Comparison of the results of in-person and mobile phone surveys for a health facility assessment in Tajikistan: a validation study protocol

Pablo Amor Fernandez*,1
Rachel Neill*,1
Ruchika Bhatia2
Jigyasa Sharma2
Kathryn Andrews2
Sven Neelsen2
Etoile Pinder3
Marifat Abdullaev4
Firuza Safarova5
Mutriba Latypova2
Mirja Channa Sjoblom2
Tashrik Ahmed1
Michael A. Peters1
Ashley Sheffel1
Tawab Hashemi1

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Peter Meredith Hansen¹

Gafur Muhsinzoda⁶

Gil Shapira²

+ Co-first authors who contributed equally

Affiliations

¹ The Global Financing Facility for Women, Children, and Adolescents, Washington, District of Columbia, USA

² The World Bank, Washington, District of Columbia, USA

³ Sanigest Internacional, Panamá, Panama

⁴ Department of Medicine, Tajik National University, Dushanbe, The Republic of Tajikistan

⁵ Consultant, Dushanbe, The Republic of Tajikistan

⁶ Ministry of Health and Social Protection of the Population of the Republic of Tajikistan, Dushanbe, The Republic of Tajikistan

Corresponding author

Rachel Neill

rneill@worldbank.org
ABSTRACT

Health facility assessments provide important data to measure the quality of health services delivered to populations. These assessments are comprehensive, resource intensive, and periodic to inform medium- to longer-term policies. However, in absence of other reliable data sources, country decision makers often rely on outdated data to address service delivery challenges that change more frequently. High-frequency phone surveys are a potential option to improve the efficiency and timeliness of collecting time-sensitive service delivery indicators in-between comprehensive in-person assessments. The objectives of this study are to assess the reliability, concurrent criterion validity, and non-response rates in a rapid phone-based health facility assessment developed by the Global Financing Facility’s FASTR initiative compared to a comprehensive in-person health facility assessment developed by the World Bank’s Service Delivery Indicators Health Program. The in-person survey and corresponding in-person item verification will serve as the gold standard. Both surveys will be administered to an identical sample of 500 health facilities in Tajikistan using the same data collection entity. To assess reliability, percent agreement, Cohens Kappa, and prevalence and bias adjusted Kappa will be calculated. To assess concurrent criterion validity, sensitivity and specificity will be calculated, with a cut-off of .7 used for adequate validity. The study will further compare response rates and dropout rates of both surveys using simple t-tests and balance tests to identify if the characteristics of the phone-based and in-person survey samples are similar after accounting for any differences in survey response rates. The results of this study will provide important insights into the reliability and validity of phone-based data collection approaches for health facility assessments. This is critical as Ministries of Health seek to establish and sustain more continuous data collection, analysis, and use of health facility-level data to complement periodic in-person assessments to improve the quality of services provided to their populations.
INTRODUCTION

Large-scale, in-person health facility assessments (HFAs) offer detailed data on health systems functioning and quality of care, making them indispensable for evidence-informed decision-making and health policy development. Several holistic health facility assessments exist – including the Harmonized Health Facility Assessment (HHFA), the Service Availability and Readiness Assessment (SARA), the Service Provision Assessment (SPA), and the Service Delivery Indicators survey (SDI) – that paint a detailed picture of health facility readiness and service delivery. These surveys are important complements to administrative data in low- and middle-income countries (LMICs) by providing detailed information on the supply-side of the health system including the availability of key health systems inputs, the functioning of health facility management, providers’ clinical knowledge, and patients’ satisfaction with care.

Large-scale in-person HFAs are, by design, meant to be comprehensive in their assessment, requiring considerable time, financial and human resources. They are conducted periodically, often every five to seven years, to inform medium-to-longer-term policies and planning. However, in absence of other frontline service delivery data sources that are reliable and comprehensive, country decision makers are often forced to rely on HFA data that is often outdated, with little ability to diagnose and address supply-side service delivery challenges that may be changing more frequently.

