

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

3    Pablo Amor Fernandez<sup>+,1</sup>

4    Rachel Neill<sup>+,1</sup>

5    Ruchika Bhatia<sup>2</sup>

6    Jigyasa Sharma<sup>2</sup>

7    Kathryn Andrews<sup>2</sup>

8    Sven Neelsen<sup>2</sup>

9    Etoile Pinder<sup>3</sup>

10   Marifat Abdullaev<sup>4</sup>

11   Firuza Safarova<sup>5</sup>

12   Mutriba Latypova<sup>2</sup>

13   Mirja Channa Sjoblom<sup>2</sup>

14   Tashrik Ahmed<sup>1</sup>

15   Michael A. Peters<sup>1</sup>

16   Ashley Sheffel<sup>1</sup>

17   Tawab Hashemi<sup>1</sup>

18 Peter Meredith Hansen<sup>1</sup>

19 Gafur Muhsinzoda<sup>6</sup>

20 Gil Shapira<sup>2</sup>

21 + Co-first authors who contributed equally

22 **Affiliations**

23 <sup>1</sup> The Global Financing Facility for Women, Children, and Adolescents, Washington, District of  
24 Columbia, USA

25 <sup>2</sup> The World Bank, Washington, District of Columbia, USA

26 <sup>3</sup> Sanigest Internacional, Panamá, Panama

27 <sup>4</sup> Department of Medicine, Tajik National University, Dushanbe, The Republic of Tajikistan

28 <sup>5</sup> Consultant, Dushanbe, The Republic of Tajikistan

29 <sup>6</sup> Ministry of Health and Social Protection of the Population of the Republic of Tajikistan,  
30 Dushanbe, The Republic of Tajikistan

31 **Corresponding author**

32 Rachel Neill

33 [rneill@worldbank.org](mailto:rneill@worldbank.org)

34

35

36 **ABSTRACT**

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40 decision makers often rely on outdated data to address service delivery challenges that change more  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

## 59 INTRODUCTION

60 Large-scale, in-person health facility assessments (HFAs) offer detailed data on health systems  
61 functioning and quality of care, making them indispensable for evidence-informed decision-making and  
62 health policy development. Several holistic health facility assessments exist – including the Harmonized  
63 Health Facility Assessment (HHFA)<sup>1</sup>, the Service Availability and Readiness Assessment (SARA)<sup>2</sup>, the  
64 Service Provision Assessment (SPA)<sup>3</sup>, and the Service Delivery Indicators survey (SDI)<sup>4</sup> – that paint a  
65 detailed picture of health facility readiness and service delivery. These surveys are important  
66 complements to administrative data in low- and middle-income countries (LMICs) by providing detailed  
67 information on the supply-side of the health system including the availability of key health systems inputs,  
68 the functioning of health facility management, providers’ clinical knowledge, and patients’ satisfaction  
69 with care.

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

76 In an ideal world, comprehensive in-person HFAs would be complemented with both high-quality  
77 administrative data and high-frequency assessments. Large scale, comprehensive in-person HFAs offer a  
78 detailed understanding of the health system and are well suited to capture systemic changes that typically  
79 unfold over longer periods. Meanwhile, data gathered from high-quality administrative sources and high-  
80 frequency assessments can provide timely insights into key indicators expected to change more  
81 frequently. This is particularly important in fragile and conflict-affected contexts and during shocks (e.g.,

82 epidemics or natural disasters) when frequent contextual changes make timely adaptation and response  
83 especially critical. For example, immediate disruptions in human resources or medical supplies – whether  
84 due to sudden crises or chronic health systems challenges – can be more swiftly managed using higher  
85 frequency, longitudinal data, while the in-person observation by skilled enumerators found in in-person  
86 HFAs is critical for detailed data on provider competencies. Thus, the integration of comprehensive HFAs  
87 with ongoing administrative data analysis and high-frequency assessments has the potential to enable a  
88 more responsive and resilient health care system.

