Measuring and valuing broader impacts in public health: Development of a sanitation-related quality of life instrument in Maputo, Mozambique

Ian Ross¹,² | Giulia Greco¹,³ | Charles Opondo⁴,⁵ | Zaida Adriano⁶ | Rassul Nala⁷ | Joe Brown⁸ | Robert Dreibelbis² | Oliver Cumming²

¹Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London, UK
²Department of Disease Control, London School of Hygiene and Tropical Medicine, London, UK
³MRC/UVRI & LSHTM Uganda Research Unit, Kampala, Uganda
⁴Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK
⁵Nuffield Department of Population Health, University of Oxford, Oxford, UK
⁶WE Consult, Maputo, Mozambique
⁷Instituto Nacional de Saúde, Maputo, Mozambique
⁸Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill, North Carolina, USA

Abstract
Two billion people globally lack access to a basic toilet. While improving sanitation reduces infectious disease, toilet users often identify privacy, safety and dignity as more important. However, these outcomes have not been incorporated in sanitation-related economic evaluations. This illustrates the general challenge of outcome measurement and valuation in the economic evaluation of public health interventions, and risks misallocating the US$ 20 billion invested in sanitation in low- and middle-income countries every year. In this study in urban Mozambique, we develop an instrument to measure sanitation-related quality of life (SanQoL). Applying methods from health economics and the capability approach, we develop a descriptive system to measure five attributes identified in prior qualitative research: disgust, health, shame, safety and privacy. Sampling individuals from the intervention and control groups of a sanitation intervention trial, we elicit attribute ranks to value a SanQoL index and assess its validity and reliability. In combination with a measure of time using a sanitation service, SanQoL can quantify incremental benefits in a sanitation-focused cost-effectiveness analysis. After monetary valuation based on willingness to pay, QoL benefits could be summed with health gains in cost-benefit analysis, the most common method in sanitation economic evaluations.
1 | INTRODUCTION

 Nearly two billion people globally lack access to basic sanitation (WHO & UNICEF, 2021). This failure to separate human excreta from human contact represents an enormous public health challenge. Every year 26 million disability-adjusted life years (DALYs), including 430,000 child deaths from diarhoea, are attributable to inadequate sanitation (Prüss-Ustün et al., 2019). From a health system perspective, averting this disease burden would be highly valuable. Toilet users, however, often identify benefits of improved sanitation for privacy, safety, reduced smells, and dignity, as more important than reducing infectious disease (Elmendorf & Buckles, 1980; Jenkins & Curtis, 2005; Mukherjee, 2001). These are attributes of good quality of life (QoL), and contribute to health in its broadest sense including mental and social wellbeing (WHO, 1948). Therefore, the QoL benefits beyond infectious disease are likely to underpin household willingness to pay (WTP) for sanitation improvements, with consequences for their measurement and valuation in cost-benefit analysis (CBA).

 Studies have identified associations between better sanitation and general measures of mental wellbeing (Caruso et al., 2018; Gruebner et al., 2012). However, the QoL outcomes such as privacy and safety are not captured in DALYs. Quality-adjusted life years (QALYs) are less-used in lower- and middle-income countries (LMICs), and have never been used to evaluate sanitation interventions. In any case, QALYs would also be unlikely to capture benefits of sanitation beyond morbidity and mortality from infectious disease. This is because changes in the QoL outcomes prioritised by toilet users would not be picked up by measures widely used to weight QALYs such as the EQ-5D, with the exception of consequences for anxiety/depression (Brazier et al., 2016; Euroqol Group, 2009).

 Despite the importance of outcomes such as privacy and safety to users, economic evaluations of sanitation interventions have never managed to specifically include them, in the absence of means for their measurement and valuation (Hutton & Chase, 2016). The only way in which sanitation-related QoL outcomes have been valued economically is through WTP for toilets (Hutton & Chase, 2016). Willingness to pay for benefit valuation is often considered dissatisfactory in a health context due to the biases it introduces (Cookson, 2003), particularly from ability to pay (Coast, Smith, & Lorgelly, 2008). The bias is likely to be toward under-valuation if a respondent’s WTP for a toilet is taken to comprise the total benefits of sanitation (e.g., morbidity, mortality, healthcare savings, time savings, privacy, safety, etc.). Accurate estimation is more likely if these benefits are estimated separately and aggregated, as is preferred in CBA (Boardman et al., 2018), and in practice all sanitation CBA studies do this (Hutton & Chase, 2016).