In an ideal world, comprehensive in-person HFAs would be complemented with both high-quality administrative data and high-frequency assessments. Large scale, comprehensive in-person HFAs offer a detailed understanding of the health system and are well suited to capture systemic changes that typically unfold over longer periods. Meanwhile, data gathered from high-quality administrative sources and high-frequency assessments can provide timely insights into key indicators expected to change more frequently. This is particularly important in fragile and conflict-affected contexts and during shocks (e.g.,
epidemics or natural disasters) when frequent contextual changes make timely adaptation and response especially critical. For example, immediate disruptions in human resources or medical supplies – whether due to sudden crises or chronic health systems challenges – can be more swiftly managed using higher frequency, longitudinal data, while the in-person observation by skilled enumerators found in in-person HFAs is critical for detailed data on provider competencies. Thus, the integration of comprehensive HFAs with ongoing administrative data analysis and high-frequency assessments has the potential to enable a more responsive and resilient health care system.

Phone surveys are one potential approach to enable high-frequency HFAs. Previous research indicates that phone-based data collection is an affordable approach to collecting both qualitative and quantitative data. During the COVID-19 pandemic, more frequent, locally adapted phone survey methodologies were utilized to monitor essential health service use and the impact of the pandemic on health facility readiness to deliver services. This demonstrated that phone-based survey approaches can provide frequent and cost-effective results in LMIC settings.

Growing evidence indicates that interview administered phone surveys of households and health workers can have utility and validity; however, there is less evidence that phone-based health facility assessments can capture comprehensive details on health facility functioning with limited bias. Two studies offer promising results. In Malawi, Pattnaik et al. conducted direct observations and register reviews at facilities to assess select aspects of phone survey validity related to family planning indicators. The study found that most health facility indicators collected by phone were above a 70% sensitivity threshold compared to in-person records and inspections and determined that phone data collection was both feasible and highly cost-effective. In Kenya, Ashigbie et al. assessed the validity of phone-assessed medicine availability indicators at health facilities compared to in-person visits. The study found strong health facility-level agreement on medicine availability between the phone survey and subsequent in-person assessment (kappa = 0.90; confidence interval (CI) 0.88-0.92); further, the phone survey was
conducted with a unit cost of $19.73 compared to $186.20 for the in-person assessment\textsuperscript{15}. Both studies concluded that phone-based surveys were valid and cost-effective compared to existing gold standards of either in-person data collection or inspection visits for a limited set of indicators\textsuperscript{14,15}.

Despite promising initial results, a key concern for phone-based surveys is minimizing respondent fatigue by keeping the survey administration short. Like in-person surveys, phone-based interviews that take a longer period are more likely to lead to respondent fatigue, which can give rise to threats to validity. For example, recent household survey research quantified that a 15-minute addition in questionnaire administration time resulted in 8-to-17\% less detailed response over the phone\textsuperscript{16}. Similarly, Abate et al. (2023) randomly assigned household survey participants to in-person or phone-based survey administration for a household nutrition consumption survey\textsuperscript{17}. They found that average per-capita consumption reported by respondents was 23\% lower and the estimated poverty headcount was twice as high in the phone-based survey compared to the in-person survey due to underreporting of food items consumed, which the authors attributed to respondent fatigue\textsuperscript{17}. These findings emphasize the importance of developing phone-specific questionnaires which should be shorter, simpler, and less cognitively burdensome to complete than typical in-person assessments\textsuperscript{11}.

An additional limitation of phone-based data collection is that the information provided by respondents is generally self-reported and cannot be directly verified by enumeration using observation. This poses specific validity questions in health facility surveys which rely on in-person verification of medical equipment, drugs, health workers, registry logs, and physical infrastructure. Moreover, more complex indicators cannot be effectively administered over the phone. For example, while a phone-based survey might approximate in-person measurement for basic infrastructure availability or the number of staff assigned to a health facility, it is likely unsuitable for assessing more intricate constructs like provider performance and adherence to clinical protocols.
Finally, phone-based data collection is optimally conducted with a single respondent, while most large-scale, comprehensive HFAs rely on multiple respondents, often starting with the health facility manager but going on to include respondents across the health facility (e.g., service providers, pharmacist, financial administrator/accountant, etc.). This means that the phone-based survey requires a careful selection of the key questions or domains that a single respondent is best suited to answer.