89         Phone surveys are one potential approach to enable high-frequency HFAs. Previous research  
90 indicates that phone-based data collection is an affordable approach to collecting both qualitative and  
91 quantitative data<sup>5</sup>. During the COVID-19 pandemic, more frequent, locally adapted phone survey  
92 methodologies were utilized to monitor essential health service use and the impact of the pandemic on  
93 health facility readiness to deliver services<sup>6,7</sup>. This demonstrated that phone-based survey approaches can  
94 provide frequent and cost-effective results in LMIC settings<sup>6</sup>.

95         Growing evidence indicates that interview administered phone surveys of households<sup>8-11</sup> and  
96 health workers<sup>12-14</sup> can have utility and validity; however, there is less evidence that phone-based health  
97 facility assessments can capture comprehensive details on health facility functioning with limited bias.  
98 Two studies offer promising results. In Malawi, Pattnaik et al. conducted direct observations and register  
99 reviews at facilities to assess select aspects of phone survey validity related to family planning indicators<sup>14</sup>.  
100 The study found that most health facility indicators collected by phone were above a 70% sensitivity  
101 threshold compared to in-person records and inspections and determined that phone data collection was  
102 both feasible and highly cost-effective<sup>14</sup>. In Kenya, Ashigbie et al. assessed the validity of phone-assessed  
103 medicine availability indicators at health facilities compared to in-person visits. The study found strong  
104 health facility-level agreement on medicine availability between the phone survey and subsequent in-  
105 person assessment ([kappa = 0.90; confidence interval (CI) 0.88-0.92]); further, the phone survey was

106 conducted with a unit cost of \$19.73 compared to \$186.20 for the in-person assessment<sup>15</sup>. Both studies  
107 concluded that phone-based surveys were valid and cost-effective compared to existing gold standards of  
108 either in-person data collection or inspection visits for a limited set of indicators<sup>14,15</sup>.

109 Despite promising initial results, a key concern for phone-based surveys is minimizing respondent  
110 fatigue by keeping the survey administration short. Like in-person surveys, phone-based interviews that  
111 take a longer period are more likely to lead to respondent fatigue, which can give rise to threats to validity.  
112 For example, recent household survey research quantified that a 15-minute addition in questionnaire  
113 administration time resulted in 8-to-17% less detailed response over the phone<sup>16</sup>. Similarly, Abate et al.  
114 (2023) randomly assigned household survey participants to in-person or phone-based survey  
115 administration for a household nutrition consumption survey<sup>17</sup>. They found that average per-capita  
116 consumption reported by respondents was 23% lower and the estimated poverty headcount was twice  
117 as high in the phone-based survey compared to the in-person survey due to underreporting of food items  
118 consumed, which the authors attributed to respondent fatigue<sup>17</sup>. These findings emphasize the  
119 importance of developing phone-specific questionnaires which should be shorter, simpler, and less  
120 cognitively burdensome to complete than typical in-person assessments<sup>11</sup>.

121 An additional limitation of phone-based data collection is that the information provided by  
122 respondents is generally self-reported and cannot be directly verified by enumeration using observation.  
123 This poses specific validity questions in health facility surveys which rely on in-person verification of  
124 medical equipment, drugs, health workers, registry logs, and physical infrastructure. Moreover, more  
125 complex indicators cannot be effectively administered over the phone. For example, while a phone-based  
126 survey might approximate in-person measurement for basic infrastructure availability or the number of  
127 staff assigned to a health facility, it is likely unsuitable for assessing more intricate constructs like provider  
128 performance and adherence to clinical protocols.

129 Finally, phone-based data collection is optimally conducted with a single respondent, while most  
130 large-scale, comprehensive HFAs rely on multiple respondents, often starting with the health facility  
131 manager but going on to include respondents across the health facility (e.g., service providers, pharmacist,  
132 financial administrator/accountant, etc.). This means that the phone-based survey requires a careful  
133 selection of the key questions or domains that a single respondent is best suited to answer.