 No quantitative measure capturing sanitation interventions’ QoL outcomes with a user-derived valuation scheme exists (Sclar et al., 2018). The only standalone measure of such outcomes is the women’s sanitation insecurity profile, which comprises 60 survey questions spread over seven factor scales measuring a mixture of practices, experiences and feelings (Caruso et al., 2017) While it is appropriate for non-economic purposes, the women’s sanitation insecurity profile’s application of equal weighting and lack of an overall score preclude its use in economic evaluation (Brazier et al., 2016). It was also developed with (and designed for) women only, while sanitation affects the general population. To inform the allocation of public funds, a respondent-weighted index with a single overall score designed for use in the general population is needed.

 The lack of an outcome measure applicable for use in economic evaluation may be contributing to misallocation of the US$ 20 billion invested annually in sanitation in LMICs (WHO, 2017). Health economists are increasingly using capability-based outcome measures as a way to broaden the evaluative space beyond the value of health alone (Coast, Smith, & Lorgelly, 2008; Greco et al., 2016). The capability approach focuses on the value of what people are able to be and do (Sen, 1980, 1993) and is well-suited to a multi-dimensional conceptualisation of QoL (Coast et al., 2015). Many capability-based measures have been developed for different economic evaluation purposes both in high-income and low-income settings (Coast, 2019; Greco, 2016; Simon et al., 2013).

 In Mozambique, where our study took place, 39% of the urban population (77% in rural) do not have basic sanitation, defined as “improved” facilities which are not shared with other households (WHO & UNICEF, 2021). A third of those use a shared improved facility. Urban sanitation services are a municipal responsibility and each city has policies and
strategies in place (CMM, 2017; MOPH, 2011), but access to basic sanitation areas is increasing at only 1.5 percentage points per year (WHO & UNICEF, 2021).

In this study, we aim to develop and value a measure of sanitation-related quality of life (SanQoL) in urban Mozambique, based on prior qualitative research in the setting (Ross et al., 2021). The underlying objective was to enable classification of “sanitation states” and their valuation, in support of economic evaluations of sanitation interventions. We include exploration of the measure's validity and reliability. The study was nested within the Maputo Sanitation (MapSan) trial (clinicaltrials.gov registration: NCT02362932), which evaluated the impact of a shared urban sanitation intervention on children’s enteric infections (Knee et al., 2021).

2 | METHODS

2.1 | Study setting

We undertook this study in Maputo, Mozambique, in the low-income neighbourhoods of the Nhlamankulu district where multi-household compounds with shared sanitation facilities are common (Brown et al., 2015). Of Maputo City’s 1.1 million population, 70% live in informal settlements (INE, 2019; UN-HABITAT, 2010). Approximately 89% use non-sewered sanitation facilities (Hawkins & Muximpua, 2015). In the control group of the MapSan trial, households used low-quality shared pit latrines, which are common in low-income areas. These latrines typically comprised unlined pits with squatting slabs made of wood or car tyres, and no water seal (u-bend) to provide a barrier against smells and flies. Few pit latrines have roofs, and the walls are often made with sections of scrap corrugated iron or plastic sheeting, with makeshift fabric doors. Together, these conditions lead to unpleasant, unsafe toilets providing little privacy.

In the intervention group, an international non-government organisation provided a subsidised pour-flush toilet with concrete superstructure, discharging to a septic tank. The intervention toilets had two design types, depending on user numbers. The first was a shared toilet with one stance (cubicle) to be used by around 15 people, at 85% subsidy. The second was a Community Sanitation Block with two stances, to be used by a minimum of 21 people, at 90% subsidy. Both STs and CSBs were built from concrete blocks, with metal doors lockable from the inside. More information on the setting, including photos of toilet types, is provided in Supplementary Material A. Our study received approval from the Comité Nacional de Bioética para a Saúde (IRB00002657) at the Ministry of Health in Mozambique and the ethics committee at the London School of Hygiene & Tropical Medicine (Ref: 14609). Informed, written consent was obtained from all participants. Deidentified individual participant data, codebook, and replication code are available online (Ross, 2021).

