nature water

Article

https://doi.org/10.1038/s44221-025-00434-7

Validity and reliability of the Sanitation-related Quality of Life index (SanQoL-5) in six countries

Received: 29 September 2024

Accepted: 28 March 2025

Published online: 28 April 2025

Check for updates

Fouzia Akter[1,2], Neiva Banze[3], Igor Capitine[3], Kondwani Chidziwisano[4], Jenala Chipungu[5], Catildo Cubai[3], Oliver Cumming[1], Robert Dreibelbis[1], Patrick Vidzo Katana[1], Cremildo Manhiça[3], Mindy Panulo[4], Priya Rampal[6], Anjali Sharma[5], Sheillah Simiyu[7], Abiy Tafesse[8], James Ben Tidwell[9], Edna Viegas[3], Blessings White[4] & Ian Ross[1]

Sustainable Development Goal 6.2 measures sanitation progress by type of toilet service. Improving people's subjective sanitation experiences is also important but rarely rigorously measured. The Sanitation-related Quality of Life index (SanQoL-5) combines answers to five simple questions (disgust, privacy, disease risk, shame and safety) into an overall score ranging from 0 to 1. Here we evaluated the validity and reliability of SanQoL-5 by interviewing 6,165 people across rural and urban areas of six countries: Ethiopia, India, Kenya, Malawi, Mozambique and Zambia. We found good evidence for construct validity, with support ($P < 0.05$) for 87% of hypothesized associations between SanQoL-5 and toilet quality characteristics. In 75 intercountry comparisons, only 9% of instances showed evidence of meaningful differential item functioning, suggesting good cross-cultural comparability. SanQoL-5 conformed to expectations in item response theory models, and we found evidence of convergent, discriminant and known groups validity. SanQoL-5 can be used in impact evaluation, monitoring, needs assessment and benefit–cost analysis.

Poor sanitation is an enduring public health challenge, with 1.5 billion people globally lacking access to a basic toilet[1]. Studies of the effectiveness of sanitation improvements often focus on disease[2] and behaviour change[3]. Sustainable Development Goal (SDG) 6.2 measures progress on sanitation by type of toilet service, which is important and objectively measurable[1]. Improvements in people's subjective sanitation experiences are also important but rarely rigorously measured. These experiences could be things that happen to, or are felt by, people while they carry out sanitation practices. Sanitation-related quality of life refers to how sanitation practices and services directly affect people's experiences, for example privacy, safety and disgust[4]. Measuring these outcomes is important because they are often rated highly (and alongside or above disease) as drivers of household sanitation decisions[5–7] and contribute to health in its broadest sense[8].

'Health-related quality of life' experiences have routinely been measured in health studies since the 1980s (ref. 9). Field-specific quality of life measurement in many areas of environmental health has been limited, but the recently developed water insecurity experiences (WISE) scales[10,11] are now delivering insights into the causes and consequences of water insecurity[12,13]. Sanitation has lagged behind, but

[1]London School of Hygiene and Tropical Medicine, London, UK. [2]Daffodil International University, Dhaka, Bangladesh. [3]Instituto Nacional de Saúde, Marracuene, Mozambique. [4]Malawi University of Business and Applied Sciences, Blantyre, Malawi. [5]Centre for Infectious Disease Research in Zambia, Lusaka, Zambia. [6]OPM India, New Delhi, India. [7]African Population and Health Research Center, Nairobi, Kenya. [8]World Vision Ethiopia, Addis Ababa, Ethiopia. [9]World Vision US, Federal Way, WA, USA. e-mail: ian.ross@lshtm.ac.uk

Table 1 | SanQoL-5 questions (descriptive system)

Attribute	Question (item)[a]	Responses
Disgust	How often do you feel disgusted while using the toilet?	Always Sometimes Never
Disease	How often do you worry that the toilet spreads diseases?	
Safety	How often do you feel unsafe while using the toilet?	
Shame	How often do you feel ashamed about using the toilet?	
Privacy	How often do you worry about being seen while using the toilet?	

[a]A preamble is as follows: 'The following questions are about your sanitation experiences in the past 30 days, meaning defecation, urination, and anything else you do in a toilet. Please respond with always, sometimes or never.' If less literate respondents struggle with a question, it can be repeated as 'Do you feel disgusted while using the toilet? How often?'. Before the SanQoL-5 questions, the respondent is asked about the last place they defecated. If the respondent practiced open defecation (OD), for example, in fields or wasteland, they are directed to OD-specific questions, for example 'How often do you worry about being seen while practising open defecation?'. Further guidance is available at www.SanQoL.org.

the Sanitation-related Quality of Life index (SanQoL-5) has now been applied in 15 populations across 9 countries. SanQoL-5 measures people's sanitation experiences across five attributes: privacy, disgust, safety, shame and disease[14]. Each attribute is measured by a short question with answers on a three-level frequency scale (Table 1). The quality of life theory underlying SanQoL-5 is the capability approach to welfare economics[15,16]. The five attributes measure outcomes people have 'reason to value' about sanitation, as identified in qualitative work[4] and prior literature[17,18]. The questions measure 'functionings' (people's achievement of capabilities in the past 30 days) rather than the capabilities themselves (the broader set from which people have freedom to choose).

The SanQoL-5 draws on methods common in health economics for developing measures to use in economic evaluation[19]. An economic evaluation purpose requires a small number of attributes or questions (typically <7), such as the 'EQ-5D', which measures and values health-related quality of life[20,21]. Attributes are selected primarily for content validity—the extent to which the most relevant and important aspects of a concept are captured[22]. Rather than using factor analysis as applied in scale development, the attributes are valued as an index anchored at 0 and 1, with weights based on preference elicitation[23]. The index then represents a given population's relative valuation of those preselected attributes.

The original SanQoL-5 development study[14] found that the index demonstrated favourable psychometric proprieties, but only in one urban setting (Maputo, Mozambique) with a modest sample size (n = 424). Since initial development, there have been refinements to the questions based on mixed-methods cognitive testing and piloting in Zambia and Ghana. Given that SanQoL-5 has now been applied in 15 populations and its questions recommended by the United Nations Children's Fund (UNICEF)/World Health Organization (WHO) Joint Monitoring Programme for gender monitoring[24], revalidation of the measure in broader settings and larger samples is warranted but has not yet been undertaken.

In this Article, we assess the validity and reliability of the SanQoL-5 index by interviewing 6,165 people across rural and urban areas in six countries: Ethiopia, India, Kenya, Malawi, Mozambique and Zambia. In doing so, we address the need for validated sanitation measures focused on quality of life, which can be simply and consistently deployed in studies and routine monitoring.

Sample characteristics

Samples ranged from 0% rural to 100% rural (Table 2), and sampling and representativeness varied by country (Methods). Samples were approximately gender-balanced in all countries apart from Malawi, where 83% of respondents were women (Table 2). The mean age was about 40 years in all countries. Samples varied in terms of types of sanitation. In Mozambique, the most common toilet (50%) was a pit latrine with concrete slab, with 32% using a flush or pour-flush toilet. In India, all those using a toilet used a flush or pour-flush. In all other countries (pour-)flush toilets were very uncommon (<1%), with most people (61–83%) using a pit latrine with wood/soil slab. Open defecation (at the last time people defecated) ranged from 3% in Malawi to 35% in India. We present distributions of SanQoL-5 attribute levels (Fig. 1), as well as histograms of SanQoL-5 index values (Supplementary Information A). All attribute levels received at least 10% of responses in all countries. SanQoL-5 value sets (weighted indices) are provided in Supplementary Information B.

Validity and reliability

There was good evidence for the construct validity of SanQoL-5. There was evidence at $P < 0.10$ for 91% of the hypothesized associations with toilet quality characteristics in individual regressions (87% at $P < 0.05$) and 86% (71% at $P < 0.05$) in concurrent regressions (Table 3). All associations were in the hypothesized direction, indicating that better-quality toilets were associated with higher SanQoL-5. Full regression output is reported in Supplementary Information C. The only hypothesized variable showing a lack of association with SanQoL-5 in more than one country was whether the toilet was shared with other households. Only 10% of negative controls were associated with SanQoL-5 at $P < 0.10$.

