Electronic supplementary material:
The online version of this article contains supplementary material.

journal of global health

© 2026 The Author(s)

Cite as: Luoma J, Haidara FC, Samake O, Adubra L, Abdou M, Alber D, Ashorn U, Cloutman-Green E, Diallo M, Juma J, Ducker C, Fan Y, Gruffud G, Hallamaa L, Ihamuotila R, Martell O, Okello C, Haapaniemi T, Vehmasto T, Murugesan A, Traore CAT, Ylikruuvi K, Klein N, Ashorn P, Sow S, Cheung YB. Development and evaluation of a subjective measurement of geographical hardness-to-reach: a survey of data collectors in a large-scale cluster randomised trial of azithromycin mass drug administration. J Glob Health. 2026;16:04322.

Development and evaluation of a subjective measurement of geographical hardness-to-reach: a survey of data collectors in a large-scale cluster randomised trial of azithromycin mass drug administration

Juho Luoma[1]*, Fadima Cheick Haidara[2]*, Oumar Samake[3], Laura Adubra[1], Mahamadou Abdou[3], Dagmar Alber[4], Ulla Ashorn[1], Elaine Cloutman-Green[5], Mamadou Diallo[2], Jane Juma[2], Camilla Ducker[6], Yue-Mei Fan[1], Gwydion Gruffudd[6], Lotta Hallamaa[1], Rikhard Ihamuotila[1], Owen Martell[6], Collins Okello[2], Tiia Haapaniemi[1], Taru Vehmasto[1], Akshaya Murugesan[1], Cheick Amadou Tidiane Traore[3], Kaisa Ylikruuvi[1], Nigel Klein[4]†, Per Ashorn[1], Samba Sow[2]‡, Yin Bun Cheung[1]‡

[1]Center for Child, Adolescent, and Maternal Health Research, Faculty of Medicine and Health Technology, Tampere University and Tampere University Hospital, Tampere, Finland

[2]Center for Vaccine Development–Mali, Bamako, Mali

[3]Ministère de la Santé et du Développement Social du Mali, Bamako, Mali

[4]Great Ormond Street Institute of Child Health, University College London, UK

[5]Great Ormond Street Hospital, London, UK

[6]Tro Da Ltd, Caerphilly, Wales, UK

*Joint first authorship.

†Deceased.

‡Joint senior authorship.

Correspondence to:
Juho Luoma
Tampere University
Medisiinarinkatu 3, 33520 Tampere
Finland
juho.luoma@tuni.fi

Background Distance from residence to the nearest health facility is a known predictor of child mortality. However, physical distance alone may not fully capture geographical hardness-to-reach. Subjective assessments from data collectors may offer additional insight into village-level accessibility and its relationship to health outcomes.

Methods This was a sub-study of Large-scale Assessment of the Key health-promoting Activities of two New Mass Drug Administration regimens with Azithromycin, a large-scale cluster-randomised trial of mass administration of azithromycin in Mali. We developed a 14-item survey and administered it to 25 data collectors to evaluate village geographical accessibility. We recorded survey completion time and respondent burden assessments to investigate the measurement's feasibility. We assessed convergent validity by correlation between two composite scores derived from survey items and objective measures (distance and travel time to the nearest health facility, the Malian national outreach programme status, and the average monthly rainfall) and predictive validity using under-five mortality as the criterion. Inter-rater reliability was evaluated by intra-cluster correlation coefficients (ICC).

Results The mean time of completing the survey was three minutes. Survey composite scores reflecting subjective village inaccessibility were associated with higher under-five mortality in the villages: one standard deviation change towards harder access associated with up to 8% higher under-five mortality rate, even after adjusting for distance to the nearest health centre. Internal consistency between the items ranged from moderate to high, with polychoric correlation coefficients ranging from 0.49 (items 'How long does it take for you to get to this village on rainy season' and 'Condition of the roads on non-rainy season') to 0.94 (items 'How hard this village is to reach' and 'How difficult it is for the residents to seek medical care'). Inter-rater reliability for all-item composite score and two-item composite score, was moderate to strong, with the average ICCs being 0.74 and 0.69, respectively). The composite scores showed expected associations with objective accessibility indicators,

with Pearson correlation coefficients between all-item composite score and distance to the nearest health facility, average monthly rainfall, and travel time by car being, 0.36 ($P<0.001$), 0.22 ($P<0.001$), and 0.35 ($P<0.001$), respectively. Pearson correlation coefficients between two-item composite score and distance to the nearest health facility, average monthly rainfall, and travel time by car were 0.34 ($P<0.001$), 0.22 ($P<0.001$), and 0.32 ($P<0.001$), respectively.

Conclusions Our findings indicate acceptable feasibility and reliability, with moderate convergent and predictive validity. This simple survey can serve as a useful tool for evaluating hard-to-reach areas and guiding resource allocation in public health interventions, including mass drug administration campaigns.

Registration clinicaltrials.gov: NCT04424511 (parent trial)

Keywords hard-to-reach; accessibility; survey; health equity

Access to healthcare remains a key determinant of child survival and well-being, especially in low- and middle-income countries (LMICs). In this context, greater distance from households to health facilities is known to be associated with higher under-five mortality, lower immunisation coverage, and delayed care-seeking for acute illnesses [1–3]. Recent findings indicate that geographical accessibility can also modify the impact of public health interventions. For instance, the effect of azithromycin mass drug administration on child mortality has been stronger in more remote areas [4]. Bed net ownership was also characterised by vast geographic disparities, highlighting the negative correlation between distance to primary national roads and distance to the nearest health facility [5].

However, distance alone cannot fully explain the complexity of healthcare accessibility. Environmental conditions, such as seasonal flooding, limited transportation, and poor infrastructure, together with social and economic constraints, may critically impede access even for communities relatively close to health facilities [6–9]. Communities with limited access to healthcare can also be more prone to having a higher disease burden, as they face overlapping issues related to poor socio-economic and infrastructural conditions [10–13]. The concept of 'hard-to-reach' populations has, therefore, gained prominence, yet remains inconsistently defined and measured [6]. Existing categorisations often rely on spatial data or modelled estimates of remoteness, which are sometimes incomplete or outdated, and may not reflect local realities [14–16]. The *Stratégie Avancée* programme in Mali, for example, classifies villages by distance to health facilities, yet this approach rarely captures the broader contextual factors that make service delivery genuinely difficult [17].