2.2 | Overall study design

We followed measure development methods common in health economics for the purpose of economic evaluation (Goodwin & Green, 2016), whereby the final measure comprises index values anchored at 0 and 1 (Drummond et al., 2015). The descriptive system comprises a set of items with categorical response scales, each representing one dimension of the construct being measured (Brazier et al., 2016). Items are selected primarily for content validity (Fayers & Machin, 2015), and an individual’s combination of responses under the descriptive system comprises their “health state” (Brazier et al., 2016) or “capability state” (Coast, Flynn, et al., 2008). Methods such as factor analysis are not applied, and instead states are valued as an index using preference elicitation.

We adapted these methods to develop a measure of SanQoL. The capability approach to welfare economics (Sen, 1980, 1993) comprised our overarching theoretical approach to defining and measuring QoL, which informed both qualitative and quantitative methods. The target population for our measure is people living in urban settings where poor sanitation is common. The primary intended use is in economic evaluation of sanitation programmes. This required the measure to have a small number of items, since trading-off more than seven attribute levels simultaneously in preference elicitation tasks is generally considered too cognitively demanding (Hensher et al., 2015; Mangham et al., 2009).

Initial qualitative research in the same Maputo setting defined the construct to be measured as “the subset of overall QoL which is directly affected by sanitation practices or services” (Ross et al., 2021). This draws on common definitions of health-related QoL (HRQoL), and is aligned with the third of the four types identified by Karimi and Brazier (2016) in their review of definitions of health and HRQoL. Specifically we build on how Peasgood et al. (2014, p. 17) describe narrower definitions of HRQoL: “the sub-set of the important or most common ways in which health or health care impact
upon well-being.” The scope of sanitation practices is as perceived by users but is assumed to extend beyond defecation and urination to include for example menstrual hygiene, as well as any related practices users consider important, such as bathing. The capability approach was reflected in the design of the qualitative study (Ross et al., 2021). For example, the topic guides focused on a broad evaluative space (“a good life”), we used focus groups to engender the deliberation encouraged by the capability approach, and we triangulated findings on attributes’ relative importance using methods from cognitive anthropology (Weller & Romney, 1988).

The qualitative research resulted in a conceptual model (Figure 1) centred on five attributes: disgust, health, privacy, safety and shame. These were the attributes toilet users in this setting considered most important, and are largely consistent with the broader qualitative literature in other countries and settings (Novotný et al., 2018). Some studies have identified convenience or water supply, for example, as important (Jenkins & Curtis, 2005; Sahoo et al., 2015). However, these were not raised by our participants, reflecting that the social and environmental context in which a toilet “commodity” (Sen, 1985) is used affects the capabilities an individual can derive from it (Ross et al., 2021).

In this paper, we present the subsequent stages of measure development (Table 1), adhering to the standards of the International Society for QoL Research for patient-reported outcome measures (Supplementary Material B) (Reeve et al., 2013). We conducted all analyses in Stata 16 (StataCorp, 2019).

2.3 Item development and piloting

We developed several possible items for each of the five attributes identified by the qualitative research (Streiner et al., 2015), aiming to focus attention on capabilities by framing questions with “can you...?” or “are you able to...?” A long-list of 36 items was reviewed internally, then by 14 external experts. We undertook piloting and cognitive interviewing during April 2019, recruiting a team of four enumerators fluent in Portuguese and Changana, the first and second most commonly spoken languages in Maputo. One author (Z. A.), a Mozambican fluent in Portuguese and English, translated items into Portuguese. (I. R.), an intermediate Portuguese speaker, extensively discussed meanings with (Z. A.) and the enumerator team. We piloted items with 64 individuals from the target population, and undertook further cognitive interviewing with 28 of the pilot respondents (Bowden et al., 2002). In cognitive interviews, after each SanQoL item, the respondent was asked to explain back the question in their own words and discuss the ease of understanding the question. Enumerators rated their explanation on a scale of full, partial or no comprehension. For each of the five attributes, we identified one item which best achieved face and content validity, considering the qualitative findings and cognitive interviews. These five items comprise the descriptive system, with responses on a four-level ordinal frequency scale: always, sometimes, rarely, never (Table 2, and in Portuguese in Supplementary Material C). The descriptive system
Therefore contains 1024 (=4^5) combinations of attribute levels, or “sanitation capability states.” Items were framed as direct questions such that “always” was consistently the best outcome.

