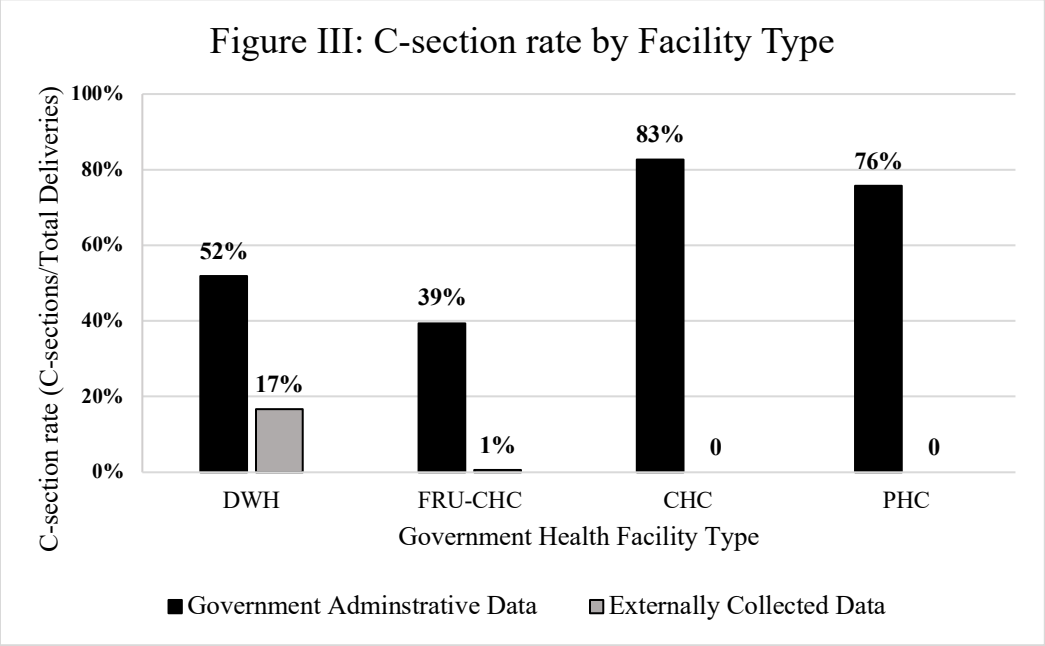
We explored further by stratifying our results by level of care (District Women Hospital (DWH), First Referral Unit-Community Health Center (FRU-CHC), Community Health Center (CHC), and Primary Health Center (PHC) and by service type (incentivized vs non-incentivized). This dual-stratification, as displayed in Table II, reveals that government administrative results tend to over-report services, most notably with incentivized services compared to externally collected results across all facilities. Reviewing the government-reported sterilization loads by facility type, DWHs report 20% more, CHCs and FRU-CHC nearly 9% more, and PHCs over 10% more cases compared to the audited data on sterilization cases. IUCD/PPIUCD insertions show less dramatic differences between government administrative and external audited datasets. Interestingly, both PHC's and CHC's slightly under-report IUCD/PPIUCD insertions as compared to audited results, though negative trend results are not significant.

Table II: Mean differences of Clinical Quality Indicators, Stratified by Level of Care and Incentive Status*

Facility Type	District Women Hospital (n=14)	First Referral Unit-Community Health Center (n=12)	Community Health Center (n=10)	Primary Health Center (n=4)
Incentivized Services: mean difference (p-value)				
C-sections	505.31 (0.01)	304.82 (0.01)	570.40 (0.04)	456.25 (0.32)
IUCD/ PPIUCD insertions	22.43 (0.74)	426.58 (0.37)	-34.70 (0.36)	-6.25 (0.54)
Female Sterilization	374.43 (0.17)	179.50 (0.00)	132.00 (0.03)	202.25 (0.06)
Non-incentivized Services: mean difference (p-value)				
Monthly Deliveries	81.00 (0.27)	29.08 (0.44)	82.56 (0.08)	12.00 (0.91)
Essential Drugs for mothers and infants	11.07 (0.00	9.00 (0.00)	11.60 (0.00)	12.50 (0.02)
Beds in Post Natal Ward	50.77 (0.00)	13.75 (0.00)	13.10 (0.01)	11.50 (0.23)
Clinical Staff	10.71 (0.22)	5.92 (0.21)	9.00 (0.11)	3.75 (0.35)
*Reported over previous 3 months: Government collected: Nov 2016-Feb 2017; External collected: May 2017-July 2017				