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

### 143 **Study Objectives and Contributions**

144 This study has three primary objectives. The first objective is to assess the inter-method reliability of  
145 the rapid phone-based HFA compared to the comprehensive in-person HFA. This objective focuses on  
146 isolating the survey mode effect – the differences in results due to the phone and in-person administration  
147 – and will determine the extent to which phone data collection is a reliable alternative mode for  
148 conducting HFAs for a subset of key PHC service delivery indicators. To minimize bias, this objective will  
149 include only the survey questions that are identical in both the rapid phone-based HFA and the  
150 comprehensive in-person HFA. By keeping questions the same and varying only the administration mode,  
151 we can effectively isolate and examine the mode effect.

152 The second objective is to assess the concurrent criterion validity of indicators in the rapid phone-  
153 based HFA compared to indicators in the comprehensive in-person HFA. As discussed earlier, phone  
154 surveys require questions that are shorter, simpler, and less cognitively demanding. While this design is  
155 crucial for phone-based assessments, it can lead to issues such as questions being interpreted differently  
156 by respondents or lack of clarity or confusion on specific terms in the asked question that may affect data  
157 quality. This objective therefore focuses on evaluating the extent to which the rapid phone-based HFA  
158 can accurately measure the same underlying indicators as that in a comprehensive in-person HFA via a  
159 simplified questionnaire deemed suitable for phone-based administration.

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

168 Taken together, these objectives will demonstrate whether the rapid phone-based HFA can  
169 approximate the results of the comprehensive in-person HFA for a select subset of facility-level indicators.  
170 As phone-based data collection in LMICs is a rapidly evolving field with growing interest following the  
171 COVID-19 pandemic, publication of this protocol aims to inform the broader scientific community of this  
172 undertaking and improve the transparency of our subsequent analysis against the study design articulated  
173 below.

## 174 **MATERIALS AND METHODS**



175 **Study Setting**

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

189 The World Bank (WB) and the Global Financing Facility for Women, Children, and Adolescents  
190 (GFF) are supporting the efforts of the Ministry of Health and Social Protection of the Republic of Tajikistan  
191 to improve the quality of PHC service delivery. As part of this initiative, they are conducting a series of  
192 surveys to assess the quality of PHC in the country. Given Tajikistan's lack of comprehensive data on PHC  
193 service delivery, the WB and GFF are supporting the implementation of a comprehensive in-person HFA  
194 to get a better understanding of the country's PHC system, identify its strengths and weaknesses, and  
195 provide comprehensive data to guide policy decisions. In addition to the comprehensive in-person HFA,  
196 the World Bank and GFF are also implementing a phone-based HFA to be conducted every 6-12 months  
197 to monitor intervention implementation, inform course correction, and identify new challenges or shocks

198 that emerge over time. Conducting comprehensive in-person HFA and rapid phone-based HFA  
199 concurrently in the same setting therefore has offered a unique opportunity to conduct a head-to-head  
200 comparison of the results from the two surveys.

### 201 **Survey instruments**

202 The two surveys that will be conducted are: (1) the GFF's Frequent Assessment and Health  
203 Systems Tools for Resilience (FASTR) Initiative's rapid-cycle health facility phone survey which is designed  
204 for interview-administered phone-based data collection at the PHC level<sup>21</sup> and which will serve as the  
205 rapid phone-based HFA and (2) the World Bank's Service Delivery Indicators (SDI) health survey which will  
206 serve as the comprehensive in-person HFA<sup>4</sup>.

207 The FASTR rapid-cycle health facility phone survey (subsequently referred to as the rapid phone-  
208 based HFA) is a PHC assessment tool that aims to monitor service availability, readiness, and functioning  
209 of PHC facilities over time with an emphasis on reproductive, maternal, newborn, child, and adolescent  
210 health and nutrition services<sup>21</sup>. The tool's measurement approach is grounded in the WHO and UNICEF's  
211 Primary Health Care Measurement Framework and Indicators<sup>22</sup>. The questionnaire was drawn from  
212 existing comprehensive HFAs<sup>1-3</sup> as well as health facility phone survey experience during the COVID-19  
213 pandemic<sup>7,23</sup>. The survey is administered to a nationally or sub-nationally representative panel sample of  
214 PHC facilities over four quarterly data collection contacts per year. Quarterly data collection enables both  
215 the tracking of presumed higher-variability indicators over time (e.g., medicine stock, utility functioning)  
216 and the distribution of annual indicators across four survey rounds to shorten survey administration and  
217 reduce respondent fatigue. The phone survey also includes questions that capture respondent  
218 perceptions, such those related to challenges faced by the facility. These perception-based questions are  
219 a useful way to receive timely feedback from facilities about challenges that they face but are outside the