People with higher sanitation levels of service tended to have higher SanQoL-5 index values, which supports 'known groups' validity (Fig. 2). In the two samples in which >15% of the sample practised OD, there was a SanQoL-5 gain from toilet use over OD (Table 3 and Fig. 2). Correlation between SanQoL-5 and sanitation visual analogue scale (SanVAS)[25] scores (Supplementary Information D) ranged from 0.28 in Malawi to 0.58 in Mozambique, all in the hypothesized direction at $P < 0.001$, which is evidence of convergent validity (Supplementary Information E). Considering discriminant validity, SanQoL-5 was independent of two EQ-5D health-related QoL variables in the two countries we collected those data (India and Kenya). Specifically, the five-level EQ-5D variable for 'problems in walking about' was not correlated with SanQoL-5 in either country (Kenya $r = -0.04$ ($P = 0.33$), India $r = -0.04$ ($P = 0.89$)). EQ-5D 'pain or discomfort' was also not correlated with SanQoL-5 (Kenya $r = -0.08$ ($P = 0.06$), India $r = -0.033$ ($P = 0.26$)). Cronbach's α ranged from 0.73 to 0.92 per country (0.85 in pooled data), indicating good internal reliability (Supplementary Information E).

Item response theory

In the pooled and country-specific item response theory (IRT) models, category characteristic curves show a distinct peak for the 'sometimes' level across all attributes (Supplementary Information F), confirming that it is appropriate to maintain three-level attributes rather than binary. The item information functions show that all attributes are giving good information across the construct (theta) with privacy and shame providing more information than the others (Supplementary Information F). The five attributes have similar 'difficulty' (no Guttman ordering), which feeds into a smooth test information function covering the breadth of the construct. Neither the pooled nor country-specific models raise any concerns.

Measurement invariance

With 5 attributes and 15 possible comparisons among 6 countries, there were 75 possible instances of differential item functioning (DIF). Among these, seven (9%) exhibited DIF in ordinal logistic regression that was 'meaningful', following the widely used[26] cut-off of 2% increase in pseudo-R^2 (Supplementary Information G). This provides good evidence of equivalent measurement and meaning across countries. While further assessment of DIF in larger representative samples is

Table 2 | Respondent and toilet characteristics

	Ethiopia (*n*=1,586)	India (*n*=1,213)	Kenya (*n*=1,000)	Malawi (*n*=1,400)	Mozambique (*n*=601)	Zambia (*n*=365)
Milieu	81% rural	87% rural	71% rural	100% rural	0% rural	100% rural
Demographic characteristics						
Respondent is female	829 (52%)	607 (50%)	514 (51%)	1167 (84%)	330 (55%)	182 (50%)
Respondent age (mean, s.d.)	41.8 (13.6)	34.8 (10.8)	35.3 (13.1)	39.9 (16.1)	40.4 (16.4)	43.3 (15.6)
Aged 18–29	286 (18%)	430 (35%)	412 (41%)	401 (29%)	205 (34%)	80 (22%)
Aged 30–44	663 (42%)	518 (43%)	338 (34%)	529 (38%)	182 (30%)	129 (35%)
Aged 45–59	419 (26%)	229 (19%)	194 (19%)	275 (20%)	111 (18%)	91 (25%)
Aged 60+	217 (14%)	36 (3%)	56 (6%)	187 (13%)	103 (17%)	65 (18%)
Household size (mean, s.d.)	5.3 (2.1)	6.2 (3.0)	4.7 (2.6)	4.5 (1.7)	5.3 (2.5)	5.8 (3.0)
Completed primary school or above	722 (91%)[a]	865 (71%)	853 (85%)	1,256 (90%)	399 (66%)	258 (71%)
Piped water on-premises	469 (30%)	114 (9%)	287 (29%)	4 (0.3%)	358 (60%)	17 (5%)
Sanitation characteristics						
Toilet type						
Flush or pour-flush toilet	5 (0.3%)	784 (65%)	209 (21%)	1 (0.1%)	195 (32%)	5 (1%)
Pit latrine with concrete slab	219 (14%)	0 (0%)	563 (56%)	194 (14%)	301 (50%)	61 (17%)
Pit latrine with wood/soil slab	961 (61%)	0 (0%)	195 (20%)	1,159 (83%)	47 (8%)	254 (70%)
Open defecation	401 (25%)	429 (35%)	32 (3%)	46 (3%)	58 (10%)	45 (12%)
Toilet shared with other households[b]	187 (16%)	87 (12%)	349 (36%)	825 (62%)	153 (28%)	127 (40%)
Toilet has solid walls[b]	358 (30%)	781 (99%)	762 (82%)	961 (72%)	430 (72%)	218 (81%)
Faeces not observed on pan/slab[b]	656 (81%)	593 (75%)	631 (78%)	1,088 (82%)	579 (96%)	254 (94%)
Pan/slab is concrete, porcelain or similar[b]	220 (19%)	784 (100%)	769 (80%)	195 (15%)	496 (91%)	64 (20%)
Water seal is present (flush or pour-flush)[b]	5 (0.4%)	801 (100%)	209 (21%)	1 (0.1%)	195 (36%)	5 (2%)
Toilet has inside lock[b]	n/a	612 (78%)	714 (77%)	65 (5%)	152 (28%)	6 (2%)

Data are *n* (%) for categorical variables and mean (s.d.) for numerical variables. Percentages for categorical variables are the percentage of those with non-missing data for that variable. n/a = not applicable. [a]In Ethiopia, data are for highest level of education 'reached' rather than 'completed', and the question was randomized to be asked of only half the sample. [b]Only observed or asked for toilet users.

recommended, these results broadly support the cross-cultural comparability of SanQoL-5.

Comparing question framings

Analyses comparing the old question framings to the current framings (Table 1) in Ethiopia and Zambia support use of the current questions, where 'always' is the worst outcome. A higher proportion of hypothesized variables had statistically significant associations in both countries (Ethiopia and Zambia) under the current questions. In no area did the old questions perform better (Supplementary Information H).

Discussion

This study evaluated multiple aspects of validity and reliability of a five-attribute Sanitation-related Quality of Life index (SanQoL-5), using rigorous psychometric methods across six countries in diverse rural and urban settings. The SanQoL-5 questions are short and simple, and together take around 1–2 min to administer. The SanQoL-5 covers a breadth of what people value about sanitation: avoiding disgust, avoiding shame, avoiding disease risks, having safety and having privacy. Rather than focusing on toilet types like SDG 6, the SanQoL-5 index captures people's sanitation-related experiences.

Our study included populations using a variety of sanitation types in rural and urban settings in six countries, with SanQoL-5 responses covering the full range of attribute levels (Fig. 1). We have presented evidence for different types of validity and reliability generated using these datasets, and our findings on measurement invariance support cross-cultural comparability. We believe that SanQoL-5 can be widely applied with adults, but at this stage there is good evidence of validity only in African countries and northern India. Validation is a continuous process, even for long-established measures[27]. Further exploration of the validity of SanQoL-5 in other world regions is required, and we recommend that piloting and/or cognitive interviews ideally be undertaken before application in new settings or languages[28]. We also recommend that users of SanQoL-5 undertake their own validity and reliability assessments, wherever possible. There was prior evidence of test–retest reliability[14], although further exploration of this is needed, and future studies should also investigate predictive validity (for example, SanQoL-5 at one timepoint 'foretelling' some subsequent outcome). Translations in several languages are available at www.SanQoL.org. Before this study, an earlier version of SanQoL-5 had been validated in one setting (urban Mozambique). Improvement of the questions based on mixed-methods research in multiple countries resulted in the updated version of SanQoL-5 that we have evaluated here, which is easier to understand and performed better in head-to-head comparisons (Supplementary Information H).