Despite the rise in phone-based surveys, promising early validation results, and these known limitations, there has not been a validation study of a nationally representative, phone-based HFA to our knowledge. This is a critical gap in the evidence – can phone-based surveys measure key constructs that are commonly measured via in-person HFAs? And what bias, if any, is introduced? This study will fill that gap by conducting a head-to-head comparison of a nationally representative, comprehensive, in-person HFA and a nationally representative, rapid, phone-based HFA. Our findings will have important implications on the feasibility of adopting phone-based surveys as a key complement to comprehensive HFAs to improve the timeliness and responsiveness of health systems data to better align with real-world decision-making needs.

**Study Objectives and Contributions**

This study has three primary objectives. The first objective is to assess the inter-method reliability of the rapid phone-based HFA compared to the comprehensive in-person HFA. This objective focuses on isolating the survey mode effect – the differences in results due to the phone and in-person administration – and will determine the extent to which phone data collection is a reliable alternative mode for conducting HFAs for a subset of key PHC service delivery indicators. To minimize bias, this objective will include only the survey questions that are identical in both the rapid phone-based HFA and the comprehensive in-person HFA. By keeping questions the same and varying only the administration mode, we can effectively isolate and examine the mode effect.
The second objective is to assess the concurrent criterion validity of indicators in the rapid phone-based HFA compared to indicators in the comprehensive in-person HFA. As discussed earlier, phone surveys require questions that are shorter, simpler, and less cognitively demanding. While this design is crucial for phone-based assessments, it can lead to issues such as questions being interpreted differently by respondents or lack of clarity or confusion on specific terms in the asked question that may affect data quality. This objective therefore focuses on evaluating the extent to which the rapid phone-based HFA can accurately measure the same underlying indicators as that in a comprehensive in-person HFA via a simplified questionnaire deemed suitable for phone-based administration.

The third objective is to compare the survey response rates and dropout rates of the rapid phone-based HFA to those of the comprehensive in-person HFA. Survey response rates are crucial for interpreting results and identifying potential biases. This is especially important for phone surveys where respondents can more easily drop out or refuse to participate in the survey. For instance, unlike for in-person surveys wherein respondents might feel more obligated to participate in a survey due to the physical presence of the interviewer, the respondents might find it easier to not answer a phone call or hang up. Additionally, phone surveys may also have higher non-response rates due to poor connectivity or time constraints faced by respondents.

Taken together, these objectives will demonstrate whether the rapid phone-based HFA can approximate the results of the comprehensive in-person HFA for a select subset of facility-level indicators.

As phone-based data collection in LMICs is a rapidly evolving field with growing interest following the COVID-19 pandemic, publication of this protocol aims to inform the broader scientific community of this undertaking and improve the transparency of our subsequent analysis against the study design articulated below.

MATERIALS AND METHODS
This study will take place in Tajikistan, a landlocked, and mountainous country in Central Asia. Tajikistan has low population density with most urban areas concentrated in the eastern part of the country\textsuperscript{18}. Over 70\% of the country’s 10.1 million people live in rural areas\textsuperscript{18,19}. Health care is overwhelmingly provided in the public sector. The primary healthcare services subject to this validation study are delivered through a combination of rural health centers (RHC) and associated health houses (HH) in rural areas, and by district and city health centers’ family medicine departments in urban settings. Primary care provision suffers from dilapidated infrastructure, shortages of equipment, and a low ratio of doctors to the total population, particularly in rural areas\textsuperscript{20}. The low ratio of doctors to the population is particularly challenging given Tajikistan’s low population density and inaccessibility in mountainous rural areas. To strengthen Tajikistan’s health service delivery system and improve population health outcomes, it is critical that health facilities can provide quality services to patients. This includes ensuring that health workers are well trained and capacitated, addressing overcrowding of health facilities, maintaining adequate supplies and drugs, and having an effective referral system in place.

The World Bank (WB) and the Global Financing Facility for Women, Children, and Adolescents (GFF) are supporting the efforts of the Ministry of Health and Social Protection of the Republic of Tajikistan to improve the quality of PHC service delivery. As part of this initiative, they are conducting a series of surveys to assess the quality of PHC in the country. Given Tajikistan’s lack of comprehensive data on PHC service delivery, the WB and GFF are supporting the implementation of a comprehensive in-person HFA to get a better understanding of the country’s PHC system, identify its strengths and weaknesses, and provide comprehensive data to guide policy decisions. In addition to the comprehensive in-person HFA, the World Bank and GFF are also implementing a phone-based HFA to be conducted every 6-12 months to monitor intervention implementation, inform course correction, and identify new challenges or shocks.
that emerge over time. Conducting comprehensive in-person HFA and rapid phone-based HFA concurrently in the same setting therefore has offered a unique opportunity to conduct a head-to-head comparison of the results from the two surveys.