To address these limitations, we developed a simple, locally informed tool to assess geographical hardness-to-reach. Unlike purely distance-based or computational measures, this tool draws on local perceptions of accessibility, allowing for a more context-specific understanding of how geography and infrastructure affect service delivery. Conceptually, 'hardness-to-reach' affects access to healthcare and service availability, which in turn affect health outcomes, such as mortality. Remoteness and infrastructure describe structural characteristics of a location, while accessibility reflects the ease of reaching services, and deprivation captures underlying socio-economic disadvantage. Hardness-to-reach thus refers to the combined and context-dependent manifestation of these factors, as experienced in practice. While public health has traditionally focused on tangible outcomes such as mortality, there is growing recognition of the value of subjective measures, such as quality of life and pain [18], as complements to objective indicators [19]. In this context, we sought to propose subjective measurement of hardness-to-reach as a meaningful and informative construct for public health. By quantifying perceived hardness-to-reach, the tool complements existing spatial methods and provides a standardised measure that can inform research and implementation planning. This approach may improve the identification of areas where additional logistical support is needed and contribute to more equitable health interventions in LMICs.

METHODS

We developed the hardness-to-reach tool (Table S1 in the **Online Supplementary Document**) within this sub-study of Large-scale Assessment of the Key health-promoting Activities of two New Mass Drug Administration regimens with Azithromycin (LAKANA) trial, a multi-year study covering 1,151 villages in Mali [20], as a simple questionnaire to be implemented by frontline data collectors (DCs) in capturing key determinants of hardness-to-reach. The 14 questionnaire items were generated by consulting two field team leaders on the reoccurring difficulties experienced during the trial, such as that accessibility varied by season, travel time and transportation posed challenges, and DCs encountered additional unexpected barriers when travelling to the villages during mass drug administration visits. Items were grouped into four thematic domains: methods of transportation, topographical difficulty, travel time, and overall subjective judgment. The lead investigator drafted the initial tool in English and refined it with a native French-speaking investigator. The draft was then reviewed by field leaders and pilot-tested with five DCs to ensure functionality, clarity in French, and consistency with prior knowledge of the locations.

Survey items are answered using either a dichotomous (yes/no) format or a five-point Likert scale, except for one item that requires specifying the mode of transportation used to reach the village when it was not accessed by car. The five-point scale items were originally coded from 1 (difficult) to 5 (easy) and subsequently reverse-coded in the analysis so that higher scores reflected greater 'hard-to-reach' status. Data were captured electronically using the CommCare system (Dimagi, Cambridge, Massachusetts, USA).

Study population and sample size considerations

We aimed to create a measure that represents a subjective assessment of geographical accessibility of all 1,151 villages included in the trial. We carried out survey rounds out at the end of 2024, coinciding with the conclusion of the trial. The survey was self-administered using data collection tablets given to respondents by the supervisors of field activities. The participants were limited to completing the survey for up to 20 villages at a time, with the option to restart the electronic questionnaire without losing previous evaluations.

To ensure geographic coverage across all 11 districts in which LAKANA operated (Selingue, Sefeto, Sagabari, Koulikoro, Kati, Kita, Kéniéba, Ouélésséboug ou, Kangaba, Kalabancoro, and Fana), 25 DCs with experience in one or more of the districts were invited, and all agreed to participate. A brief information sheet for informed consent was provided for the DCs before they were given the questionnaire. However, to preserve confidentiality, the names or positions of the DCs were not recorded in the electronic system and informed consent was obtained verbally. The DCs were not assigned to specific villages to assess, but were rather allowed to select the villages themselves. This helped ground the assessments in direct field experience and, alongside administration through multiple rounds, minimised the respondent burden. After evaluating the villages of their choosing, DCs were asked to assess how laborious the survey process felt.

Sample size calculation

Assuming concurrent validity in terms of Pearson correlation coefficient (r) between the tool score and validity criteria variables was approximately 0.5, rating at least 220 villages would offer a precision level (width of 95% confidence interval (CI)) of ±0.1 for the estimates. This sample size also offered precision level of ±0.05 for inter-rater reliability estimate, *i.e.* the intraclass correlation coefficient (ICC), assuming it to be approximately 0.8.

Descriptive statistics

We calculated the mean and standard deviation (SD) values for each item and composite scores and evaluated the potential floor and ceiling effects, *i.e.* do survey responses on items accumulate on the highest or lowest value.

Feasibility

To assess respondent burden, we captured the completion time per village from the digital survey application and asked DCs to rate the burden by choosing one of three options: 'The questionnaire was laborious', 'The questionnaire was moderate', and 'The questionnaire was light'.

Predictive and convergent validity

We used the under-five mortality from LAKANA trial as the outcome for assessing the predictive validity of the tool, as it reflects the cumulative effects of a complex system of health-related factors, and as it is a policy-relevant endpoint that directly aligns with the end goal of saving lives. While individual factors such as treatment seeking behaviour or facility attendance may provide additional information, they were not available for the study area. In LAKANA, DCs visited villages in eight three-month cycles and recorded the vital status of all household members at each visit. For this analysis, we aggregated the sums of deaths and person-time of children under the age of five years per village.

We calculated two composite scores: one including all survey items and a reduced score using only two items – the estimate of village residents' access to healthcare and the overall estimate of hardness-to-reach. We selected those two items because they were generic and applicable irrespective of settings and users. To test if this minimal item set could approximate the full set, we inspected the Pearson's r between the two-item composite score and all-items composite score, as well as between the two-item composite score and the sum of the remaining items.

We selected objective accessibility indicators for evaluating convergent validity based on existing literature [5,6] and observations from the LAKANA trial, such as heavy rains, deteriorated road or terrain conditions, or long travel distances. The selected indicators were the great circle, or 'as-the-crow-flies'; distance from the nearest health facility in kilometres; national outreach programme status (*Stratégie Avancée*, standard/advanced); monthly rainfall during the LAKANA trial time frame (2020–2024) in mean mm in month; and travel time to the nearest health facility by car in minutes. We calculated the as-the-crow-flies distances by using aggregated GPS-coordinates of households participating in LAKANA trial and the Haversine method [21] to estimate the distance between a village and the closest health centre. We retrieved rainfall information *via* Open-Meteo [22] and travel times by car *via* OpenStreetMap [23].