Additional discrepancies exist between maternal health indicators found in the government administrative data versus external audit data. For example, C-sections are heavily incentivized in India and we find over 55% of deliveries are C-sections in the government reported data compared to a C-section rate of 8% in audited data (Figure III). Stratification of C-sections results by facility level of care reveals much higher C-sections across all facility types in the government administrative data versus the audited data. Specifically, according to the external audit findings, no C-sections occur at lower-level facilities (PHC and CHCs) which is expected since they are not equipped for these procedures, despite government data reporting C-sections. However, our stratified analyses also reveal that the FRU-CHCs where laboring women are first sent after visiting their local PHC or CHC, are also performing very few C-sections according to

the external audit (less than 5 C-sections reported over 3 months across all 12 FRU-CHC facilities). This finding dramatically contradicts the government administrative reported data. The audited data by level of care shows that only DWHs perform C-sections, though also much fewer (n=505.31, p<0.001) than reported in the government administrative data.



Stratified results for non-incentivized services also reveal differences by level of care between the two datasets (Table II). However, the discordance between government data and external audit data is much less compared to incentivized services, and especially minor among the monthly delivery loads. However, slightly larger differences are suggested among key infrastructure, medicine, and staffing indicators. For example, beds in post-natal wards and essential drugs for mothers and infants appear over-reported across all facilities. The largest discrepancies are found in the 14 DWHs with over 50 more beds and 10 more staff reported on average than actually found during the external audits (p<0.001). There also seems to be a trend

Table III: Cohen Kappa Results for Categorical Indicators

Indicators	Cohen's Kappa Score	Quality of Agreement
Medicines		
Sodium Chloride	0.13	slight
Calcium Gluconate	0.28	fair
Gentamicin for mothers	0.30	fair
Metronidazole	0.26	fair
Misoprostol	0.19	slight
Azithromycin	0.26	fair
Nifedipine	0.30	fair
Zinc Tablets	0.35	fair
Domperidone	0.13	slight
Vaccines		
Measles vaccine	0.30	fair
Pentavalent vaccine	0.30	fair
Staffing		
Specialist doctors in facility	0.21	fair
Pathologist in facility	0.25	fair

* only significant results presented (p -value<0.05)

that clinical staff numbers are over-reported in PHCs, CHCs, FRU-CHCs, and DWHs by the government data sources.

To further assess concordance between the government reported and externally audited health facility data we performed the Cohen Kappa test on all categorical individual indicators. Table III presents only those agreement results that were significant by individual item. Concordance ranged from “slight” to “fair” on items related to availability of essential medicines and vaccines for mother while those scores for items, such as specialist doctor and pathologist available in facility, all show fair agreement levels. These results cross-validate the t-test overall scores presented above and further support the finding that the non-incentivized services show strong concordance across government reported health data and externally audited facility health data in UP.

Discussion

While this study found high concordance between most maternal health indicators across government administrative data and externally collected data, the results suggest significant over-reporting by government sources on indicators that are incentivized at the facility-level or provider-level. In line with our original hypotheses, C-sections and sterilizations were particularly over-reported in government administrative data. Potential explanations for differences in reporting include the structure of India's national health system that incentivizes health providers for certain types of procedures or outcomes, such as assisted deliveries, family planning services, and completed female sterilizations [21]. Incentives that are tied to performance whether in economics, education or health, are typically associated with little or no quality improvement and often with inaccurate data [32-34]. It is also important to note that larger discrepancies occurred at higher-level facilities – namely district hospitals – compared to other facilities. Potential reasons for this could include higher funding needs for district hospitals given the higher client volume. It is also possible that district hospitals may have less capacity for thorough record keeping compared to smaller hospitals, and therefore, results in differential reporting by government reporting and external auditing.

Governments around the world are increasingly driven by consumers and policy-makers to make government data publicly available, accessible, and reliable [2–4,35]. Identifying and understanding the gaps and limitations of government health data contribute to the GoI's objective towards health systems strengthening and making verifiable, routinely collected health

data the gold standard to achieve GoI's goal towards universal health coverage in achievement of Sustainable Development Goals [36]. India increasingly relies on its own internally generated funding and data collection sources to gather health data (moving away from reliance on large scale donors or research institutes). Additional measures to safeguard data quality are essential [14]. Such safeguards could also improve overall health systems operations, patient wellbeing and health outcomes, and public confidence in government health data, by avoiding perverse incentives where certain medical procedures are tied to higher reimbursable values by government and/or private health insurance programs.