**Survey instruments**

The two surveys that will be conducted are: (1) the GFF’s Frequent Assessment and Health Systems Tools for Resilience (FASTR) Initiative’s rapid-cycle health facility phone survey which is designed for interview-administered phone-based data collection at the PHC level and which will serve as the rapid phone-based HFA and (2) the World Bank’s Service Delivery Indicators (SDI) health survey which will serve as the comprehensive in-person HFA.

The FASTR rapid-cycle health facility phone survey (subsequently referred to as the rapid phone-based HFA) is a PHC assessment tool that aims to monitor service availability, readiness, and functioning of PHC facilities over time with an emphasis on reproductive, maternal, newborn, child, and adolescent health and nutrition services. The tool’s measurement approach is grounded in the WHO and UNICEF’s Primary Health Care Measurement Framework and Indicators. The questionnaire was drawn from existing comprehensive HFAs as well as health facility phone survey experience during the COVID-19 pandemic. The survey is administered to a nationally or sub-nationally representative panel sample of PHC facilities over four quarterly data collection contacts per year. Quarterly data collection enables both the tracking of presumed higher-variability indicators over time (e.g., medicine stock, utility functioning) and the distribution of annual indicators across four survey rounds to shorten survey administration and reduce respondent fatigue. The phone survey also includes questions that capture respondent perceptions, such those related to challenges faced by the facility. These perception-based questions are a useful way to receive timely feedback from facilities about challenges that they face but are outside the
scope of this validation study. As of August 2024, the tool has been implemented in Burkina Faso, Vietnam, Senegal, Bangladesh, Tajikistan, and Madagascar.

The SDI (subsequently referred to as the comprehensive in-person HFA) health survey program began in 2008 and has been implemented in over a dozen countries across the globe. The survey was revamped in 2019 to better align with the recent developments in the literature on measuring quality of PHC service delivery. Building on the frameworks for high quality health systems, the survey now assesses processes of care and patient outcomes along with structural inputs to provide a comprehensive picture of PHC service delivery in a country. The SDI survey includes three questionnaires: health facility questionnaire, health care provider questionnaire, and outpatient exit interview. Only the health facility questionnaire will be included in the validation study. The SDI comprehensive in-person HFA, and its corresponding in-person verification of the availability and functionality of structural inputs, will be considered the gold standard for the purpose of this study.

**Questionnaire adaptation**

The two tools were adapted to Tajikistan’s context via several steps. The initial step involved modifying response options, revising question framing, and adding or removing questions to align with the country’s PHC needs and priorities. This included incorporating details such as locally available infrastructure, the country’s health workforce structure, clinical guidelines for diagnosis and treatment of common conditions, and the expected list of tracer equipment, medical supplies, and drugs available at different types of health facilities. Additional questions (including entirely new sections in the questionnaires) were developed with input from subject matter experts from WB, GFF, and Tajikistan health systems stakeholders.

This preliminary contextualized questionnaire was shared with a joint working group comprising of local experts, practitioners, and government to capture Tajikistan’s unique primary health care
context and to ensure appropriateness for different levels of Tajikistan’s health system. The surveys then underwent field testing at selected facilities to further improve the questionnaires based on respondent feedback and identify and address issues that emerge during the pilot. This involved improving option choices for specific questions, improving question framing such as adding clearer and easy to understand definitions of certain terms, and deleting questions that might not be relevant at the facility level.

The final versions of the rapid phone-based and comprehensive in-person HFA questionnaires then underwent a detailed review process to ensure accuracy and consistency in translation. The English versions of both questionnaires were translated to Russian by the same translator, reviewed by local context experts, and back translated to English to identify any discrepancies that might affect question clarity.

**Indicator mapping**

After adaptation, a systematic indicator alignment process across the rapid phone-based HFA and the comprehensive in-person HFA was conducted. First, all survey items (survey questions) from the seven survey modules of the rapid phone-based HFA (services, infrastructure, human resources, medical supplies and equipment, leadership and coordination, community engagement, quality improvement) were transferred into a Microsoft Excel document. Next, all the survey items from the comprehensive in-person HFA questionnaire that measured the same underlying indicator as the rapid phone-based HFA were identified and mapped to the relevant survey items in the rapid phone-based HFA in the same excel document. At the conclusion of this exercise, 102 survey items from the rapid phone-based HFA were mapped to 129 survey items from the comprehensive in-person HFA.