We used Poisson mixed-effects models with random intercepts for DCs to assess the association between under-five mortality and both the items and the composite scores, using village-level person-years as an offset variable. To assess the predictive value of the survey responses independent of geographic proximity, additional analyses adjusted for distance to the nearest health facility. We used linear mixed effects regression models with the objective accessibility indicators as dependent variables and the survey responses as independent variables with random intercepts for DCs to evaluate convergent validity. Variables in the model were standardised to have zero-mean and SD of one such that the coefficients have an interpretation of Pearson's r. To evaluate known-group validity, we compared the composite scores between the outreach strategy status using boxplots. We also established a mixed effects model with random intercept for DCs and a logistic link function to evaluate how one-unit change in the composite score relates to the odds of being in the advanced group.

Internal consistency and inter-rater reliability

We used polychoric correlation to assess relations between survey items, separately for villages that are reachable by vehicle or not, as some items are conditional on that. For inter-rater reliability, we considered only villages assessed by more than one DC, as we wanted to evaluate it distinctly from the performance of the measurement scale in practice. We computed ICCs for each individual Likert scale item on the survey tool, as well as for the composite scores. Because multiple DCs could assess the same village, and because not all 25 DCs evaluated every village, our study had an unbalanced, incomplete observational design for an inter-rater reliability study.

Furthermore, we predicted we would have two systematic sources of variance: villages and DCs. Therefore, we used two metrics to evaluate inter-rater reliability: individual ICC and average ICC, calculated from a two-way random-effects model [24–26]. The individual ICC has the interpretation of a single, 'typical' DC providing the ratings whereas the average ICC has the interpretation of taking the average rating of multiple ratings on a single village. A commonly used cut-off for acceptable inter-rater reliability ranges between 0.40 and 0.50 [27,28]. Here, we use 0.50 as a cut-off. Technical details of the inter-rater reliability estimation models are described elsewhere (**Online Supplementary Document**).

We carried out all statistical analyses in *R*, version 4.3.2 (R Foundation for Statistical Computing, Vienna, Austria). We report our findings per the STROBE reporting guidelines [29].

RESULTS

Feasibility

There were 1,697 assessments covering 1,142 villages, of which 388 were rated by at least two DCs. The mean number of assessments per village was 1.49; villages with only one assessment and those with multiple assessments were spatially evenly distributed (Figure S1 and Table S2 in the **Online Supplementary Document**). A single DC assessed approximately 68 villages, spreading through 4–6 survey rounds of about 15 village assessments. Completing the all-item survey form for a single village took approximately 3 minutes (range = 1.5–7). When DCs assessed fewer than 10 villages in a round, respondent burden was rated as light, moderate, and laborious in 29 (60%), 9 (19%), and 10 (21%) response rounds, respectively. In rounds with 10 or more villages, respondent burden was rated as light, moderate, and laborious in 40 (43%), 14 (15%), and 39 (42%) response rounds, respectively.

Descriptive statistics

The mean values of the all-items and two-item composite scores were 22.4 (SD = 6.4) and 6.2 (SD = 1.9), respectively (**Table 1**). In the thematic domain of subjective judgement, the items of access to healthcare, overall estimate, and equipment transportation difficulty in rainy and non-rainy seasons, had mean values of 3.1 (SD = 1.0), 3.1 (SD = 1.0), 3.4 (SD = 1.1), and 2.8 (SD = 1.0),

Table 1. Descriptive statistics of the survey scores on 1,697 village assessments

	Thematic domain	Possible range	x̄ (SD)	% at max	% at min
Composite score, all items	Composite score	8–38*	22.4 (6.4)	2.0	2.7
Composite score, two items	Composite score	2–10†	6.2 (1.9)	2.0	2.7
Estimate of village residents' access to healthcare	Subjective judgement	1–5	3.1 (1.0)	8.5	5.1
Overall estimate of hardness to reach	Subjective judgement	1–5	3.1 (1.0)	6.7	4.8
How difficult is it to transport equipment to the village in the rainy season?	Subjective judgement	1–5	3.4 (1.1)	17.4	3.3
How difficult is it to transport equipment to the village in the non-rainy season?	Subjective judgement	1–5	2.8 (1.0)	2.9	10.0
Difficulty of the terrain on non-rainy season	Topographical difficulty	1–5	4.1 (0.8)	32.7	1.8
Difficulty of the terrain on rainy season	Topographical difficulty	1–5	4.6 (0.5)	56.3	0.0
Condition of the roads on rainy season	Topographical difficulty	1–5	3.4 (1.1)	16.7	3.4
Condition of the roads in the non-rainy season	Topographical difficulty	1–5	2.7 (0.9)	1.3	10.5
Estimate of travel time to the village on rainy season	Travel time	1–5	1.7 (0.7)	0.2	44.3
Estimate of travel time to the village on non-rainy season	Travel time	1–5	1.3 (0.6)	0.2	69.8
Can the village be reached by large, motorised vehicle?	Method of transportation	No = 1, yes = 0	96.8 (17.6)‡	3.2	96.8
Are there any public transportation possibilities for the residents?	Method of transportation	No = 1, yes = 0	79.1 (40.7)‡	79.1	20.9

SD – standard deviation, x̄ – mean
*8 = Very easy to reach, 38 = very difficult to reach.
†2 = very easy to reach, 10 = very difficult to reach.
‡Expressed in %.

respectively. Regarding clustering of values on either on the lowest or the highest, there was little floor or ceiling effect, except that 17.4% of the ratings on the item on equipment transportation difficulty in rainy season were at the ceiling.

Approximately 3.2% of the responses indicated that the village was not reachable by car (alternative reported methods of transportation for reaching the village are listed in Table S3 in the **Online Supplementary Document**), while 79.1% indicated that there are no public transportation possibilities for the residents. Within the thematic domain of topographical difficulty, the mean score of was 3.4 (SD = 1.1) for the item 'Condition of the roads on rainy season' and 2.7 (SD = 0.9) for the item 'Condition of the roads on non-rainy season'. There was ceiling effect in the items, whereby 16.7% and 13.0% of the ratings per each item, respectively, were at the ceiling.