220 scope of this validation study. As of August 2024, the tool has been implemented in Burkina Faso, Vietnam,  
221 Senegal, Bangladesh, Tajikistan, and Madagascar.

222 The SDI (subsequently referred to as the comprehensive in-person HFA) health survey program  
223 began in 2008 and has been implemented in over a dozen countries across the globe<sup>24</sup>. The survey was  
224 revamped in 2019 to better align with the recent developments in the literature on measuring quality of  
225 PHC service delivery<sup>25,26</sup>. Building on the frameworks for high quality health systems, the survey now  
226 assesses processes of care and patient outcomes along with structural inputs to provide a comprehensive  
227 picture of PHC service delivery in a country. The SDI survey includes three questionnaires: health facility  
228 questionnaire, health care provider questionnaire, and outpatient exit interview. Only the health facility  
229 questionnaire will be included in the validation study. The SDI comprehensive in-person HFA, and its  
230 corresponding in-person verification of the availability and functionality of structural inputs, will be  
231 considered the gold standard for the purpose of this study.

### 232 **Questionnaire adaptation**

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

241 This preliminary contextualized questionnaire was shared with a joint working group comprising  
242 of local experts, practitioners, and government to capture Tajikistan's unique primary health care

243 context and to ensure appropriateness for different levels of Tajikistan’s health system. The surveys  
244 then underwent field testing at selected facilities to further improve the questionnaires based on  
245 respondent feedback and identify and address issues that emerge during the pilot. This involved  
246 improving option choices for specific questions, improving question framing such as adding clearer and  
247 easy to understand definitions of certain terms, and deleting questions that might not be relevant at the  
248 facility level.

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

#### 254 **Indicator mapping**

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

264 Following the mapping, a question-by-question revision activity was conducted to identify the two  
265 types of survey items mapped across both the questionnaires: (1) ‘direct’ survey items and (2) ‘indirect’

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

279 Survey items were classified as 'indirectly mapped' when they intended to measure similar  
280 underlying indicators in both surveys, but differed in the way they were asked or framed due to  
281 feasibility for phone administration. This typically occurred when comprehensive in-person HFA survey  
282 items were too lengthy or detailed to be administered over the phone, leading to simpler alternatives  
283 being included in the rapid phone-based HFA. For example, in measuring HIV/AIDS service provision, the  
284 rapid phone-based HFA asks whether the health facility provides HIV/AIDS services, while the  
285 comprehensive HFA asks whether the health facility was able to provide HIV/AIDS diagnosis services and  
286 HIV/AIDS treatment services in the 3 months prior to data collection. In this case, the underlying  
287 indicator "HIV/AIDS service availability" was common in both the surveys but was captured differently in  
288 each, leading to it being "indirectly mapped."

289 Since the "indirectly mapped" survey items differed in question framing, they were not suitable  
 290 for the inter-method reliability test (objective 1) to isolate the mode effect (phone v/s in-person). Any  
 291 difference in the survey results could be attributed to various factors, such as respondents interpreting  
 292 the questions differently, rather than solely due to the mode of administration. Therefore, these survey  
 293 items were suitable only for assessing objective two (validity).

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

298 Table 1. Number of indicators, their categorization, and relationship to study objectives

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

<b>All Areas</b>	<b>89</b>	<b>22 (25%)</b>	<b>67 (75%)</b>
------------------	-----------	-----------------	-----------------

299

300 **Data collection**

301 Data collection will take place from August to October 2024. This study was approved by the  
 302 Government of the Republic of Tajikistan Ministry of Health and Social Protection Biomedical Ethics  
 303 Committee, with the Order No. 60 and number N2148 on May 23<sup>rd</sup>, 2024. Informed consent will be  
 304 obtained from all survey respondents.