There are several possible applications of the SanQoL-5 index. First, SanQoL-5 can be used as an outcome in impact evaluation (for example, differences compared with a counterfactual), as already done in several studies[29–31]. Second, it can be used in the monitoring and evaluation of programmes (for example, differences in a group over time). SanQoL-5

https://doi.org/10.1038/s44221-025-00434-7

was already used for this purpose by the non-government organizations World Vision and Water & Sanitation for the Urban Poor, and in a high-frequency monitoring study of container-based sanitation[32]. Third, it can be used in needs assessment, for example characterizing the scale and nature of sanitation problems in a population. Fourth, it can be used in economic evaluation of sanitation programmes, as in a cost-effectiveness analysis in Mozambique[33] and a benefit–cost analysis in progress in Malawi[29]. It is for this economic purpose that SanQoL-5 was designed as a weighted index[20], thereby capturing the value of sanitation to people. This design feature was a necessary condition for allowing SanQoL-5 gains to be given monetary value in benefit–cost analysis, based on willingness to pay.

Fig. 1 | Distributions of SanQoL-5 attribute levels by country. a, Ethiopia, $n = 1{,}570$. **b**, India, $n = 1{,}212$. **c**, Kenya, $n = 988$. **d**, Malawi, $n = 1{,}400$. **e**, Mozambique, $n = 601$. **f**, Zambia, $n = 365$.

In all of these uses, SanQoL-5 provides complementary subjective information to objective 'quality of service' measures (for example, roof or wall quality in the 'sanitation quality index'[34]). Subjectivity is a characteristic of all quality of life measures[35]. It can be helpful to know, for example, whether people's subjective perception of disease risk has changed as a result of a programme, as compared with actual disease cases. These two things may not always be closely related[36].

In all six settings evaluated here, people with progressively higher levels of sanitation service tended to have progressively higher SanQoL-5 (Fig. 2). This is evidence of known groups validity but also demonstrates the potential of SanQoL-5 to evaluate relative QoL gains arising from different sanitation programmes and policies. There was diversity of SanQoL-5 experience within each sanitation service category (Fig. 2). This is unsurprising because each contains a variety of individuals with their own characteristics and experiences, as well as toilet subtypes in different states of condition, noting the conceptual model underlying SanQoL-5 (ref. 4).

Sharing toilets with other households is often assumed to deliver worse outcomes than private toilets[37]. On the one hand, therefore, it was unexpected that the variable for sharing was not statistically significantly associated with SanQoL-5 in Ethiopia and Zambia. On the other hand, these samples were predominantly rural (81% and 100%, respectively), and it is plausible that any negative consequences of sharing are more acute in dense urban settings. Sharing toilets may be more palatable in sparsely populated rural areas with smaller numbers of sharing households. Among those who shared toilets, the median number of households sharing was 2 in all countries (predominantly rural settings) except Mozambique (urban) where it was 3. Further exploration in urban settings of the relationship between sharing and SanQoL-5 is required.

Table 3 | *P* values on coefficients for hypothesized associations in GLMM regressions, individually and concurrently

	Ethiopia (n=774)		India (n=789)		Kenya (n=796)		Malawi (n=1,328)		Mozambique (n=539)		Zambia (n=268)	
	ind.	conc.	ind.	conc.	ind.	conc.	ind.	conc.	ind.	conc.	ind.	conc.
Hypothesized to be associated with SanQoL-5												
Toilet has solid walls	**<0.001**	**0.001**	n/a (<1% don't have)		0.157	0.572	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.015**	**0.047**
Faeces not observed on pan/slab	**0.002**	**0.004**	**0.020**	**0.057**	**0.001**	**0.006**	**<0.001**	**<0.001**	n/a (4% have)		n/a (6% have)	
Pan/slab is concrete, porcelain or similar	**<0.001**	**<0.001**	n/a (0% don't have)		n/a (water seal overlap)		n/a (14% have)		n/a (9% don't have)		**0.035**	**0.062**
Water seal present (flush or pour-flush)	n/a (<1% have)		n/a (0% don't have)		**0.004**	**0.062**	n/a (<1% have)		**<0.001**	**<0.001**	n/a (2% have)	
Toilet has inside lock	n/a		**<0.001**	**<0.001**	**0.005**	**0.047**	n/a (5% have)		**<0.001**	**0.006**	n/a (2% have)	
Toilet not shared with other households	**0.092**	0.718	n/a (11% share)		**<0.001**	**0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.789	0.522
Any toilet versus open defecation	**<0.001**	n/a	**<0.001**	n/a	n/a (OD 4%)	n/a	n/a (OD 3%)	n/a	n/a (OD 10%)	n/a	n/a (OD 12%)	n/a
Negative controls												
Partner	**0.861**	n/a	0.018	n/a	**0.269**	n/a	**0.923**	n/a	n/a	n/a	n/a	n/a
Household size	**0.136**	n/a	**0.513**	n/a	**0.197**	n/a	**0.310**	n/a	**0.946**	n/a	**0.666**	n/a

Individual (ind.) models regressed on SanQoL-5 index values and the indicated variable. Concurrent (conc.) models regressed on all hypothesized variables at once. *P* values less than 0.10 are emboldened, expect for negative contols where *P* values greater than 0.10 are emboldened. *N* per country is the sample for the concurrent results, and less than the full sample per country because of missing data (for example, no toilet observations for people practising open defecation). 'Any toilet versus open defecation' is not included in concurrent models because it would result in zero observations (those practising OD have no data for toilet variables). We assessed a covariate for a given country only if ≥15% of the sample with non-missing data was in each category, to ensure a minimum of statistical power. No adjustments were made for multiple comparisons. Full regression output for concurrent models is in Supplementary Information C alongside *P* values to further decimal places and exact *P* values for $P < 0.001$. All statistical tests are two-tailed.