The mean estimated travel time between the village and health centre was 2–3 (SD = 1) hours during the rainy season and 1–2 (SD = 1) hours during non-rainy season. There was floor effect regarding the travel time, with 44.3% and 69.8%, respectively, being at the floor, which indicated travel time being less than one hour.

Internal consistency

Polychoric correlation coefficients among the items were ≥0.49. The highest correlation (0.94) was observed between 'access to healthcare' and 'overall estimate of hardness-to-reach', while the lowest (0.49) was between 'travel time estimate on rainy season' and 'road condition estimate during non-rainy season' (**Figure 1**; Figure S2 in the **Online Supplementary Document**). The two-item composite score showed a correlation of 0.94 with the all-item composite score and 0.89 with the sum of remaining items.

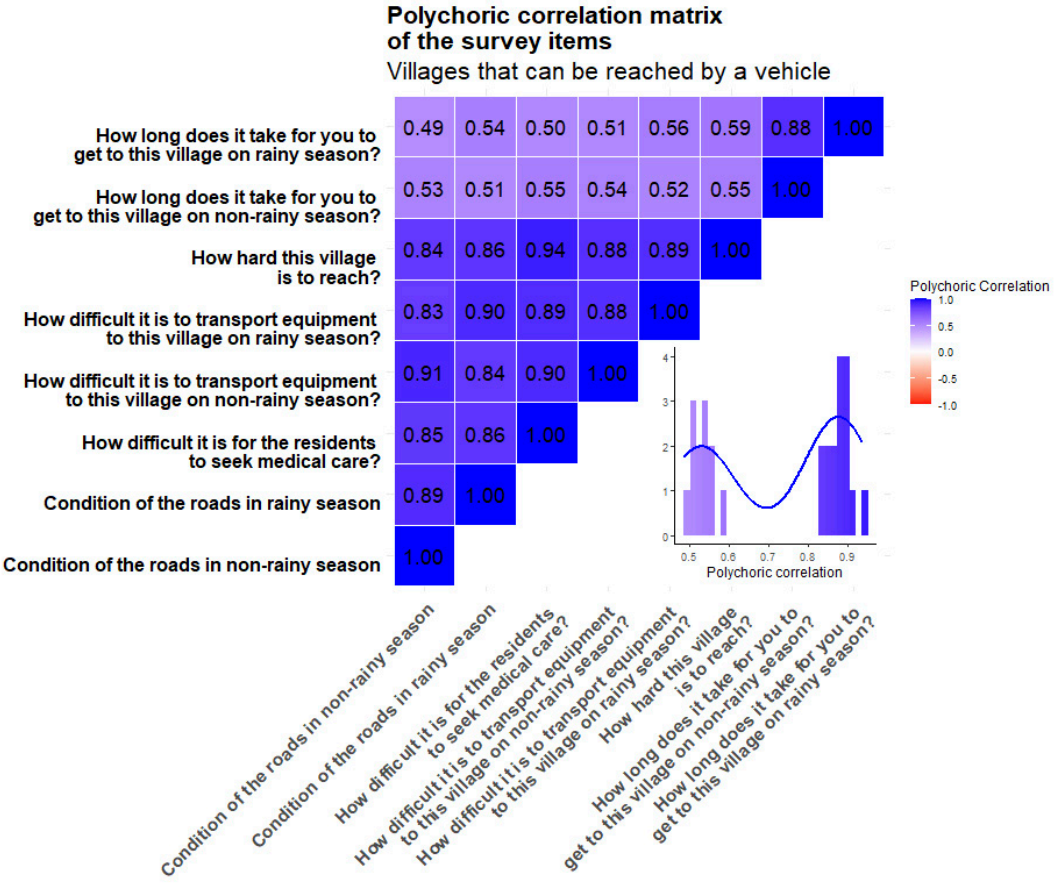


Figure 1. Polychoric correlation matrix of the five-point survey items for villages that can be reached by a vehicle.

Predictive, known group, and convergent validity

The mixed-effects model yielded an unadjusted risk ratio (RR) of 1.17 (95% CI = 1.12–1.21, $P < 0.001$) for the association between the all-items composite score and under-five mortality (Table 2). After adjustment for distance to nearest health facility, the adjusted RR was 1.08 (95% CI = 1.03–1.13, $P < 0.001$). For the two-item composite score, the unadjusted RR was 1.16 (95% CI = 1.11–1.20, $P = 0.001$) and the distance-adjusted RR was 1.07 (95% CI = 1.03–1.12, $P = 0.001$).

Table 2. Association between under-five mortality rate of villages and two versions of survey composite score (higher score means harder to reach) and distance to nearest health facility

Model*	Independent variables	Unadjusted RR per SD (95% CI)	Adjusted RR per SD (95% CI)	Adjusted *P*-value
Model 1	All-item composite score	1.17 (1.12–1.21)	1.08 (1.03–1.13)	<0.001
	Distance to health facility†		1.19 (1.14–1.25)	<0.001
Model 2	Two-item composite score	1.16 (1.11–1.20)	1.07 (1.03–1.12)	0.001
	Distance to health facility†		1.20 (1.15–1.25)	<0.001

aRR – adjusted risk ratio, CI – confidence interval, RR – risk ratio, SD – standard deviation
*Models 1 and 2 indicate either all-item or two-item variable being the main independent variable.
†SD = 4.7.

The 'advanced' category in outreach strategy had a mean all-items composite score of 23.7, while the 'standard' category had a mean composite score of 19.0 (Figure 2). Furthermore, an increase in the composite score was associated with an increase in the odds of belonging to the advanced group; the odds ratio per SD increase in the composite score was 2.32 (95% CI = 2.03–2.65, $P < 0.001$). The means of two-item composite score in the advanced and standard categories were 6.5 and 5.3, respectively, and the odds ratio for one SD increase in the two-item score was 2.05 (95% CI = 1.81–2.35, $P < 0.001$).