305 **Sampling**

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

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

		<b># Facilities</b>	<b># Selected</b>
	Dushanbe	15	15
Oblast	Sughd	616	151
	Khaktlon	1,123	150
	Gorno-Badakhshan Autonomous Region	227	84

	Districts of Republican Subordination	713	100
Type	City Health Center	41	41
	District Health Center	53	53
	Rural Health Center	895	214
	Health House	1,705	192
	<b>TOTAL</b>	<b>2,694</b>	<b>500</b>

314           The enumerators will obtain verbal consent for the rapid phone-based HFA and written consent  
315 for the comprehensive in-person HFA from the respondents before starting the survey. If a respondent  
316 declines to participate, the reason for their refusal will be documented, and the facility will be replaced  
317 by a facility from the same strata whenever available. In the event of a non-response, health facilities  
318 will be replaced from a "replacement list" curated from the sample frame for each survey, following the  
319 same protocol of random sampling selection. Furthermore, in the event of a non-response,  
320 replacements for the two surveys will be done independently. This means that if a facility does not  
321 participate in the rapid phone-based HFA and needs to be replaced with another health facility, it will  
322 not be simultaneously replaced in the comprehensive in-person HFA, and vice-versa, unless the non-  
323 participation is due to the closure of the facility. As a result, the final samples may differ slightly due to  
324 varying response rates.

325           In both surveys, the officer-in-charge of the health facility will serve as the primary respondent.  
326 This individual is best positioned to provide comprehensive insights into the facility's overall readiness.  
327 However, in situations where the designated officer-in-charge is unable to participate or believes  
328 another staff member would be better suited to answer specific sections or questions, the survey  
329 response could be delegated to a deputy or another qualified health worker.



330 Data collection procedures

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

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

343 The rapid phone-based HFA is designed as a short 30 to 45 minutes survey, requiring no  
344 preparation by the officer-in-charge. Respondents will have the option to either start answering the  
345 survey during the first call, or to schedule it for a more convenient day or time. If a respondent is  
346 unreachable, the rapid-cycle survey protocol requires at least five contact attempts on different days  
347 and at varying times to complete the interview, maximizing communication success and reducing  
348 potential biases introduced by timing constraints. This will be particularly important for remote facilities  
349 with low phone connectivity, where the respondent may have sporadic access to a stable phone  
350 network. In instances where immediate completion of the interview is not feasible due to  
351 telecommunications issues or respondent availability, follow-up calls may be scheduled for a later date  
352 or time.

353 For the comprehensive in-person HFA, enumerator teams will contact the officers-in-charge of  
354 sampled facilities two weeks before data collection, to inform them about the survey, coordinate the  
355 date of the first visit, and share the ‘advance questionnaire’. The advance questionnaire is a subset of  
356 the comprehensive in-person HFA and comprises of questions that require reference to facility data  
357 records. This will be shared with the officer-in-charges at least two weeks prior to data collection day,  
358 allowing respondents to prepare responses, thereby saving time and improving data quality on the day  
359 of data collection.

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

### 371 Timing of survey administration

372 The comprehensive in-person HFA and the rapid phone-based HFA will be conducted as close in  
373 timing as possible, ideally within a two-week period for each health facility. The close timing of the  
374 surveys minimizes the risk of external factors (for example, a health facility receiving new medical  
375 supplies) impacting the data recorded in each survey and will ensure a more reliable comparison.

376 However, there is a risk that conducting both surveys too close will introduce respondent fatigue that  
377 might affect results. Therefore, the order of the surveys for each health facility will be randomized to  
378 control for some of the real changes that may occur due to the order, with 50% of facilities receiving the  
379 rapid phone-based HFA first and 50% of the facilities receiving the comprehensive in-person survey first.  
380 This will allow for a more robust analysis. It will also allow us to conduct a longitudinal analysis of  
381 responses for the rapid phone-based HFA and the comprehensive in-person HFA in each facility to  
382 identify possible confounding effects depending on which survey the facility receives first. We would  
383 expect this to be a possibility especially in facilities that receive the in-person survey first, as respondent  
384 recall may be primed by the recent in-person verification of survey items and therefore, respondents  
385 may answer the rapid phone-based HFA with greater validity.