Following the mapping, a question-by-question revision activity was conducted to identify the two types of survey items mapped across both the questionnaires: (1) ‘direct’ survey items and (2) ‘indirect’
Survey items were categorized as ‘directly mapped’ when the framing and underlying indicator of the survey item were identical (or nearly identical) across both surveys. This required a one-to-one match of a question in the rapid phone-based HFA and the comprehensive in-person HFA. These survey items were commonly the ones wherein the comprehensive in-person HFA question structure was appropriate for phone administration. For instance, both surveys include the same question to identify the source of electricity at health facilities (‘What is the health facility’s main source of electricity?’). We also considered questions with identical wording across both surveys but which include enumerator verification through direct observation in the comprehensive in-person HFA "directly mapped", because we consider in-person verification as a component of the mode effect. This includes questions on the availability and functioning of facility inputs, such as private toilets, where the enumerators in the comprehensive in-person HFA can verify the response due to being physically present in the health facilities. Directly mapped indicators will be included in objective one (reliability/mode effect) and objective two (validity) testing.

Survey items were classified as 'indirectly mapped' when they intended to measure similar underlying indicators in both surveys, but differed in the way they were asked or framed due to feasibility for phone administration. This typically occurred when comprehensive in-person HFA survey items were too lengthy or detailed to be administered over the phone, leading to simpler alternatives being included in the rapid phone-based HFA. For example, in measuring HIV/AIDS service provision, the rapid phone-based HFA asks whether the health facility provides HIV/AIDS services, while the comprehensive HFA asks whether the health facility was able to provide HIV/AIDS diagnosis services and HIV/AIDS treatment services in the 3 months prior to data collection. In this case, the underlying indicator "HIV/AIDS service availability" was common in both the surveys but was captured differently in each, leading to it being "indirectly mapped."
Since the "indirectly mapped" survey items differed in question framing, they were not suitable for the inter-method reliability test (objective 1) to isolate the mode effect (phone v/s in-person). Any difference in the survey results could be attributed to various factors, such as respondents interpreting the questions differently, rather than solely due to the mode of administration. Therefore, these survey items were suitable only for assessing objective two (validity).

Table 1 provides a summary of the distribution of both types of survey items. An item-by-item mapping is provided in S1 Appendix. The majority (75%) of the questions were deemed ‘indirectly mapped’ which in and of itself is a finding, demonstrating that few questions are likely to be suitable for identical administration across phone and in-person HFAs.

<table>
<thead>
<tr>
<th>Survey Module</th>
<th>Total number of questions in the study</th>
<th>Direct indicators (reliability and validity objectives)</th>
<th>Indirect indicators (validity objective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SERV Services</td>
<td>14</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>INF Infrastructure</td>
<td>15</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>HR Human Resources</td>
<td>8</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>SUP Medical supplies and equipment</td>
<td>40</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>LC Leadership and coordination</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>COM Communication</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>QI Quality Improvement</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>
Data collection

Data collection will take place from August to October 2024. This study was approved by the Government of the Republic of Tajikistan Ministry of Health and Social Protection Biomedical Ethics Committee, with the Order No. 60 and number N2148 on May 23\textsuperscript{rd}, 2024. Informed consent will be obtained from all survey respondents.

Sampling

The sampling methodology for both surveys is designed to be representative of the PHC health facilities in Tajikistan at the national and oblast (region) levels. Since the role of private sector in providing PHC services is minimal in Tajikistan, the sample frame consists of all public PHC health facilities in the country comprising of City and District Health Centers (family medicine departments), Rural Health Centers, and Health Houses. The sample was stratified across all oblasts. To obtain quality results from the comparison of the rapid phone-based HFA and the comprehensive in-person HFA, each survey will be conducted in the identical sample of 500 facilities described in Table 2.