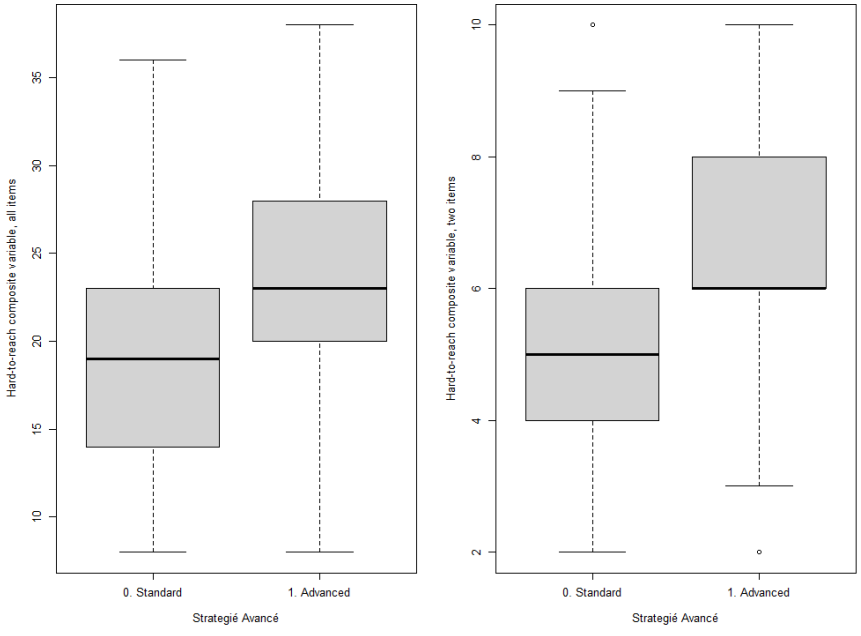


Figure 2. Boxplot visualisation of the key variable values in 'standard' and 'advanced' *Stratégie Avancée* groups.

Regarding convergent validity, the all-item composite score was significantly correlated with the as-crow-flies distance to nearest health facility (Pearson's $r = 0.36$; 95% CI = 0.31–0.40, $P < 0.001$), average monthly rainfall (Pearson's $r = 0.22$; 95% CI = 0.16–0.26, $P < 0.001$), and travel time to the nearest health facility by a vehicle (Pearson's $r = 0.35$, 95% CI = 0.30–0.39). The two-item composite score was significantly correlated with the as-the-crow-flies distance to nearest health facil-

ity (Pearson's $r=0.34$; 95% CI=0.30–0.39, $P<0.001$), average monthly rainfall (Pearson's $r=0.22$; 95% CI=0.17–0.27, $P<0.001$), and travel time to the nearest health facility by a vehicle (Pearson's $r=0.32$; 95% CI=0.27–0.37) (Table S4 in the **Online Supplementary Document**).

Inter-rater reliability

The individual ICC of all-item composite score and two-item composite score were 0.57 (95% CI=0.51–0.63) and 0.51 (95% CI=0.44–0.57), respectively (**Table 3**). Most of the other items in the survey gave individual ICC values between 0.4 and 0.5.

The harmonic mean number of DCs per village was 2.30. The all-item and two-item composite scores had average ICCs of 0.74 and 0.69, respectively. Similarly, most other items had average ICCs about 0.2 larger than their individual ICC. The difference indicates the improved performance of the measure by using 2–3 DCs to rate a village as compared to using only one DC to conduct the assessment.

Table 3. ICC on each five-point scale item and the composite score

	Individual ICC (95% CI)	Average ICC (95% CI)
Composite variable, all items	0.57 (0.51–0.63)	0.74 (0.68–0.78)
Composite variable, two items	0.51 (0.44–0.57)	0.69 (0.62–0.74)
Estimate of village residents' access to healthcare	0.49 (0.42–0.55)	0.66 (0.60–0.72)
Overall estimate of hardness to reach	0.49 (0.41–0.55)	0.67 (0.60–0.72)
How difficult is it to transport equipment to the village in the rainy season?	0.48 (0.40–0.54)	0.65 (0.58–0.71)
How difficult is it to transport equipment to the village in the non-rainy season?	0.46 (0.39–0.53)	0.65 (0.58–0.70)
Difficulty of the terrain on rainy season	0.64 (0.05–0.99)	0.68 (0.06–0.99)
Difficulty of the terrain on non-rainy season	0 (-0.64–0.63)	0 (-0.80–0.64)
Condition of the roads on rainy season	0.41 (0.33–0.47)	0.58 (0.50–0.65)
Condition of the roads in the non-rainy season	0.41 (0.33–0.48)	0.58 (0.49–0.65)
Estimate of travel time to the village on rainy season	0.32 (0.25–0.39)	0.50 (0.41–0.58)
Estimate of travel time to the village on non-rainy season	0.23 (0.15–0.30)	0.39 (0.27–0.48)

CI – confidence interval, ICC – intra-class coefficient

DISCUSSION

We developed and evaluated a new tool designed to measure the level of geographical hardness-to-reach of villages and evaluated its feasibility and measurement performance. Each DC who participated in the implementation of the tool had worked in the LAKANA trial in Mali and, although they rated multiple villages, they generally considered the survey to be feasible to implement in terms of respondent burden. Markedly, survey items and their derived composite scores indicated a 4–8% higher under-five mortality per one SD change, even when the model was adjusted for the distance from the nearest health facility, a usual predictor of increased under-five mortality. The tool's composite scores were associated with national outreach programme, *Stratégie Avancée* [17], classifications in an expected manner. The 'advanced' category had higher composite scores than the 'standard' category, indicating that the score effectively captures dimensions of accessibility. Convergent validity was supported by statistically significant associations between the composite score and objective indicators, including as-the-crow-flies distance and average monthly rainfall. While we did not pre-specify which level of coefficients would be acceptable, referring to the conventional classifications of 0.1, 0.3 and 0.5 as small, medium, and large [30] would suggest that our results indicate a medium level of evidence for convergent validity.