The proper use of validated public health facility data presents promising avenues to evaluate healthcare efficiency and effectiveness and to provide updated health information to policymakers in a cost-effective and publicly accessible and credible manner [13]. Even basic steps can be made to enhance India's RSBY insurance scheme's ability to track quality of care, uncover low quality and learn from high quality, and thereby improve overall facility-based provision of health care in India [15]. The GoI is actively working to address issues around "perverse incentives" by applying a multi-strategic approach that involves government-sponsored health insurance schemes. In September 2018, for example, the GoI instituted a national health insurance protection program in UP entitled "Ayushman Bharat Yojana" (ABY) to provide health insurance to 500 million Indians with an ancillary government entity, named NITI-Aayog, to perform external audits bimonthly [36]. ABY provides coverage up to 50,000 INR rupees (roughly USD 7000) per poor family per year for secondary and tertiary care hospitalization. To date, more than 14 million beneficiaries have already been admitted and received treatment under this scheme [37].

In terms of limitations of this study, the sample size is relatively small. While the facility-reported data was collected across 727 facilities (as described in Figure 1), due to resource constraints and the parent study design, only 40 sites completed a Q+ facility assessment survey by an external data collection team. To ameliorate this potential limitation, the 40 study sites were chosen to represent geographic variation as well as level of care across the state. We examined how these 40 sites compared to the remaining high-volume facilities not included (n=206) and found no significant differences in the facilities that were and were not included in our study. Second, the different timing of questionnaire could change responses naturally. This seasonality could bias the discordance results, especially with regards to childbirths, a well-established, worldwide example of seasonality [38,39]. However, as our analysis demonstrated, vaginal delivery loads did not show significant discordance between the government administrative data and the external data collection. Lastly, surveys were administered by different enumerators allowing for potential method administration variability, and respondents were not necessarily the same for each survey, though all survey respondents were facility-based, government hired health providers. Future studies should combine observational data with facility staff reports and interviews conducted by external data collectors.

This study still has a number of important programmatic and policy implications. First, governments, donors and researchers should all be critical of incentivized-reporting when designing studies, determining funding decisions and allocating resources. This study demonstrates that indicators such as C-section rates and female sterilizations are highly misreported, particularly by lower-level facilities. Allocating resources to facilities that have a genuine need for services, medicines, and equipment will lead to higher efficiency and equity in healthcare. Second, programs should set up standardized monitoring systems across all health

facilities in India. This data system would include integrating audited data such as research-collected data, as well as routine administrative data, such as public insurance claims data. Importantly, training for implementation of these data collection systems should be rigorous, including standardized training for data collectors and government officials who are reporting on data. This training should also be tailored towards the level of care and stressing accurate reporting of all health indicators. Quality checks for data monitoring systems, like those common at many research institutions, should also be built into existing government data collection mechanisms, with ongoing quality checks performed.

Conclusion

This study highlights key similarities and striking discrepancies between government administrative health data and externally collected data from the same public health facilities in Uttar Pradesh on maternal health clinical quality. From a health systems research perspective, this study suggests that non-incentivized indicators may have higher validity for broader research questions. With rapid digital advancements changing the global health landscape of how we, whether as funders, politicians or researchers, think about and use government data sources, the importance of verifiably credible government data cannot be overstated. These data have the potential to be highly critical in informing large-scale quality improvements to the healthcare system to ultimately improve the overall health, patient experience, and well-being of women and newborns.

Acknowledgements

We are grateful to all the members of the SPARQ team at the University of California, San Francisco, National Health Mission of Uttar Pradesh (Mission Director, Mr. Pankaj Kumar), and Community Empowerment Lab (Mr. Vinay Pratap Singh). We thank all of our local data collectors and study participants in Uttar Pradesh, India.

Author contributions

BP conceived the manuscript, refined the tools, conducted the data analysis, and led the writing. SS and SM supported data collection and preparation, analysis, and writing. SYC provided editorial review and assisted with data interpretation. MS and FK designed the study, developed the tools, and supported writing of the manuscript. All authors read and approved the final manuscript.