https://doi.org/10.1038/s44221-025-00434-7

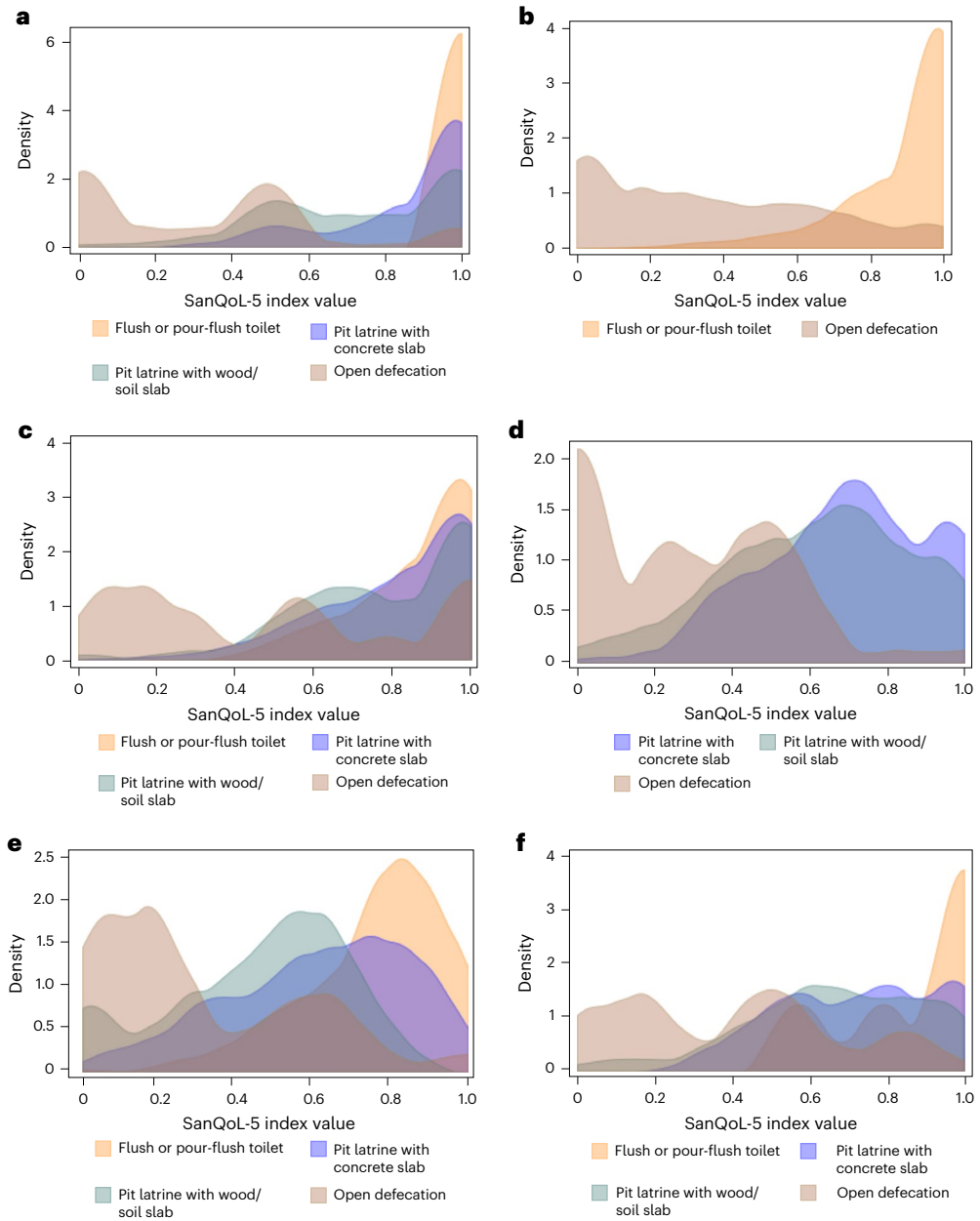

Fig. 2 | SanQoL-5 index kernel density distributions by toilet type. a, Ethiopia. **b**, India. **c**, Kenya. **d**, Malawi. **e**, Mozambique. **f**, Zambia.

SanQoL-5 has thus far only been used in adult populations, but it could be useful in adolescent or child populations, for example, in the context of school sanitation as well as households. About one-sixth of the world's population are adolescents[38], who may experience sanitation in different ways to adults[39,40]. Further work on content validity and ease of understanding is required for its use among children and adolescents. Questions may also need amending, as in the youth version of the EQ-5D[41].

The SanQoL-5 captures five dimensions of sanitation-related quality of life and makes no claim to measure all aspects of sanitation-related QoL that may be important. Users requiring more granularity might include other longer measures alongside it. For example, scales in the Agency, Resources, and Institutional Structures for Sanitation-related Empowerment (ARISE) family capture many aspects of sanitation-related QoL among women in more detail, but with a large number of questions that take more time[42]. Users are reminded that SanQoL-5 development followed design principles common in measures for economic evaluation[23], with attributes selected for content validity[22] rather than based on factor analysis. As with any measure development effort, alternative methods might have delivered a different instrument. As above, we recommend measuring quality of service alongside QoL outcomes[34].

A priority for future research is a more detailed exploration of which toilet types or characteristics are associated with the biggest gains in SanQoL-5, to inform policy and programming decisions. A further priority is the investigation of gender differences in SanQoL-5 (in particular, intrahousehold differences), as investigated for water and food security[43,44]. A strength of SanQoL-5 is that it is applicable to any gender, meaning it can identify gaps or inequalities between women and men.

Strengths of our study include the diversity of countries, rural and urban milieus and toilet types used, as well as the variety of analytical methods for assessing different aspects of validity and reliability. Limitations include that data were not collected for some aspects of reliability, for example test–retest, although this was assessed in the earlier SanQoL-5 study[14]. Responsiveness of SanQoL-5 to changes in

sanitation services over time could also not be assessed in this study, although it was demonstrated in an earlier study[45]. Other limitations include that, although samples were large enough for validity and reliability assessment, they were in relatively small geographic areas within each country (apart from Kenya, which was nationally representative). Furthermore, no high-income countries were included, but evidence suggests that there may be sanitation-related quality of life deficits in those countries (e.g., among groups often excluded from sanitation services owing to poverty or discrimination)[46].

The SanQoL-5 index provides a short and simple measure capturing the outcomes people value about sanitation, which are also what often motivate toilet purchases and upgrades. A single overall score, combining five important experiences, is practical for assessing the impact of sanitation improvements. Monitoring for SDG 6 focuses on toilet types, but achieving and sustaining progress on sanitation will require efficient resource allocation, which takes account of people's experiences, too. Understanding which programme designs and technologies are associated with the largest gains in SanQoL-5 can help to target investments to where they will see the greatest uptake and economic returns.

Methods

Study settings

We use data from previous studies in Ethiopia, India, Kenya, Malawi[29], Mozambique[47] and Zambia. The Ethiopia sample comprised 1,586 people from 24 communities (81% rural) across six districts (*woredas*) in three regions of the country. The India sample comprised 1,213 people from 60 communities (87% rural) representative of two states (Bihar and Uttar Pradesh), specifically two people (one male, one female) per household in 607 households. The Kenya sample comprised 1,000 people from 60 communities (71% rural) in a nationally representative sample of 600 households with a secondary respondent in 400 households. The Malawi sample comprised 1,400 people from 70 rural villages in Chiradzulu district. The Mozambique sample comprised 601 people, half from 24 urban blocks (*quarteirões*) in Maputo City and half from 18 blocks in the large town of Dondo in Sofala province. The Zambia sample comprised 365 people from nine rural villages in the Chongwe district. Our final sample includes 6,165 households across the four sites, representing heterogeneous geographies, cultures and sanitation infrastructure availability. Random sampling of households was used in different ways in all sites. Further details of underlying studies are in Supplementary Information I.

SanQoL-5 data and weighting

The SanQoL-5 questions are presented in Table 1. Answers are combined into a single score ranging from 0 to 1. Higher SanQoL-5 scores are better, with 1 denoting 'full sanitation capability' (maximum QoL) and 0 'no sanitation capability' (minimum QoL). With a three-level response to each of the five questions, there are 243 (= 3^5) possible combinations of SanQoL-5 attribute levels. The rationale for non-equal weights is that, in a given population, reduced disgust might hold greater value for people on average than improved privacy. These preferences are important to account for in the economic applications that SanQoL-5 is designed for, for example, benefit–cost analysis[48]. Therefore, rather than assuming that disgust has the same value as privacy, preferences can be elicited from the relevant population using methods such as discrete choice. Because the weights are elicited from people themselves, the SanQoL-5 index represents the value of sanitation to people in that population.

The set of preference weights for the 243 attribute combinations is known as a value set. Four of our studies apply value sets generated within the study, using a discrete choice experiment (Mozambique), attribute scoring (Malawi) and attribute ranking (Ethiopia and Zambia) (Supplementary Information J). The India and Kenya samples apply the discrete choice experiment value set. The SanQoL-5 index represents a given population's relative valuation of the attributes, so weights are typically slightly different in different countries, as with health-related QoL indices such as EQ-5D[20].

Overall study design

We apply a combination of classical test theory and IRT to assess different aspects of validity and reliability. First, we assessed construct validity—whether an instrument measures the construct it intends to measure. We took a predictive approach to construct validity, by testing hypotheses about how SanQoL-5 would covary with hypothesized variables. Second, we assessed convergent validity—whether two instruments aiming to measure similar constructs are correlated (an aspect of construct validity). We assessed this by correlation between SanQoL-5 and a SanVAS with scores ranging from 0 to 100 (Supplementary Information D)[25]. We used Spearman's rank correlation (r) because, like EQ-5D index values[49], SanQoL-5 index values are not usually normally distributed in a given population. We hypothesized that there would be moderate correlation ($0.4 > r < 0.6$), following norms for health VAS[50]. Third, we assessed discriminant validity (the opposite concept to convergent) by correlation between SanQoL-5 and the two EQ-5D questions (on mobility and pain) included in the India and Kenya questionnaires[21]. We used Spearman's rank correlation for the same reason, hypothesizing no correlation ($r = 0$). Fourth, we assessed known groups validity—whether an instrument can discriminate between two groups expected to differ in terms of the outcome (another aspect of construct validity). We explored this by assessing whether people with higher levels of sanitation service tended to have higher SanQoL-5 index values. Finally, we assessed internal reliability—how consistently different questions in a measure capture the same construct[35]. We assessed internal reliability using Cronbach's α (>0.7)[51] and item-total correlation (>0.4)[52]. In statistical tests, $P < 0.05$ in a two-tailed test was considered statistically significant evidence of association.