#### 386 **Data analysis**

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

#### 389 **Objective One: Inter-Reliability Analysis**

390 The mode effect (phone v. in-person) of the rapid phone-based HFA compared to the  
391 comprehensive in-person HFA will be analyzed using the “directly mapped” survey items. The results will  
392 be analyzed and reported in three ways: as a simple percent agreement of responses, using the Cohen’s  
393 Kappa to account for the percent of agreement that is expected by chance, and using a prevalence and  
394 bias adjusted kappa (PABAK) to account for prevalence and bias in the sample. Calculating PABAK is  
395 especially important for a HFA, where many of the survey items are expected to be available and  
396 therefore, the expected agreement by chance for many indicators is high due to high prevalence<sup>27</sup>.  
397 Taken together, these three measures will provide a nuanced understanding of the results<sup>28</sup>.

398           There are some differences on what is considered adequate for reliability. While literature  
399 suggests that a Kappa coefficient larger than 0.61 would be reliable, it is acknowledged that these  
400 thresholds can be somewhat arbitrary. To address this, a scale for the level of agreement<sup>29</sup> will be used:  
401 almost perfect agreement above 0.90, strong 0.80 to 0.90, moderate 0.60 to 0.79, weak 0.40 to 0.59,  
402 and minimal 0.20 to 0.39.

#### 403 Objective Two: Concurrent Criterion Validity Analysis

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

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

419           Regression analysis will also be conducted to identify any heterogeneity effects due to facility  
420 characteristics. Subgroup analysis will be conducted based on (1) oblast, (2) rural versus urban, and (3)

421 facility type to identify if any underlying, facility characteristics influence the reliability and validity of the  
422 rapid, phone-based HFA results. This will identify whether validity is different in particular sub-groups.  
423 For example, the rapid, phone-based HFA may have improved validity in smaller facilities where one  
424 respondent is likely to be familiar with all the tracer items and could be more challenging in a larger  
425 facility with multiple departments.

#### 426 Objective Three: Non-Response Rate

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

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

## 442 **DISCUSSION**

443 To the best of our knowledge, this study is the first to assess the reliability and validity of a rapid  
444 phone-based HFA against a comprehensive in-person HFA. This will fill an important gap in global evidence  
445 on the reliability and validity of phone-based approaches to administer HFAs to monitor a subset of key  
446 facility-level indicators. This insight is especially critical given the rapid rise of phone-based survey  
447 approaches in LMICs following the COVID-19 pandemic<sup>6</sup> as an innovative method for more frequent,  
448 lower-cost, facility-level, primary data collection for select indicators.

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

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

## 467 **CONCLUSION**

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

## 478 **ACKNOWLEDGEMENTS**

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

487           Funding for this study was received from the Global Financing Facility for Women, Children, and  
488 Adolescents and the World Bank Group. The findings, interpretations, and conclusions expressed in this

489 paper are entirely those of the authors and do not necessarily represent the views of the World Bank, its  
490 executive directors, and the governments of the countries they represent.

#### 491 **AUTHOR CONTRIBUTIONS**

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

#### 497 **REFERENCES**

- 498 1. The World Health Organization. Harmonized Health Facility Assessment (HHFA). Published  
499 2023. Accessed November 6, 2023. [https://www.who.int/data/data-collection-](https://www.who.int/data/data-collection-tools/harmonized-health-facility-assessment/introduction)  
500 [tools/harmonized-health-facility-assessment/introduction](https://www.who.int/data/data-collection-tools/harmonized-health-facility-assessment/introduction)
- 501 2. The World Health Organization. Service availability and readiness assessment (SARA).  
502 Published 2023. Accessed November 6, 2023. [https://www.who.int/data/data-collection-](https://www.who.int/data/data-collection-tools/service-availability-and-readiness-assessment-(sara))  
503 [tools/service-availability-and-readiness-assessment-\(sara\)](https://www.who.int/data/data-collection-tools/service-availability-and-readiness-assessment-(sara))
- 504 3. The DHS Program - Service Provision Assessments (SPA). The DHS Program. Published  
505 November 7, 2023. Accessed November 6, 2023.  
506 <https://dhsprogram.com/methodology/Survey-Types/SPA.cfm>