Disclosure statement

All authors declare no conflicts of interest.

Ethics and consent

Ethical review and clearance were provided by the Institutional Review Boards of the University of California, San Francisco in California and the Community Empowerment Lab in Lucknow, Uttar Pradesh, India.

Funding information

This research was funded by the Bill and Melinda Gates Foundation (OPP1127467). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Alexopoulos C, Zuiderwijk A, Charapabidis Y, et al. Designing a Second Generation of Open Data Platforms: Integrating Open Data and Social Media. In: Janssen M, Scholl HJ, Wimmer MA, et al., editors. Electron. Gov. [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014 [cited 2018 Jul 25]. p. 230–241. Available from: http://link.springer.com/10.1007/978-3-662-44426-9_19.

[2] Attard J, Orlandi F, Scerri S, et al. A systematic review of open government data initiatives. Gov. Inf. Q. 2015;32:399–418.

[3] Chattapadhyay S. Access and Use of Government Data by Research and Advocacy Organisations in India: A Survey of (Potential) Open Data Ecosystem. Proc. 8th Int. Conf. Theory Pract. Electron. Gov. [Internet]. New York, NY, USA: ACM; 2014 [cited 2018 Jul 25]. p. 361–364. Available from: http://doi.acm.org/10.1145/2691195.2691262.

[4] Jetzek T, Avital M, Bjorn-Andersen N. The Value of Open Government Data: A Strategic Analysis Framework. 2012;12.

[5] Hut-Mossel L, Welker G, Ahaus K, et al. Understanding how and why audits work: protocol for a realist review of audit programmes to improve hospital care. BMJ Open [Internet]. 2017 [cited 2018 Oct 8];7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5541620/.

[6] Sandefur J, Glassman A. The Political Economy of Bad Data: Evidence from African Survey & Administrative Statistics - Working Paper 373 [Internet]. Cent. Glob. Dev. 2014 [cited 2018 May 18]. Available from: https://www.cgdev.org/publication/political-economy-bad-data-evidence-african-survey-administrative-statistics-working.

[7] Masood A, Lodhi RN. Factors Affecting the Success of Government Audits: A Case Study of Pakistan. Univers. J. Manag. 2015;11.

[8] Chalkidou K, Marquez P, Dhillon PK, et al. Evidence-informed frameworks for cost-effective cancer care and prevention in low, middle, and high-income countries. Lancet Oncol. 2014;15:e119–e131.

[9] Commission on Evidence-Based Policy Making. Commission on Evidence-Based Policymaking Releases Final Report [Internet]. 2017. p. 1. Available from: https://www.cep.gov/news/sept6news.html.

[10] OCED. Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges [Internet]. Paris: OECD Publishing; 2013 [cited 2018 May 15]. Available from: .

[11] Anderson GL, Burns CJ, Larsen J, et al. Use of administrative data to increase the practicality of clinical trials: Insights from the Women's Health Initiative. Clin. Trials Lond. Engl. 2016;13:519–526.

[12] Braham RA, Finch CF. The reliability of team-based primary data collectors for the collection of exposure and protective equipment use data in community sport. Br. J. Sports Med. 2004;38:E15.

[13] Davis P, Milne B, Parker K, et al. Efficiency, effectiveness, equity (E3). Evaluating hospital performance in three dimensions. Health Policy. 2013;112:19–27.

[14] Downey L, Rao N, Guinness L, et al. Identification of publicly available data sources to inform the conduct of Health Technology Assessment in India. F1000Research. 2018;7:245.

[15] Morton M, Nagpal S, Sadanandan R, et al. India's Largest Hospital Insurance Program Faces Challenges In Using Claims Data To Measure Quality. Health Aff. Proj. Hope. 2016;35:1792–1799.

[16] Otieno CF, Kaseje D, Ochieng' BM, et al. Reliability of community health worker collected data for planning and policy in a peri-urban area of Kisumu, Kenya. J. Community Health. 2012;37:48–53.

[17] Lancet T. Stemming the global caesarean section epidemic. The Lancet. 2018;392:1279.

[18] World Health Organization. WHO Statement on Caesarean Section Rates. Human Reproduction Programme (HRP): Research for Impact. Available from: https://apps.who.int/iris/bitstream/handle/10665/161442/WHO_RHR_15.02_eng.pdf;jsessionid=C0795F80C382D0461C13C6532EBD6B4E?sequence=1.