Hypotheses for construct validation

We prespecified hypotheses about the presence of associations between SanQoL-5 index values and a set of toilet characteristics (hypothesized variables)[53–55]. These were predominantly fieldworker observations of toilet characteristics including: walls being solid; faeces not being observed on the pan/slab; the pan/slab being concrete or similar; a water seal being present; the toilet having an inside lock; and the toilet not being shared with other households. Variables were binary coded such that positive regression coefficients are hypothesized (1 = better outcome, 0 = worse). For example, we hypothesized that solid walls are more likely to provide privacy and safety than makeshift or absent walls, and solid walls would have a positive correlation with SanQoL-5. In making hypotheses, we drew on the literature on sanitation and mental well-being, as well as motives for sanitation behaviours[17,18]. Further details and rationales for hypothesized variables are provided in Supplementary Information K.

We also included negative controls hypothesized not to be strongly associated with SanQoL-5 (ref. 56), namely household size and whether the respondent had a partner. These are imperfect, because we were limited by what was asked in the original surveys. For example, household size could influence SanQoL-5 if it means more people are sharing a toilet. However, we would not hypothesize household size to be a strong predictor of SanQoL-5 in samples of only around 1,000 people.

Construct validity analyses were completed for each country separately. We assessed a binary variable for a given country only if ≥15% of the sample with non-missing data was in each category, to ensure a minimum of statistical power. We tested hypotheses using generalized linear mixed models (GLMMs) in Stata 18. In India and Kenya, where there were two respondents per household, we used three-level GLMMs with random effects at the household and community level. In other countries, we used two-level GLMMs, with the exception of Zambia where there were only nine clusters, so we used wild bootstrap inference with linear regression[57]. We clustered standard

errors at the community level. We regressed on SanQoL-5 index values per country, including as a covariate each hypothesized variable in turn. We also explored the consequences of accounting for covariance between toilet characteristics, by including all hypothesized variables as covariates concurrently.

Item response theory

We used the graded response model (GRM) to assess the psychometric properties of each attribute and its contribution to the information function for unweighted SanQoL-5. GRM is widely used in the evaluation of health-related QoL measures because it allows polytomous variables, that is, with multiple attribute levels[58,59]. GRM is not part of the Rasch family because it allows discrimination to vary across items[35]. For IRT analyses, we pooled data across countries, as well as running models for individual countries where $n \geq 1{,}000$ (Ethiopia, India, Kenya and Malawi)[35]. Based on the GRM, we present item information and test information functions, as well as category characteristic curves.

Measurement invariance via DIF

For measures to be compared across countries or settings, it is important that there is equivalence of measurement and meaning. We explored measurement invariance using DIF by ordinal logistic regression because SanQoL-5 attributes are polytomous. Specifically, we followed the approach of Penton et al.[26] based on level sum scores (LSS), as recommended for EQ-5D[60]. LSS is the sum of attribute level scores and can be thought of as an unweighted SanQoL-5 index score (Supplementary Information G). With 6 countries, there were 15 possible country pairs. With 5 attributes, there were 75 possible instances of DIF overall. For each of the 15 pairs, we ran 2 models. For model 1, we ran ordinal logistic regression on each attribute score (ranging from 0 to 2) for those two countries only, including LSS as an independent variable. In model 2, we ran the same regression but also including a dummy variable for the two countries (for example, 0 for Kenya and 1 India). We calculated the difference in pseudo-R^2 between models 1 and 2, interpreting a difference of >2% as 'meaningful' DIF between those two countries (if the coefficient on country dummy had $P < 0.05$). This is the same cut-off used by Penton et al.[26] and earlier authors[61,62]. We took the more conservative approach by not first 'purifying' LSS as some studies do[63].

Comparing question framings

In the first two studies in which the SanQoL-5 was used[14,45], the questions had been framed such that 'always' was the best outcome. For example, 'Can you use the toilet without feeling disgusted?'. Mixed-methods cognitive and piloting work in support of the Zambia study identified this framing as challenging to understand in local languages without further explanation (as well as other languages spoken by the team, for example, Hindi). To facilitate a comparison, we included the old ('always = best') questions alongside the new/current question framing (Table 1) in Zambia. A third of the Ethiopia sample ($n = 506$), which undertook fieldwork at a similar time, were also asked both sets of questions. A further analysis in our present study was therefore comparing the performance of the 'always = best' and 'always = worst' framings, using the same validity and reliability methods as above. For example, we tested the construct validity hypotheses under the two question framings for the five SanQoL-5 attributes and compared results. For a fair comparison in Ethiopia, we compared results only for the $n = 506$ who completed both question formulations (rather than the full $n = 1{,}586$ sample)

Ethics

The Malawi study received prior approval from the National Committee on Research in the Social Sciences and Humanities (ref: NCST/RTT/2/6) in Malawi. The Mozambique study received prior approval from the Comité Institucional de Ética do Instituto Nacional de Saúde (ref: 028/CIE-INS/2023) in Mozambique. The Zambia study received prior approval from the University of Zambia Biomedical Research Ethics Committee (ref: UNZA-1389/2020). The India study received prior approval from Convergent Institutional Review Board (ref: 2023-24/019) in India. The Kenya study received prior approval from the AMREF Ethical and Scientific Review Committee (ref: P1508-2023) in Kenya. The Ethiopia data were collected as part of an internal evaluation by World Vision, who secured a prior approval letter from each district sampled for data collection. Use of the Ethiopia data was approved by the London School of Hygiene & Tropical Medicine (LSHTM) because anonymized data had been made openly available online by World Vision at https://osf.io/x5myz/ before this study commenced. The protocol covering Ethiopia and Zambia was approved by the LSHTM MSc Research Ethics Committee (Ref: 29049), while the LSHTM Observations/Interventions Research Committees approved the studies in India/Kenya (ref: 29640), Malawi (ref: 28249) and Mozambique (ref: 28190). Informed consent was obtained from all research participants before studies commenced. Participants were not compensated in Ethiopia, India, Kenya, Malawi or Zambia. In Mozambique, participants were given a paper calendar. This study was performed in line with the principles of the Declaration of Helsinki.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Datasets to reproduce results for all countries are available via the Open Science Framework at https://doi.org/10.17605/OSF.IO/N43FK (ref. 64).

Code availability

Code to reproduce results is available via the Open Science Framework at https://doi.org/10.17605/OSF.IO/N43FK (ref. 64).

References

1. *Progress on Household Drinking Water, Sanitation and Hygiene 2000–2022: Special Focus on Gender* (UNICEF & WHO, 2023).
2. Wolf, J. et al. Effectiveness of interventions to improve drinking water, sanitation, and handwashing with soap on risk of diarrhoeal disease in children in low-income and middle-income settings: a systematic review and meta-analysis. *Lancet* **400**, 48–59 (2022).
3. Garn, J. V. et al. The impact of sanitation interventions on latrine coverage and latrine use: a systematic review and meta-analysis. *Int. J. Hyg. Environ. Health* **220**, 329–340 (2017).
4. Ross, I. et al. How does sanitation influence people's quality of life? Qualitative research in low-income areas of Maputo, Mozambique. *Soc. Sci. Med.* **272**, 113709–113709 (2021).
5. Elmendorf, M. & Buckles, P. K. *Appropriate Technology for Water Supply and Sanitation: Sociocultural Aspects of Water Supply and Excreta Disposal* (World Bank, 1980).
6. Jenkins, M. & Curtis, V. Achieving the 'good life': why some people want latrines in rural Benin. *Soc. Sci. Med.* **61**, 2446–2459 (2005).
7. Mukherjee, N. *Achieving Sustained Sanitation for the Poor: Policy and Strategy Lessons from Participatory Assessments in Cambodia, Indonesia and Vietnam* (World Bank, 2001).
8. *Constitution of The World Health Organization* (WHO, 1948).
9. Wilson, I. B. & Cleary, P. D. Linking clinical variables with health-related quality of life: a conceptual model of patient outcomes. *JAMA* **273**, 59–65 (1995).
10. Young, S. L. et al. The Household Water InSecurity Experiences (HWISE) scale: development and validation of a household water insecurity measure for low-income and middle-income countries. *BMJ Global Health* **4**, e001750 (2019).