Table 2. Sample for the rapid-phone based and comprehensive in-person HFAs

<table>
<thead>
<tr>
<th>Oblast</th>
<th># Facilities</th>
<th># Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dushanbe</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Sughd</td>
<td>616</td>
<td>151</td>
</tr>
<tr>
<td>Khaktlon</td>
<td>1,123</td>
<td>150</td>
</tr>
<tr>
<td>Gorno-Badakhshan Autonomous Region</td>
<td>227</td>
<td>84</td>
</tr>
</tbody>
</table>
The enumerators will obtain verbal consent for the rapid phone-based HFA and written consent for the comprehensive in-person HFA from the respondents before starting the survey. If a respondent declines to participate, the reason for their refusal will be documented, and the facility will be replaced by a facility from the same strata whenever available. In the event of a non-response, health facilities will be replaced from a "replacement list" curated from the sample frame for each survey, following the same protocol of random sampling selection. Furthermore, in the event of a non-response, replacements for the two surveys will be done independently. This means that if a facility does not participate in the rapid phone-based HFA and needs to be replaced with another health facility, it will not be simultaneously replaced in the comprehensive in-person HFA, and vice-versa, unless the non-participation is due to the closure of the facility. As a result, the final samples may differ slightly due to varying response rates.

In both surveys, the officer-in-charge of the health facility will serve as the primary respondent. This individual is best positioned to provide comprehensive insights into the facility's overall readiness.

However, in situations where the designated officer-in-charge is unable to participate or believes another staff member would be better suited to answer specific sections or questions, the survey response could be delegated to a deputy or another qualified health worker.
Data collection procedures

Both surveys will involve comprehensive training for enumerators to familiarize them with survey protocols, questionnaire administration, and ethical considerations for each assessment. This will be followed by additional pilot testing the instrument in a few health facilities to identify any necessary changes and to provide enumerators with field experience before starting data collection.

At least one week before data collection commences, a data collection supervisor will contact the PHC network managers (individuals who manage several public health facilities) via phone to introduce themselves and both surveys. They will also request the names and contact information of the relevant sampled health facilities under their supervision. The PHC managers should have received the official notice from the Government regarding the surveys and that their facilities may have been included in the sample. If they do not have the official communication, the supervisor will share with them a copy of the official letter. PHC network managers will then be responsible for notifying officers-in-charge of sampled health facilities.

The rapid phone-based HFA is designed as a short 30 to 45 minutes survey, requiring no preparation by the officer-in-charge. Respondents will have the option to either start answering the survey during the first call, or to schedule it for a more convenient day or time. If a respondent is unreachable, the rapid-cycle survey protocol requires at least five contact attempts on different days and at varying times to complete the interview, maximizing communication success and reducing potential biases introduced by timing constraints. This will be particularly important for remote facilities with low phone connectivity, where the respondent may have sporadic access to a stable phone network. In instances where immediate completion of the interview is not feasible due to telecommunications issues or respondent availability, follow-up calls may be scheduled for a later date or time.
For the comprehensive in-person HFA, enumerator teams will contact the officers-in-charge of sampled facilities two weeks before data collection, to inform them about the survey, coordinate the date of the first visit, and share the ‘advance questionnaire’. The advance questionnaire is a subset of the comprehensive in-person HFA and comprises of questions that require reference to facility data records. This will be shared with the officer-in-charges at least two weeks prior to data collection day, allowing respondents to prepare responses, thereby saving time and improving data quality on the day of data collection.

If facilities have not received the official communication about their participation in the survey, PHC managers will be contacted to ensure that all selected facilities receive the official letter requesting their participation. For the comprehensive in-person HFA, one week before data collection, enumerator teams will again contact the officers-in-charge to confirm the date of the first visit and request completion of the advance questionnaire. This follow-up will also provide an opportunity to address any questions the officers-in-charge may have about the survey and to reschedule the first visit, if necessary. Furthermore, if on the day of data collection, the facility is closed, or the officer-in-charge is not present, or the data cannot be collected for any other reason, the enumerator teams will coordinate and reschedule the visit. If, however, the officers-in-charge refuse to participate in the study, their non-response along with their reason for refusal will be recorded, and the facility will be replaced using random sampling from the required strata.