- 507 4. The World Bank. Health Service Delivery Indicators. World Bank. Published 2023. Accessed  
508 November 6, 2023. [https://www.worldbank.org/en/programs/service-delivery-](https://www.worldbank.org/en/programs/service-delivery-indicators/health)  
509 [indicators/health](https://www.worldbank.org/en/programs/service-delivery-indicators/health)
- 510 5. Burnard P. The telephone interview as a data collection method. *Nurse Education Today*.  
511 1994;14(1):67-72. doi:10.1016/0260-6917(94)90060-4
- 512 6. Arita S, Ba MF, Traoré Z, Bonnet E, Faye A, Ridde V. Use of interviewer-administered  
513 telephone surveys during infectious disease outbreaks, epidemics and pandemics: a scoping  
514 review. *BMJ Glob Health*. 2023;8(5):e011109. doi:10.1136/bmjgh-2022-011109
- 515 7. Peters MA, Ahmed T, Azais V, et al. Resilience of front-line facilities during COVID-19:  
516 evidence from cross-sectional rapid surveys in eight low- and middle-income countries.  
517 *Health Policy and Planning*. 2023;38(7):789-798. doi:10.1093/heapol/czad032
- 518 8. Pariyo GW, Greenleaf AR, Gibson DG, et al. Does mobile phone survey method matter?  
519 Reliability of computer-assisted telephone interviews and interactive voice response non-  
520 communicable diseases risk factor surveys in low and middle income countries. *PLoS ONE*.  
521 2019;14(4):e0214450. doi:10.1371/journal.pone.0214450
- 522 9. Kakietek JJ, Eberwein JD, Stacey N, Newhouse D, Yoshida N. Foregone healthcare during the  
523 COVID-19 pandemic: early survey estimates from 39 low- and middle-income countries.  
524 *Health Policy and Planning*. 2022;37(6):771-778. doi:10.1093/heapol/czac024

- 525 10. Etang A, Himelein K. Monitoring the Ebola Crisis Using Mobile Phone Surveys. In:  
526 Hoogeveen J, Pape U, eds. *Data Collection in Fragile States*. Springer International  
527 Publishing; 2020:15-31. doi:10.1007/978-3-030-25120-8\_2
- 528 11. Gourlay S, Kilic T, Martuscelli A, Wollburg P, Zezza A. Viewpoint: High-frequency phone  
529 surveys on COVID-19: Good practices, open questions. *Food Policy*. 2021;105:102153.  
530 doi:10.1016/j.foodpol.2021.102153
- 531 12. Chen X, Mohan D, Maïga A, et al. Validity of using mobile phone surveys to evaluate  
532 community health worker program in Mali. *BMC Med Res Methodol*. 2021;21:115.  
533 doi:10.1186/s12874-021-01317-7
- 534 13. Hazel E, Amouzou A, Park L, et al. Real-Time Assessments of the Strength of Program  
535 Implementation for Community Case Management of Childhood Illness: Validation of a  
536 Mobile Phone-Based Method in Malawi. *The American Journal of Tropical Medicine and*  
537 *Hygiene*. 2015;92(3):660-665. doi:10.4269/ajtmh.14-0396
- 538 14. Pattnaik A, Mohan D, Chipokosa S, et al. Testing the validity and feasibility of using a mobile  
539 phone-based method to assess the strength of implementation of family planning programs  
540 in Malawi. *BMC Health Serv Res*. 2020;20:221. doi:10.1186/s12913-020-5066-1
- 541 15. Ashigbie PG, Rockers PC, Laing RO, et al. Phone-based monitoring to evaluate health policy  
542 and program implementation in Kenya. *Health Policy and Planning*. 2021;36(4):444-453.  
543 doi:10.1093/heapol/czab029