11. Bethancourt, H. J., Frongillo, E. A. & Young, S. L. Validity of an abbreviated Individual Water Insecurity Experiences (IWISE-4) scale for measuring the prevalence of water insecurity in low- and middle-income countries. *J. Water Sanit. Hyg. Dev.* https://doi.org/10.2166/washdev.2022.094 (2022).
12. Wutich, A. et al. Water sharing is a distressing form of reciprocity: shame, upset, anger, and conflict over water in twenty cross-cultural sites. *Am. Anthropol.* **124**, 279–290 (2022).
13. Stoler, J. et al. Cash water expenditures are associated with household water insecurity, food insecurity, and perceived stress in study sites across 20 low- and middle-income countries. *Sci. Total Environ.* **716**, 135881 (2020).
14. Ross, I. et al. Measuring and valuing broader impacts in public health: development of a sanitation-related quality of life instrument in Maputo, Mozambique. *Health Econ.* **31**, 466–480 (2021).
15. Sen, A. in *Tanner Lectures on Human Values* Vol. 1 (ed. McMurrin, S. M.) 197–220 (Cambridge Univ. Press, 1980).
16. Sen, A. in *The Quality of Life* (eds Sen, A. & Nussbaum, M.) 30–54 (Clarendon Press, 1993).
17. Sclar, G. D. et al. Exploring the relationship between sanitation and mental and social well-being: a systematic review and qualitative synthesis. *Soc. Sci. Med.* **217**, 121–134 (2018).
18. Novotný, J., Hasman, J. & Lepič, M. Contextual factors and motivations affecting rural community sanitation in low- and middle-income countries: a systematic review. *Int. J. Hyg. Environ. Health* **221**, 121–133 (2018).
19. Goodwin, E. & Green, C. A systematic review of the literature on the development of condition-specific preference-based measures of health. *Appl. Health Econ. Health Policy* **14**, 161–183 (2016).
20. Brazier, J., Ratcliffe, J., Salomon, J. & Tsuchiya, A. *Measuring and Valuing Health Benefits for Economic Evaluation* (Oxford Univ. Press, 2016).
21. EQ-5D-5L Health Questionnaire. *EuroQol Research Foundation* https://euroqol.org/wp-content/uploads/2023/12/Sample_UK-English-EQ-5D-5L-Paper-Self-Complete-v1.2-ID-24700.pdf (2009).
22. Magasi, S. et al. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. *Qual. Life Res.* **21**, 739–746 (2012).
23. Drummond, M., Stoddard, G. L. & Torrance, G. W. *Methods for the Economic Evaluation of Health Care Programmes* 4th edn (Oxford Univ. Press, 2015).
24. Caruso, B. et al. *Priority Gender-Specific Indicators for WASH Monitoring under SDG Targets 6.1 and 6.2: Recommendations for National and Global Monitoring* (UNICEF & WHO, 2024).
25. Cheung, H. H. et al. Validity of a visual analogue scale to measure and value perceived level of sanitation—evidence from Ghana and Mozambique. *Health Policy Plan* https://doi.org/10.1093/heapol/czae092 (2024).
26. Penton, H., Dayson, C., Hulme, C. & Young, T. An investigation of age-related differential item functioning in the EQ-5D-5L using item response theory and logistic regression. *Value Health* **25**, 1566–1574 (2022).
27. Feng, Y. S., Kohlmann, T., Janssen, M. F. & Buchholz, I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. *Qual. Life Res.* **30**, 647–673 (2021).
28. Castillo-Díaz, M. & Padilla, J.-L. How cognitive interviewing can provide validity evidence of the response processes to scale items. *Soc. Indic. Res.* **114**, 963–975 (2013).
29. Chidziwisano, K. et al. The Water, Sanitation, and Hygiene (WASH) for everyone intervention study: protocol for a controlled before-and-after trial. *JMIR Res. Prot.* https://doi.org/10.2196/68280 (2024).
30. Ross, I. et al. Impact of a sanitation intervention on quality of life and mental wellbeing in low-income urban neighbourhoods of Maputo, Mozambique. *BMJ Open* https://doi.org/10.1136/bmjopen-2022-062517 (2022).
31. Bick, S. et al. Impact of a school-based water and hygiene intervention on child health and school attendance in Addis Ababa, Ethiopia: a cluster-randomised controlled trial. *BMC Med.* **22**, 348 (2024).
32. Lewis, A. R. et al. Comparative sanitation data from high-frequency phone surveys across 3 countries. *Data Brief.* **55**, 110635 (2024).
33. Ross, I. *Measuring and Valuing Quality of Life in the Economic Evaluation of Sanitation Interventions*. PhD thesis, London School of Hygiene & Tropical Medicine (2021); https://doi.org/10.17037/PUBS.04661119
34. Meili, D. et al. Indicators for sanitation quality in low-income urban settlements: evidence from Kenya, Ghana, and Bangladesh. *Soc. Indic. Res.* **162**, 683–720 (2021).
35. Fayers, P. & Machin, D. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes* 3rd edn (Wiley, 2015).
36. Novotný, J., Humňalová, H. & Kolomazníková, J. The social and political construction of latrines in rural Ethiopia. *J. Rural Stud.* **63**, 157–167 (2018).
37. Tidwell, J. B. et al. Where shared sanitation is the only immediate option: a research agenda for shared sanitation in densely populated low-income urban settings. *Am. J. Trop. Med. Hyg.* https://doi.org/10.4269/ajtmh.20-0985 (2020).
38. *World Population Prospects 2022: Summary of Results* (UN DESA, 2022).
39. Nallari, A. 'All we want are toilets inside our homes!': the critical role of sanitation in the lives of urban poor adolescent girls in Bengaluru, India. *Environ. Urb.* **27**, 73–88 (2015).
40. Sahoo, K. C. et al. Sanitation-related psychosocial stress: a grounded theory study of women across the life-course in Odisha, India. *Soc. Sci. Med.* **139**, 80–89 (2015).
41. Wille, N. et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual. Life Res.* **19**, 875–886 (2010).
42. Sinharoy, S. S., McManus, S., Conrad, A., Patrick, M. & Caruso, B. A. The Agency, Resources, and Institutional Structures for Sanitation-related Empowerment (ARISE) Scales: development and validation of measures of women's empowerment in urban sanitation for low- and middle-income countries. *World Dev.* **164**, 106183 (2023).
43. Tsai, A. C. et al. Population-based study of intra-household gender differences in water insecurity: reliability and validity of a survey instrument for use in rural Uganda. *J. Water Health* **14**, 280–292 (2016).
44. Coates, J. C., Webb, P., Houser, R. F., Rogers, B. L. & Wilde, P. 'He said, she said': who should speak for households about experiences of food insecurity in Bangladesh? *Food Secur.* **2**, 81–95 (2010).
45. Tidwell, J. B., Nyarko, K. B., Ross, I., Dwumfour-Asare, B. & Scott, P. Evaluation of user experiences for the Clean Team Ghana container-based sanitation service in Kumasi, Ghana. *J. Water Sanit. Hyg. Dev.* https://doi.org/10.2166/washdev.2022.013 (2022).
46. Brown, J. et al. The effects of racism, social exclusion, and discrimination on achieving universal safe water and sanitation in high-income countries. *Lancet Glob. Health* **11**, e606–e614 (2023).
47. Katana, P. V. et al. Valuing an index of sanitation-related quality of life in urban Mozambique: a discrete choice experiment. *Value Health Reg. Issues* **47**, 101087 (2025).
48. Whitehead, S. J. & Ali, S. Health outcomes in economic evaluation: the QALY and utilities. *Br. Med. Bull.* **96**, 5–21 (2010).
49. Parkin, D., Devlin, N. & Feng, Y. What determines the shape of an EQ-5D index distribution? *Med. Decis. Making* **36**, 941–951 (2016).