**Timing of survey administration**

The comprehensive in-person HFA and the rapid phone-based HFA will be conducted as close in timing as possible, ideally within a two-week period for each health facility. The close timing of the surveys minimizes the risk of external factors (for example, a health facility receiving new medical supplies) impacting the data recorded in each survey and will ensure a more reliable comparison.
However, there is a risk that conducting both surveys too close will introduce respondent fatigue that might affect results. Therefore, the order of the surveys for each health facility will be randomized to control for some of the real changes that may occur due to the order, with 50% of facilities receiving the rapid phone-based HFA first and 50% of the facilities receiving the comprehensive in-person survey first. This will allow for a more robust analysis. It will also allow us to conduct a longitudinal analysis of responses for the rapid phone-based HFA and the comprehensive in-person HFA in each facility to identify possible confounding effects depending on which survey the facility receives first. We would expect this to be a possibility especially in facilities that receive the in-person survey first, as respondent recall may be primed by the recent in-person verification of survey items and therefore, respondents may answer the rapid phone-based HFA with greater validity.

Data analysis

Analysis will be conducted on a paired sample of each matched health facility that participated in both surveys.

Objective One: Inter-Reliability Analysis

The mode effect (phone v. in-person) of the rapid phone-based HFA compared to the comprehensive in-person HFA will be analyzed using the “directly mapped” survey items. The results will be analyzed and reported in three ways: as a simple percent agreement of responses, using the Cohen’s Kappa to account for the percent of agreement that is expected by chance, and using a prevalence and bias adjusted kappa (PABAK) to account for prevalence and bias in the sample. Calculating PABAK is especially important for a HFA, where many of the survey items are expected to be available and therefore, the expected agreement by chance for many indicators is high due to high prevalence27. Taken together, these three measures will provide a nuanced understanding of the results28.
There are some differences on what is considered adequate for reliability. While literature suggests that a Kappa coefficient larger than 0.61 would be reliable, it is acknowledged that these thresholds can be somewhat arbitrary. To address this, a scale for the level of agreement will be used: almost perfect agreement above 0.90, strong 0.80 to 0.90, moderate 0.60 to 0.79, weak 0.40 to 0.59, and minimal 0.20 to 0.39.

**Objective Two: Concurrent Criterion Validity Analysis**

This objective focuses on concurrent criterion validity – the extent to which a new measurement technique agrees with a gold standard when the assessments are undertaken at approximately the same time. Both the “directly mapped” and the “indirectly mapped” indicators will be dichotomized to understand the differences in the results between the two surveys with the comprehensive in-person HFA being the gold standard. Sensitivity and specificity will be used to assess whether each indicator meets the predefine validation threshold of 0.7. This has been chosen as the validation threshold because of the specific intent to use low cost, high frequency surveys to generate timely signals as a complement to, rather than replacement for, periodic in-person surveys which are still needed for their depth, comprehensiveness, and rigor. We will also compare results stratified by direct versus indirect indicators and by survey domain (e.g., infrastructure, human resources) to identify any differences.

Regression analysis will be conducted to assess whether the values are associated with any of the seven domains of the rapid phone-based HFA, facility characteristics, respondent characteristics and potentially enumerator characteristics and timing between the two surveys. Sensitivity checks will also be conducted to see if the duration of time between the two surveys is correlated with the discordant answers.

Regression analysis will also be conducted to identify any heterogeneity effects due to facility characteristics. Subgroup analysis will be conducted based on (1) oblast, (2) rural versus urban, and (3)
facility type to identify if any underlying, facility characteristics influence the reliability and validity of the rapid, phone-based HFA results. This will identify whether validity is different in particular sub-groups. For example, the rapid, phone-based HFA may have improved validity in smaller facilities where one respondent is likely to be familiar with all the tracer items and could be more challenging in a larger facility with multiple departments.

**Objective Three: Non-Response Rate**

Although both surveys will initially include the same sample of 500 facilities, we expect that the final samples for each HFA will differ in the end due to varying non-response rates. A simple-t test will be used to assess whether there is a significant difference in the response rates between the rapid phone-based HFA and the comprehensive in-person HFA. Balance tests will also be used to assess whether the characteristics of the final sample of included facilities (the original sample plus any replacements) for the comprehensive in-person HFA and the rapid phone-based HFA sample remain similar across both surveys.