[19] Sandall J, Tribe RM, Avery L, et al. Short-term and long-term effects of caesarean section on the health of women and children. The Lancet. 2018;392:1349–1357.

[20] Fadel SA, Ram U, Morris SK, *et al.* Facility Delivery, Postnatal Care and Neonatal Deaths in India: Nationally-Representative Case-Control Studies. *PLoS One* 2015; **10**: e0140448.

[21] National Health Mission. Record of Proceeding: Uttar Pradesh 2017-2018 [Internet]. Ministry of Health and Welfare, Government of India; 2017 [cited 2018 Sep 4]. Available from: http://nhm.gov.in/nrhm-in-state/state-program-implementation-plans-pips/uttar-pradesh.html.

[22] Singh S, Doyle P, Campbell OM, Mathew M, Murthy GVS. Referrals between Public Sector Health Institutions for Women with Obstetric High Risk, Complications, or Emergencies in India - A Systematic Review. *PLoS One* 2016; 11: e0159793.

[23] Afulani PA, Diamond-Smith N, Golub G, et al. Development of a tool to measure person-centered maternity care in developing settings: validation in a rural and urban Kenyan population. Reprod. Health [Internet]. 2017 [cited 2018 Oct 4];14. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5610540/.

[24] Afulani P, Phillips B, Moyer C, et al. Person-Centered Maternity Care in Low- and Middle-Income Countries: Analysis of Data from Kenya, Ghana, and India. Lancet Glob. Health. 2018;

[25] Donabedian A. Evaluating the Quality of Medical Care. Milbank Q. 2005;83:691–729.

[26] Health Statistics and Information Systems. Service Availability and Readiness Assessment (SARA): an annual monitoring system for service delivery [Internet]. World Health Organization (WHO); 2015 [cited 2018 Jun 4]. p. 178. Report No.: 2.2. Available from: https://www.who.int/healthinfo/systems/SARA_Reference_Manual_Full.pdf.

[27] Hill S. Putting the priorities first: medicines for maternal and child health. Bull. World Health Organ. 2012;90:87.

[28] Nesbitt RC, Lohela TJ, Manu A, et al. Quality along the Continuum: A Health Facility Assessment of Intrapartum and Postnatal Care in Ghana. PLOS ONE. 2013;8:e81089.

[29] Sheffel A, Karp C, Creanga AA. Use of Service Provision Assessments and Service Availability and Readiness Assessments for monitoring quality of maternal and newborn health services in low-income and middle-income countries. BMJ Glob. Health [Internet]. 2018 [cited 2019 Mar 8];3. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6267320/.

[30] Demographic and Health Surveys Program. SPA Overview. DHS Program. 2018 [cited 2019 Mar 7]. Available from: https://dhsprogram.com/What-We-Do/Survey-Types/SPA.cfm.

[31] Stata Statistical Software: Release 12. College Station, TX: StataCorp LP; 2015.

[32] Chace Z. How Perverse Incentives Drive Up Health Care Costs [Internet]. Planet Money. NPR; 2014 [cited 2018 Jun 4]. Available from: https://www.npr.org/2014/01/16/262946913/how-perverse-incentives-drive-up-health-care-costs.

[33] Weeks J. "Perverse Incentives" and the Triple Aim: Overcoming the Troubled Path to Economic Integration for Integrative Medicine and Health. Glob. Adv. Health Med. 2015;4:7–9.

[34] Palmer N, Mueller DH, Gilson L, et al. Health financing to promote access in low income settings—how much do we know? The Lancet. 2004;364:1365–1370.

[35] Martin EG, Begany GM. Opening government health data to the public: benefits, challenges, and lessons learned from early innovators. J. Am. Med. Inform. Assoc. 2017;24:345–351.

[36] National Portal of India. Ayushman Bharat - National Health Protection Mission [Internet]. Gov. India. 2018 [cited 2018 Sep 17]. Available from: http://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=1518544.

[37] National Health Authority. Everyday PM-Jay Update. Government of India. 2019 [cited 2019 Mar 7]. Available from: https://www.pmjay.gov.in.

[38] Bernard RP, Bhatt RV, Potts DM, et al. Seasonality of birth in India. J. Biosoc. Sci. 1978;10:409–421.

[39] Lam DA, Miron JA. Seasonality of births in human populations. Soc. Biol. 1991;38:51–78.