50. Whynes et al. Correspondence between EQ-5D health state classifications and EQ VAS scores. *Health Qual. Life Outcomes* **6**, 94 (2008).
51. Nunnally, J. C. *Psychometric Theory* 2nd edn (McGraw Hill, 1978).
52. Ware, J. E., Brook, R. H., Williams, K. N., Stewart, A. L. & Davies-Avery, A. *Conceptualisation and Measurement of Health for Adults in the Health Insurance Study. Vol. 1: Model of Health and Methodology* (Rand Corporation, 1980).
53. Savoia, E., Fantini, M. P., Pandolfi, P. P., Dallolio, L. & Collina, N. Assessing the construct validity of the Italian version of the EQ-5D: preliminary results from a cross-sectional study in North Italy. *Health Qual. Life Outcomes* **4**, 47 (2006).
54. Al-Janabi, H. et al. An investigation of the construct validity of the ICECAP-A capability measure. *Qual. Life Res.* **22**, 1831–1840 (2013).
55. Smith, G. T. On construct validity: issues of method and measurement. *Psychol. Assess.* **17**, 396–408 (2005).
56. Arnold, B. F. & Ercumen, A. Negative control outcomes: a tool to detect bias in randomized trials. *JAMA* **316**, 2597–2598 (2016).
57. Cameron, A. C., Gelbach, J. B. & Miller, D. L. Bootstrap-based improvements for inference with clustered errors. *Rev. Econ. Stat.* **90**, 414–427 (2008).
58. Reeve, B. B. et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med. Care* **45**, S22–S31 (2007).
59. Liu, T. H., Ho, A. D., Hsu, Y. T. & Hsu, C. C. Validation of the EQ-5D in Taiwan using item response theory. *BMC Public Health* **21**, 2305 (2021).
60. Devlin, N., Parkin, D. & Janssen, B. *Methods for Analysing and Reporting EQ-5D Data* (Springer, 2020).
61. Bjorner, J. B., Kosinski, M. & Ware, J. E. Jr. Calibration of an item pool for assessing the burden of headaches: an application of item response theory to the headache impact test (HIT). *Qual. Life Res.* **12**, 913–933 (2003).
62. Keetharuth, A. D. et al. An item response theory analysis of an item pool for the recovering quality of life (ReQoL) measure. *Qual. Life Res.* **30**, 267–276 (2021).
63. Scott, N. W. et al. Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health Qual. Life Outcomes* **8**, 81 (2010).
64. Ross, I. Data for validity and reliability of the Sanitation-related Quality of Life index (SanQoL-5) in six countries. *Open Science Framework* https://doi.org/10.17605/OSF.IO/N43FK (2025).

Acknowledgements

This work received no specific funding. The data collection underlying this work was funded as follows: (1) Ethiopia—World Vision internal evaluation; (2) India and Kenya—Bill and Melinda Gates Foundation (INV-058166) (3) Malawi—World Vision external evaluation (34730); (4) Mozambique—Bill and Melinda Gates Foundation (OPP1137224); (5) Zambia—World Vision external evaluation. F.A. conducted analysis of the Ethiopia data as part of a MSc dissertation at LSHTM funded by a Chevening scholarship from the UK government's Foreign, Commonwealth and Development Office. I.R. acknowledges the support of a postdoctoral fellowship from the Reckitt Global Hygiene Institute in the period when the paper was drafted. We greatly appreciated the cooperation of survey participants, as well as the efforts of the fieldworkers in each underlying study.

Author contributions

We use the Contributor Roles Taxonomy (CRediT) to indicate the contributions of each author. F.A.: methodology, data curation, writing—review and editing, and formal analysis. N.B.: writing—review and editing. I.C.: writing—review and editing. K.C.: writing—review and editing. J.C.: writing—review and editing. C.C: writing—review and editing. O.C.: writing—review and editing. R.D.: methodology, and writing—review and editing. P.V.K.: writing—review and editing. C.M.: writing—review and editing. M.P.: writing—review and editing. P.R.: writing—review and editing. A.S.: methodology, and writing—review and editing. S.S.: writing—review and editing. A.T.: writing—review and editing. J.B.T.: writing—review and editing. E.V.: writing—review and editing. B.W.: writing—review and editing. I.R.: conceptualization, methodology, data curation, investigation, software, formal analysis, visualization, writing—original draft, and writing—review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44221-025-00434-7.

Correspondence and requests for materials should be addressed to Ian Ross.

Peer review information *Nature Water* thanks Josef Novotný and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Ian Ross

Last updated by author(s): Mar 24, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	n/a
Data analysis	Code to reproduce results is available at https://doi.org/10.17605/OSF.IO/N43FK

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Datasets to reproduce results for all countries are available at https://doi.org/10.17605/OSF.IO/N43FK

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	In the countries these data were collected (especially in face-to-face surveys as we have done) it is generally considered inappropriate to ask people about their gender identity. Instead, fieldworkers assign a gender to the participant based on how they present in relation to local gender norms. Hence we refer to gender and not sex. Samples were approximately gender-balanced in all countries apart from Malawi, where 83% of respondents were women. Overall findings therefore apply to women and men. Gender is a variable in all 6 datasets (available online – see above). In this manuscript, the aim is to investigate validity in broad and diverse samples, so relatively equal numbers of women and men is desired. It is not common in validity research to investigate for sub-groups separately. Therefore, we do not undertake gender-based analyses in this study. Gender inequality in outcomes is a completely different matter, and is discussed in the manuscript as a future research priority.
Reporting on race, ethnicity, or other socially relevant groupings	We do not report any data on race, ethnicity, or other socially relevant groupings
Population characteristics	See "Behavioural & social sciences" questions
Recruitment	Recruitment varied across the 6 underlying secondary datasets. In general, however, the fieldworker would knock at the door of a household and enquire whether someone was available to be interviewed. If the protocol included quotas (e.g. 50/50 men/women, or by toilet type, as in Mozambique) then a person fitting the quota would be asked for. It is plausible that people who were more likely to be at home at the time of interview were more likely to be selected than the general population. We are not conducting an intervention trial, and achieving approximate gender balance was our main concern, which was achieved.
Ethics oversight	The Malawi study received prior approval from the National Committee on Research in the Social Sciences and Humanities (ref: NCST/RTT/2/6) in Malawi. The Mozambique study received prior approval from the Comité Institucional de Ética do Instituto Nacional de Saúde (ref: 028/CIE-INS/2023) in Mozambique. The Zambia study received prior approval from the University of Zambia Biomedical Research Ethics Committee (ref: UNZA-1389/2020). The India study received prior approval from Convergent IRB (ref: 2023-24/019) in India. The Kenya study received prior approval from the AMREF Ethical and Scientific Review Committee (ref: P1508-2023) in Kenya. The Ethiopia data were collected as part of an internal evaluation by World Vision, who secured a prior approval letter from each district sampled for data collection. Use of the Ethiopia data was approved by LSHTM since anonymised data had been made openly available online by World Vision at https://osf.io/x5myz/ before this study commenced. The protocol covering Ethiopia and Zambia was approved by the LSHTM MSc Research Ethics Committee (Ref: 29049), while the LSHTM Observations/Interventions Research Committees approved the studies in India/Kenya (Ref: 29640), Malawi (Ref: 28249) and Mozambique (Ref: 28190).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative cross-sectional
Research sample	The sample is 6,165 people across rural and urban areas of six countries. Ethiopia (n=1,586), India (n=1,213), Kenya (n=1,000), Malawi (n=1,400), Mozambique, (n=601), Zambia (n=365). Women are 50-55% of participants in all countries except Malawi (84%). The mean age ranges from 35-43 across the countries – full age categorisation is reported in Table 2. All data were secondary datasets collected for different purposes as outlined in the methods section. The Ethiopia sample comprised 1,586 people from 24 communities (81% rural) across six districts (woredas) in three regions of the country. Data were collected by World Vision. The India sample comprised 1,213 people from 60 communities (87% rural) representative of two states (Bihar and Uttar Pradesh), specifically two people (one male, one female) per household in 607 households. Data were collected by OPM India. The Kenya sample comprised 1,000 people from 60 communities (71% rural) in a nationally representative sample of 600 households with a secondary respondent in 400 households. Data were collected by Ipsos Kenya. The Malawi sample comprised 1,400 people from 70 rural villages in Chiradzulu district. Data were collected by the Malawi University of Business and Applied Sciences The Mozambique sample comprised 601 people, half from 24 urban blocks (quarteirões) in Maputo City and half from 18 blocks in

the large town of Dondo in Sofala province. Data were collected by the Instituto Nacional de Saude..
The Zambia sample comprised 365 people from nine rural villages in the Chongwe district. Data were collected by the Centre for Infectious Disease Research in Zambia

Sampling strategy

Since in all countries, sample sizes were selected for the purpose of the original study, no sample size calculation is appropriate for the purpose of our study. A minimum of 200 participants is often recommended for construct validity testing (Fayers & Machin, 2016), and sample sizes in all countries exceed this. Further details of sampling strategy in each country follows below.