- 544 16. Abay KA, Berhane G, Hoddinott J, Hirrfot KT. Assessing Response Fatigue in Phone Surveys :  
545 Experimental Evidence on Dietary Diversity in Ethiopia. *Policy Research Working Paper*  
546 *Series*. 2021;9636. Accessed November 6, 2023.  
547 <https://ideas.repec.org//p/wbk/wbrwps/9636.html>
- 548 17. Abate GT, De Brauw A, Hirvonen K, Wolle A. Measuring consumption over the phone:  
549 Evidence from a survey experiment in urban Ethiopia. *Journal of Development Economics*.  
550 2023;161:103026. doi:10.1016/j.jdeveco.2022.103026
- 551 18. *Tajikistan*. The World Bank Group  
552 [https://documents1.worldbank.org/curated/en/470931511944745629/pdf/121732-BRI-](https://documents1.worldbank.org/curated/en/470931511944745629/pdf/121732-BRI-P154478-PUBLIC-Tajikistan-Snapshot-Print.pdf)  
553 [P154478-PUBLIC-Tajikistan-Snapshot-Print.pdf](https://documents1.worldbank.org/curated/en/470931511944745629/pdf/121732-BRI-P154478-PUBLIC-Tajikistan-Snapshot-Print.pdf)
- 554 19. World Population Dashboard -Tajikistan. United Nations Population Fund. Published 2023.  
555 Accessed November 6, 2023. <https://www.unfpa.org/data/world-population/TJ>
- 556 20. The World Health Organization European Region, The European Observatory on Health  
557 Systems, and Policies. *Tajikistan*. World Health Organization; 2022.
- 558 21. The Global Financing Facility and the World Bank Group. *FASTR: The GFF's Approach to*  
559 *Rapid Cycle Analytics and Data Use*. Accessed June 29, 2024. [https://data.gffportal.org/key-](https://data.gffportal.org/key-theme/rapid-cycle-analytics-and-data-use)  
560 [theme/rapid-cycle-analytics-and-data-use](https://data.gffportal.org/key-theme/rapid-cycle-analytics-and-data-use)
- 561 22. *Primary Health Care Measurement Framework and Indicators: Monitoring Health Systems*  
562 *through a Primary Health Care Lens*. The World Health Organization and the United Nations  
563 Children's Fund (UNICEF); 2022. <https://www.who.int/publications/i/item/9789240044210>

- 564 23. Continuity of essential health services: Facility assessment tool. World Health Organization.  
565 Published May 12, 2021. Accessed April 1, 2024. [https://www.who.int/publications-detail-](https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-HCF-assessment-EHS-2021.1)  
566 [redirect/WHO-2019-nCoV-HCF-assessment-EHS-2021.1](https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-HCF-assessment-EHS-2021.1)
- 567 24. Andrews, Kathryn, Conner, Ruben, Gatti, Roberta, Sharma, Jigyasa. *The Realities of Primary*  
568 *Care Variation in Quality of Care Across Nine Countries in Sub-Saharan Africa*. World Bank  
569 Group Human Development Global Practice; 2021:1-44.
- 570 25. The Primary Health Care Performance Initiative. *The PHCPI Conceptual Framework*.; 2022.  
571 Accessed August 2, 2024. <https://www.improvingphc.org/phcpi-conceptual-framework>
- 572 26. Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the Sustainable  
573 Development Goals era: time for a revolution. *The Lancet Global Health*. 2018;6(11):e1196-  
574 e1252. doi:10.1016/S2214-109X(18)30386-3
- 575 27. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *Journal of Clinical Epidemiology*.  
576 1993;46(5):423-429. doi:10.1016/0895-4356(93)90018-V
- 577 28. Cohen, Jacob. A COEFFICIENT OF AGREEMENT FOR NOMINAL SCALES. *Educational and*  
578 *Psychological Measurement*. 1960;XX(1):37-46.
- 579 29. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia Medica*. 2012;22(3):276-  
580 282.