Subjective ratings and objective indicators may overlap, so the former may better be considered complementary, rather than as a replacement for the latter. Indeed, these findings indicate that the tool may capture meaningful aspects of geographical accessibility not otherwise captured by simple distance metrics, which often overlook barriers such as road conditions, terrain, weather,

and transport availability [31,32]. They could offer information relevant to strategic planning and insight for targeting health interventions in remote areas.

The two-item composite score correlated strongly with both the all-item composite score and the sum of remaining items, which indicates that the two-item version of the tool may be a meaningful approximation of the full-length tool. Furthermore, we found that the internal consistency among survey items was high and the individual inter-rater reliability was moderate, with the latter having the potential to be further improved by using multiple raters to assess the same locations, as average ICC values were stronger and highest for the composite scores. However, even individual ICCs for both composite scores had point estimates above 0.50.

Several limitations should be noted. First, and perhaps most importantly, the survey was conducted at the end of the trial period, meaning that most assessments were based on retrospective recall, rather than real-time observation. This introduces the possibility of recall bias [33]: since the DCs were in contact with the village residents, the residents' reports of health outcomes may have influenced the DC's evaluations. The retrospective dimension can also produce recall error [34]: since LAKANA was a multiyear trial, the DCs may have experienced inaccuracies for villages that were visited earlier due to the time since the visit. It is also possible that local infrastructure and circumstances have shifted since the visit, making the subjective evaluation outdated. Nevertheless, in villages evaluated by multiple DCs, inter-rater reliability was moderate, with higher ICCs for composite scores, indicating a consistent underlying signal despite individual variability. Recall errors are unlikely to be systematically aligned across DCs and are more plausibly driven by individual differences. Aggregating ratings is therefore expected to reduce such error and improve robustness. Operational factors may also have influenced results, and immediate assessment during visits could improve reliability, which warrants further study.

The item generation process was not comprehensive, as cognitive interviewing was not conducted and only one round of pilot testing was performed. However, the items were designed to be simple and unambiguous, which reduced the likelihood of misinterpretation. The pilot also yielded positive feedback from the five participating DCs. Furthermore, the DCs selected the villages they evaluated to ensure familiarity, which introduces potential selection bias. However, its impact is likely limited, as each DC assessed a large number of villages and overall coverage reached 99%. While some bias cannot be excluded, it is unlikely to have materially influenced the findings.

The terrain difficulty items showed low variability, leading to near-singular mixed-effects models with variance components close to zero. Consequently, the non-rainy season item yielded an ICC close to zero, while the rainy season item showed wide 95% CIs. Limited variability was also indicated by observed ceiling effect. Despite slightly less ceiling effect in the non-rainy item, within-village variability remained low, contributing to unstable ICC estimates, including negative 95% CI bounds.

Individual items rated by a single respondent yielded ICCs of approximately only 0.5. Given that the ICC can be interpreted as a sample size issue due to measurement imprecision [35], the use of the individual item rated by a single respondent requires doubling the sample size to achieve the same power or precision. In contrast, the use of composite scores and multiple raters increased ICC values to approximately 0.7. These findings should be considered when selecting items and designing data collection strategies, particularly with regard to the use of composite measures and the number of raters.

Although internal consistency was high, polychoric correlation exceeding 0.9 between 'overall estimate of hardness-to-reach' and 'access to healthcare' suggests potential redundancy. However, the items capture distinct dimensions, as a village may have good health access but still be considered hard-to-reach due to remoteness. Finally, while the *Stratégie Avancée* is an informative component for known-group validity analysis, it overlaps conceptually with the tool, potentially inflating group differences. These findings should, therefore, be interpreted as supportive, rather than definitive, and thus warrant future validation using more independent external criteria.

The number of DCs involved in the survey was relatively small. Despite this limitation, the study benefited from the geographic diversity covered by the DCs, who collectively provided assessments across a wide range of environments. To address the unbalanced and incomplete design, we applied current methodological recommendations for estimating ICCs [24].

A secondary analysis of the MORDOR trial found that each additional kilometre of distance from a health facility was associated with an approximately 5% increase in the incidence rate ratio of under-five mortality, with 95% CIs ranging from 3% to 7% [4]. Our results align with this finding when they are transformed to kilometre scale from SD scale. This is particularly relevant when considered in the context of scale up activities of azithromycin mass drug administration programming from Resiliency through Azithromycin for Child Survival network [36–39] and this tool could support ongoing efforts to identify and prioritise hard-to-reach populations. Moreover, as the importance of engagement of community stakeholders is widely recognised [40], and as the tool relies upon the subjective assessment of local individuals, the usage of this tool could bring into the implementation an aspect of engagement of the raters. For example, health workers, school teachers, outreach vaccine workers, and volunteers would act as co-investigators, rather than simply data collectors. In this context, a key contribution of this work is to propose and validate the concept that subjective measurement of hardness-to-reach can contribute information relevant to public health. It also provides relevant information for future researchers to make informed choices who would want to use either the full tool, composite scores, or single items. However, more validation research is needed to establish its usability across countries, personnel backgrounds, and cultural contexts.

CONCLUSIONS

This study has demonstrated the validity and reliability of the proposed tool for subjective assessment of geographical hardness-to-reach of communities and its ability to differentiate locations with varying mortality levels. A two-item composite score can improve feasibility and reduce respondent burden without much loss in validity and reliability, as compared to using the version of composite score based on 14 items. Some of the 14 items were specific to the LAKANA trial context such as those related to transportation of equipment. In contrast, the two-item composite score relies on generic items that are broadly applicable across settings. Therefore, we see higher applicability of the two-item versions, with the caveat that further evidence is needed to establish performance across contexts. Future validation in other low-income settings and possibly other types of raters will be valuable for further refinement.

Acknowledgements: We thank Awa Traore, Uma Onwuchekwa, Fatoumata Diallo, Mamoudou Kodio, Djouma Keita, Moussa Traore, and Kevin Wilson for technical support; Elisabeth Diallo, Béatrice Abossolo, and Modibo Diarra for external monitoring; Karen Kotloff, Anthony Solomon, and Thomas Lietman for scientific advice; Robert Black, Paul Milligan, Julia Bielicki, Queen Dube, and Alassane Dicko for serving on the data and safety monitoring board; and the trial communities and the staff at the Malian Ministry of Health and Social Development for their constant support. We also express our gratitude to the LAKANA data collectors who took part in this study.