In case a significant difference is found in response rate or sample characteristics, regression analysis will be used to assess associations with discordance in survey completion. This will be done by creating a binary variable to indicate whether the comprehensive in-person HFA was completed in a health facility, while the rapid phone-based HFA wasn’t. This binary variable will then regressed on facility characteristics and whether the rapid phone-based HFA was initiated before or after the comprehensive in-person HFA. Since the non-response rate is expected to be higher for the rapid phone-based HFA, this analysis will only be done for health facilities which complete the comprehensive in-person HFA but not the rapid phone-based HFA.

**DISCUSSION**
To the best of our knowledge, this study is the first to assess the reliability and validity of a rapid phone-based HFA against a comprehensive in-person HFA. This will fill an important gap in global evidence on the reliability and validity of phone-based approaches to administer HFAs to monitor a subset of key facility-level indicators. This insight is especially critical given the rapid rise of phone-based survey approaches in LMICs following the COVID-19 pandemic as an innovative method for more frequent, lower-cost, facility-level, primary data collection for select indicators.

Most immediately, the findings of this study will be used to refine the GFF FASTR Initiative’s rapid health facility phone survey assessment prior to its scale-up in multiple countries in Sub-Saharan Africa and Asia in 2025 and beyond. The revised version of the survey instrument will then be made publicly available on the GFF’s data portal (https://data.gffportal.org/) alongside a survey manual which will enable the broader public health research and practice community to utilize the toolset. Within the study context and the GFF’s FASTR initiative more broadly, the intent is to use low cost, high frequency phone surveys as a complement to, rather than a substitute for, periodic and more comprehensive in-person facility assessments.

More broadly, findings from this study will be used to provide measurement and methodological guidance for HFAs. The findings will provide novel insights on the complementarities of various methods, which HFA domains and sub-domains can be administered via phone with high reliability and validity, and which HFA domains are likely to require in-person administration or other verification approaches. It is possible that the findings could be robust enough to generate correction factors for phone-based surveys that can be applied to approximate in-person measures in contexts where traditional data collection approaches are not possible (such as during health emergencies or conflict-impacted areas). Finally, the findings on response rates will be published to inform future implementation guidance on the feasibility of different methods of data collection given HFA operational constraints and measurement needs in different contexts and when each method is likely to be particularly useful.
CONCLUSION

HFAs provide critical information on the quality of health service delivery, making them a vital source of data to inform the design, implementation, and evaluation of health care policies and programs. Validation studies are critical to informing how and when different types of data collection approaches are appropriate and what can be feasibly measured. This research will improve understanding on the validity of phone-based data collection approaches for HFAs to collect data on a select subset of key service delivery indicators, providing practical insights on the appropriateness, feasibility, and complementarity of different survey modalities in capturing health facility service availability and readiness. This is an important step towards building and sustaining continuous data collection, analysis, and data use for decision making across all levels of the health system in order to improve the quality of service provision, in Tajikistan and globally.

ACKNOWLEDGEMENTS

For their support in this study, the authors wish to thank the Government of the Republic of Tajikistan and the Ministry of Health and Social Protection of the Population of the Republic of Tajikistan (MOHSP), in particular Dr. Davlatmurod Olimov, Deputy Head of the State Health and Social Protection Supervision Service (Khadamot) under MOHSP, and Dr. Ziyodullo Idrisov, Lead Specialist of the MOHSP Directorate for Health Reforms, PHC and International Cooperation, and as well as the members of the Joint Working Group on Adaptation of Tools for Primary Health Care Quality Survey in Tajikistan. Finally, the authors wish to thank Talip Kilic of the World Bank Group for his comments on an earlier draft of the study protocol.

Funding for this study was received from the Global Financing Facility for Women, Children, and Adolescents and the World Bank Group. The findings, interpretations, and conclusions expressed in this
paper are entirely those of the authors and do not necessarily represent the views of the World Bank, its executive directors, and the governments of the countries they represent.

**AUTHOR CONTRIBUTIONS**

PAF, RN, TA, MP, GS, RB, KA, JS, AS, ML, EP, GM, and SN conceptualized and designed the study. MA and FS provided resources and materials and contributed to conceptualization and design. MS, TH and PH led funding acquisition and contributed to the conceptualization of the study. PAF, RN, and RB wrote the first draft of the manuscript. All authors provided critical review and editing of the manuscript and approve the manuscript for publication.

**REFERENCES**


https://ideas.repec.org/p/wbk/wbrwps/9636.html


18. Tajikistan. The World Bank Group