Ethiopia

Study Design
The Ethiopia study was a cross-sectional survey to assess the reach of Water, Sanitation and Hygiene (WASH) Business Centres (WBCs) on household level access to WASH resources. A report ranking WBCs was used to select six WBCs, which are based in the main town of a district (woreda). Three high-performing and three medium-performing WBCs were selected, two in each of three regions of Ethiopia. In each of the six sampled districts, four communities (kebeles) adjacent to and including the WBC community were sampled, leading to a total of 24 communities. Data were collected by World Vision Ethiopia.

Participants
A total of 1586 households were sampled (81% in rural areas), employing systematic random sampling of every fifth household, starting from the centre of the community. The inclusion criteria for participants were: (A) aged 18 years and above; (B) self-reports as the main decision maker regarding toilet construction and maintenance; (C) resident in sampled community; and, (D) available and willing to be interviewed. No sample size calculation was undertaken.

India

Study Design
The India study was a cross-sectional survey to evaluate intra-household gender differences in SanQoL-5. Thirty enumeration areas (EAs) were randomly sampled in each state (60 in total), accounting for level of urbanisation and agro-climatic zones (3 in Bihar, 5 in UP). In Bihar, three districts (Bhojpur, Darbhanga and Sheohar) were randomly sampled and, within each, nine rural EAs and one urban EA. To sample the rural EAs in each district, Primary Health Centres were sampled and, from each, one Sub-Centre was sampled. From each Sub-Centre catchment, two villages were sampled. To sample an urban EA in each district, a ward within the district headquarters was randomly sampled. In UP, five districts (Azamgarh, Bijnor, Farrukhabad, Hamirpur, and Rae Bareilly) were randomly sampled and, within each, five rural EAs and one urban EA. These states were selected because of their diversity in types of sanitation. The sample for primary respondents aimed to be representative of each state. Data were collected by OPM India.

Participants
A total of 1,213 people were sampled from 607 households (87% in rural areas) across two states (Bihar and Uttar Pradesh). Eligible primary respondents were anyone aged 18 or over available to be interviewed, with a 50/50 quota of women and men per enumeration area. The total number of residential dwellings was determined via the village head or ward counsellor, and that number divided by 10 (primary respondents per EA) to determine the sampling interval. The first household was sampled near a prominent landmark, and subsequent households sampled in a clockwise direction. Where there was >1 eligible respondent, a “next birthday” rule was used. Eligible secondary respondents were someone of the opposite sex to the primary respondent, but resident in the same household. The person closest in age as possible to the primary respondent was sampled.

Kenya

Study Design
The Kenya study was a cross-sectional survey to evaluate intra-household gender differences in SanQoL-5. Sixty enumeration areas were randomly sampled to be nationally representative. The first stage was to undertake a stratified random sample of sub-counties to reflect the national urban/rural population distribution (71% rural). For example, 6 sub-counties were sampled from Nairobi city, and 3 from Nakuru, which are amongst the most populous and urbanised counties. Only 1 sub-county was sampled from each of Vihiga and Garissa, which are amongst the smaller counties by population. The enumeration areas were villages or urban blocks, and 60 randomly sampled within sub-counties, with 10 households on average sampled per enumeration areas (10*60=600). Of Kenya’s 47 counties, 36 were randomly sampled and EAs sampled within them. Data were collected by Ipsos Kenya.

Participants
A total of 1,000 people were sampled from 600 households (71% in rural areas) in a nationally representative sample of 600 households with a secondary respondent in 400 households. Eligible primary respondents were anyone aged 18 or over available to be interviewed, with a 50/50 quota of women and men per enumeration area. Eligible secondary respondents were someone of the opposite sex to the primary respondent, but resident in the same household. The person closest in age as possible to the primary respondent was sampled.

Malawi

Study Design
The Malawi study was a controlled before-and-after (CBA) trial, with the present study using data from the baseline survey. The study included 3 Traditional Authorities (TAs), or sub-districts, in Chiradzulu district. For the first and second TAs (which were to receive the intervention), 20 clusters (villages) were randomly selected each. In the third TA (control) 30 clusters were randomly selected. Data were collected by the Malawi University of Business and Applied Sciences.

Participants
In each cluster, 20 households randomly selected in March-April 2023, providing a total of 1400 households. Inclusion criteria were that households needed at least one adult aged 18 who able to provide consent, and the head of the household needed to be a permanent resident.

Mozambique

Study Design
This study was primarily designed as a discrete choice experiment to explore how people trade off attributes of SanQoL. The study took place in the city of Maputo (population 1.1 million) and the large town of Dondo (population 100,000). Neighbourhoods (bairros) of the two conurbations selected on the basis of achieving a diversity in type of toilet used. Neighbourhoods selected in Maputo were both in the Polana Caniço area, while those in Dondo were Macharote and Nhamainga. Data were collected by the Instituto Nacional de Saúde.

Participants
The study aimed to achieve gender balance and diversity in toilet type. The study population was adults aged 18+. In Maputo, households were sampled based on data from the Health and Demographic Surveillance System (HDSS) in Polana Caniço. The study team generated lists of households according to HDSS data on toilet type. They then visited households in order of a randomly reordered list until quotas by toilet type and gender were reached, in April-May 2023. In Dondo, a similar approach was followed based on sanitation data from the Census of the Catchment Population.

Zambia

Study Design
The study in Zambia was a cross-sectional study in the Kapululwe area. A two-stage cluster sampling design was used in which the first stage consisted of selecting nine villages based on probability proportion to size. Data were collected by the Centre for Infectious Disease Research in Zambia.

Participants
40 households were sampled from each village in March-April 2021, by sampling every other household after spinning a pen to identify the initial direction. 365 households were selected in total.

Data collection

Data collection was via tablet computers, using OpenDataKit software or similar. It was not intended that anybody was present except the participant and researcher, but we cannot exclude the possibility that children or family members were sometimes listening. There was no experimental condition. Fieldworkers were not aware of the hypotheses we test in the present study, but they knew the studies were about sanitation.

Timing

Ethiopia: December 2021 - January 2022
India: November - December 2023
Kenya: December 2023 - January 2024
Malawi: March - April 2023
Mozambique: April - May 2023
Zambia: March - April 2021

Data exclusions

We did not exclude any participants.

Non-participation

We do not have data on response rate for many of the secondary datasets, with the exception of Mozambique where it was 99%. People in predominantly rural and/or low-income settings are often generous with their time. Since these were cross-sectional studies there is no "loss to follow-up".

Randomization

This is a cross-sectional study, and there are no experimental groups. The rationale for the 6 binary covariates hypothesised to be associated with SanQoL-5 are reported in Supplementary Material K, alongside the exactly wording of the relevant questions. We only included a covariate in a model for a given country if ≥15% of the sample with non-missing data was in each category, to ensure a minimum of statistical power.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	no plants used - seems an error within form
Novel plant genotypes	no plants used - seems an error within form
Authentication	no plants used - seems an error within form