Ethics statement: The trial was conducted in accordance with Good Clinical Practice guidelines, the principles of the Declaration of Helsinki, and regulatory requirements in Mali. Ethics approval was obtained from the Mali Institutional Review Board, *i.e.* the *Comité d'Éthique de l'Université des Sciences, des Techniques et des Technologies de Bamako* (approval number N2019/181/CE/FMPOS/FAPH). The study was initially approved on 27 December 2019, with the most recent annual reapproval (approval number N2025/200/USTBB) granted on 13 October 2025. The study also received approval from the Ethics Committee of the Pirkanmaa Hospital District.

Data availability: The data are available *via* Zenodo (https://doi.org/10.5281/zenodo.18312097) upon reasonable request to the LAKANA Trial Principal Investigators at lakana.psg@lists.tuni.fi. It is aggregated at a pseudonymised cluster level and contains no information traceable to individuals. This Zenodo project includes the LAKANA hard-to-reach data set.csv (aggregated cluster-level data from LAKANA and hard-to-reach tool evaluations) and the LAKANA hard-to-reach data dictionary.xlsx (variable descriptions). The data are available under the terms of the Creative Commons

Attribution 4.0 International license. The underlying analytical code for the hard-to-reach study is available at GitHub (https://github.com/LAKANA-Trial-team/Hard-to-reach/tree/v3.0.0) and archived on Zenodo (https://doi.org/10.5281/zenodo.18325366). The code is distributed under an OSI-approved open-source license.

Funding: This study was supported by grants from the Gates Foundation (OPP1210821, INV-003354, and INV-005877). Pfizer Inc. donated azithromycin and placebo. The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Authorship contributions: JL and YBC designed and conceptualised the study. JL designed and tested the electronic data capture application and was responsible for data curation and software management. FCH, OS, and SS organised the survey rounds and coordinated data collection. JL pre-processed and analysed the data. JL, KY, and LA had directly accessed and verified the underlying data and proof-read the analytical code. YBC and PA supervised the work and methodology. All authors participated in writing, reviewing, and editing the manuscript, and have read and agreed to the published version. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication

Disclosure of interest: The authors completed the ICMJE Disclosure of Interest Form (available upon request from the corresponding author) and disclose no relevant interests.

Additional material
Online Supplementary Document

PAPERS

REFERENCES

1 Feikin DR, Nguyen LM, Adazu K, Ombok M, Audi A, Slutsker L, *et al.* The impact of distance of residence from a peripheral health facility on pediatric health utilisation in rural western Kenya. Trop Med Int Health. 2009;14:54–61. **Medline:19021892 doi:10.1111/j.1365-3156.2008.02193.x**

2 Karra M, Fink G, Canning D. Facility distance and child mortality: a multi-country study of health facility access, service utilization, and child health outcomes. Int J Epidemiol. 2017;46:817–26. **Medline:27185809**

3 Kibret GD, Demant D, Hayen A. The effect of distance to health facility on neonatal mortality in Ethiopia. BMC Health Serv Res. 2023;23:114. **Medline:36737761 doi:10.1186/s12913-023-09070-x**

4 Chao DL, Arzika AM, Abdou A, Maliki R, Karamba A, Galo N, *et al.* Distance to Health Centers and Effectiveness of Azithromycin Mass Administration for Children in Niger: A Secondary Analysis of the MORDOR Cluster Randomized Trial. JAMA Netw Open. 2023;6:e2346840. **Medline:38100110 doi:10.1001/jamanetworkopen.2023.46840**

5 Stanton MC, Bockarie MJ, Kelly-Hope LA. Geographical Factors Affecting Bed Net Ownership, a Tool for the Elimination of Anopheles-Transmitted Lymphatic Filariasis in Hard-to-Reach Communities. PLoS One. 2013;8:e53755. **Medline:23308281 doi:10.1371/journal.pone.0053755**

6 Ozawa S, Yemeke TT, Evans DR, Pallas SE, Wallace AS, Lee BY. Defining hard-to-reach populations for vaccination. Vaccine. 2019;37:5525–34. **Medline:31400910 doi:10.1016/j.vaccine.2019.06.081**

7 Kaewkungwal J, Apidechkul T, Jandee K, Khamsiriwatchara A, Lawpoolsri S, Sawang S, *et al.* Application of Mobile Technology for Improving Expanded Program on Immunization Among Highland Minority and Stateless Populations in Northern Thailand Border. JMIR Mhealth Uhealth. 2015;3:e4. **Medline:25589367 doi:10.2196/mhealth.3704**

8 Uddin MJ, Sjamsuzzaman M, Horng L, Labrique A, Vasudevan L, Zeller K, *et al.* Use of mobile phones for improving vaccination coverage among children living in rural hard-to-reach areas and urban streets of Bangladesh. Vaccine. 2016;34:276–83. **Medline:26647290 doi:10.1016/j.vaccine.2015.11.024**

9 Kaji A, Parker DM, Chu CS, Thayatkawin W, Suelaor J, Charatrueangrongkun R, *et al.* Immunization Coverage in Migrant School Children Along the Thailand-Myanmar Border. J Immigr Minor Health. 2016;18:1038–45. **Medline:26483180 doi:10.1007/s10903-015-0294-x**

10 Watson-Jones D, Mugo N, Lees S, Mathai M, Vusha S, Ndirangu G, *et al.* Access and Attitudes to HPV Vaccination amongst Hard-To-Reach Populations in Kenya. PLoS One. 2015;10:e0123701. **Medline:26115523 doi:10.1371/journal.pone.0123701**

11 Kofie RY, Attua EM, Nabila JS. Poverty and socio-economic consequences of Buruli ulcer (Mycobacterium ulcerans) in the Ga West District of Ghana. Nor Geogr Tidsskr – Nor J Geogr. 2008;62:210–21. **doi:10.1080/00291950802335855**

12 Bal M, Rana R, Das A, Khuntia HK, Somalkar N, Sahoo N, *et al.* Neglected malaria parasites in hard-to-reach areas of Odisha, India: implications in elimination programme. Malar J. 2021;20:482. **Medline:34949205 doi:10.1186/s12936-021-04010-8**

REFERENCES

13 Forson AO, Awuah RB, Mohammed AR, Owusu-Asenso CM, Akosah-Brempong G, Abdulai A, *et al.* Coverage of preventive measures and surveillance for neglected tropical diseases in hard-to-reach communities in Ghana. BMC Public Health. 2023;23:1784. **Medline:37710219 doi:10.1186/s12889-023-16652-1**

14 Adeleye VM, Ashinze P, Okeoyo T, El Nabbout V, Phiri EC, Gautam G, *et al.* Advancing public health through Spatial Data Infrastructures: a review of global practices, governance and policy recommendations. Discov Public Health. 2025;22:189. **doi:10.1186/s12982-025-00563-0**

15 State Smart Transportation Initiative. Measuring Accessibility. 2026. Available: https://ssti.us/measuring-accessibility/. Accessed: 12 September 2025.

16 Australian Bureau of Statistics. Remoteness Areas. 21 March 2023. Available: https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/remoteness-structure/remoteness-areas. Accessed: 11 September 2025.

17 Ministere de la Sante et de L'hygiene Publique. Plan Strategicque National Des Soins Essentiels Dans La Communaute. Bamako, Mali: Ministere de la Sante et de L'hygiene Publique; 2015. Available: https://www.unicef.org/mali/media/2291/file/psn%202026-2020.pdf. Accessed: 8 October 2025.

18 Elasy TA, Gaddy G. Measuring subjective outcomes: Rethinking reliability and validity. J Gen Intern Med. 1998;13:757–61. **Medline:9824522 doi:10.1046/j.1525-1497.1998.00228.x**

19 Cleary PD. Subjective and Objective Measures of Health: Which is Better When? J Health Serv Res Policy. 1997;2:3–4. **Medline:10180651 doi:10.1177/135581969700200102**

20 Adubra L, Alber D, Ashorn P, Ashorn U, Cheung YB, Cloutman-Green E, *et al.* Testing the effects of mass drug administration of azithromycin on mortality and other outcomes among 1–11-month-old infants in Mali (LAKANA): study protocol for a cluster-randomized, placebo-controlled, double-blinded, parallel-group, three-arm clinical trial. Trials. 2023;24:5. **Medline:36597115 doi:10.1186/s13063-022-06966-7**

21 Hijmans RJ. geosphere: Spherical Trigonometry. cran.r-project.org. 2026. Available: https://cran.r-project.org/web/packages/geosphere/index.html. Accessed: 4 June 2025.

22 Pisel T. openmeteo. github.com. Available: https://github.com/tpisel/openmeteo. Accessed: 4 June 2025.

23 Giraud T. osrm: Interface Between R and the OpenStreetMap-Based Routing Service OSRM. J Open Source Softw. 2022;7:4574. **doi:10.21105/joss.04574**

24 Ten Hove D, Jorgensen TD, Van Der Ark LA. Updated guidelines on selecting an intraclass correlation coefficient for interrater reliability, with applications to incomplete observational designs. Psychol Methods. 2024;29:967–79. **Medline:36048052 doi:10.1037/met0000516**

25 Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull. 1979;86:420–8. **Medline:18839484 doi:10.1037/0033-2909.86.2.420**

26 McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1:30–46. **doi:10.1037/1082-989X.1.1.30**

27 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15:155–63. **Medline:27330520 doi:10.1016/j.jcm.2016.02.012**

007828 Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6:284–90. **doi:10.1037/1040-3590.6.4.284**

29 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. Ann Intern Med. 2007;147:573–7. **Medline:17938396 doi:10.7326/0003-4819-147-8-200710160-00010**

30 Machin D, Campbell MJ, Tan SB, Tan SH. Sample Size Tables for Clinical Studies. 1st ed. London, UK: Wiley; 2008.

31 Dotse-Gborgbortsi W, Nilsen K, Ofosu A, Matthews Z, Tejedor-Garavito N, Wright J, *et al.* Distance is 'a big problem': a geographic analysis of reported and modelled proximity to maternal health services in Ghana. BMC Pregnancy Childbirth. 2022;22:672. **Medline:36045351 doi:10.1186/s12884-022-04998-0**

32 Ludrosky JM, Newhouse A, Hudnall E, Sheree A, Perle JG. When a Straight Line Is Not the Most Direct Method: an Evaluation of Straight Line Versus True Distance Metrics for Patients in Rural Settings. J Behav Health Serv Res. 2023;50:214–20. **Medline:35945480 doi:10.1007/s11414-022-09812-5**

33 Furnham A, Boo HC. A literature review of the anchoring effect. J Socio-Econ. 2011;40:35–42. **doi:10.1016/j.socec.2010.10.008**

34 Loftus EF, Pickrell JE. The Formation of False Memories. Psychiatr Ann. 1995;25:720–5. **doi:10.3928/0048-5713-19951201-07**

35 Cheung YB. Statistical Analysis of Human Growth and Development. 1st edition. New York, USA: Chapman and Hall/CRC; 2013.

36 REACH Network. REACH approved for national scale-up in Burkina Faso. August 2025. Available: https://reachnetwork.africa/en/scale-up-approved-burkina-faso. Accessed: 8 October 2025.

37 REACH Network. REACH Mali begins national azithromycin distribution in Ségou. August 2024. Available: https://reachnetwork.africa/en/reach-mali-launch. Accessed: 8 October 2025.

38 Niger REACH. Azithromycin for child survival in Niger – implementation and research. 26 March 2024. Available: https://reachnetwork.africa/en/azithromycin-for-child-survival-in-niger-implementation-and-research. Accessed: 8 October 2025.

39 REACH Network. SARMAAN II planning to treat 18 million children in 11 states in Nigeria. 2024. Available: https://reachnetwork.africa/en/sarmaan-18-million-children. Accessed: 8 October 2025.

40 Gobat N, Slack C, Hannah S, Salzwedel J, Bladon G, Burgos JG, *et al.* Better engagement, better evidence: working in partnership with patients, the public, and communities in clinical trials with involvement and good participatory practice. Lancet Glob Health. 2025;13:e716–31. **Medline:40155109 doi:10.1016/S2214-109X(24)00521-7**