### Valuation

Valuation aims to aggregate item responses into a single score, weighted by the relative value of attributes elicited from the target population. We undertook a quantitative survey in May 2019 with the same enumerator team, using the mWater (2019) application on smartphones. We aimed to recruit at least 400 people aged over 18 living on trial-enrolled compounds, stratified by intervention status and gender. This meant a range of outcomes was likely, due to the diversity of low-quality toilet types in control compounds and the high-quality toilets in use in intervention compounds. We recruited a maximum of two people (one man, one woman) from each compound, on condition that they were not from the same household. In addition to SanQoL items, we collected data on water supply and sanitation services, demographic characteristics, and calculated an asset-based wealth index following standard methods (Vyas & Kumaranayake, 2006).

Finally, we asked respondents to rank the five SanQoL attributes on a visual analogue scale, according to their relative importance (Supplementary Material D).

With a view to subsequent valuation by discrete choice experiment, our interim valuation strategy was based on the rank sum method (Stillwell et al., 1981), following previous measures (de Kruijk & Rutten, 2007; Greco, 2016). This required scoring attributes on a 0-3 scale based on item responses (0 = never, 1 = rarely, 2 = sometimes, 3 = always), and estimating attribute weights from the rank data (Equation 1). We calculated SanQoL index values by combining attribute scores and weights, then rescaling to anchor at 0 and 1 (Equation 2). Zero represents “no sanitation capability” and one represents “full sanitation capability,” building on the Investigating Choice Experiments and Capabilities measures’ anchor points (Coast, Flynn, et al., 2008).

**Equation 1** – attribute weights for a population:

\[
W_j = \frac{N - R_j + 1}{\sum_{i=1}^{N} (N - R_i + 1)}
\]

### TABLE 1 Stages followed in measure development

<table>
<thead>
<tr>
<th>Stage</th>
<th>Objective</th>
<th>Method</th>
<th>Main output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conceptual model</td>
<td>To identify what people most value about sanitation, using qualitative research</td>
<td>In-depth interviews ((n = 19)) and focus groups ((n = 8))</td>
<td>Conceptual model</td>
</tr>
<tr>
<td>2. Piloting</td>
<td>To ensure items are understood, and identify floor/ceiling effects</td>
<td>Pilot interviews ((n = 64)) and cognitive interviews ((n = 28))</td>
<td>Five-item descriptive system</td>
</tr>
<tr>
<td>3. Valuation</td>
<td>To estimate the relative value of attributes</td>
<td>Rank sum method ((n = 424)) with a view to subsequent discrete choice experiment</td>
<td>SanQoL index values anchored at 0 and 1</td>
</tr>
<tr>
<td>4. Validity &amp; reliability</td>
<td>To assess the extent to which the instrument measures what it intends to measure (validity) and does so consistently (reliability)</td>
<td>Quantitative survey amongst intervention and control groups of a trial ((n = 424))</td>
<td>Evidence of construct validity, convergent validity, and test-retest reliability.</td>
</tr>
</tbody>
</table>

**Abbreviation:** SanQoL, sanitation-related quality of life.

**TABLE 2 Descriptive system for the SanQoL measure**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Questionnaire item</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disgust</td>
<td>Can you use the toilet without feeling disgusted?</td>
<td>• Always&lt;br&gt;• Sometimes&lt;br&gt;• Rarely&lt;br&gt;• Never</td>
</tr>
<tr>
<td>Health</td>
<td>Can you use the toilet without worrying that it spreads diseases?</td>
<td>• Always&lt;br&gt;• Sometimes&lt;br&gt;• Rarely&lt;br&gt;• Never</td>
</tr>
<tr>
<td>Privacy</td>
<td>Can you use the toilet in private, without being seen?</td>
<td>• Always&lt;br&gt;• Sometimes&lt;br&gt;• Rarely&lt;br&gt;• Never</td>
</tr>
<tr>
<td>Shame</td>
<td>Can you use the toilet without feeling ashamed for any reason?</td>
<td>• Always&lt;br&gt;• Sometimes&lt;br&gt;• Rarely&lt;br&gt;• Never</td>
</tr>
<tr>
<td>Safety</td>
<td>Are you able to feel safe while using the toilet?</td>
<td>• Always&lt;br&gt;• Sometimes&lt;br&gt;• Rarely&lt;br&gt;• Never</td>
</tr>
</tbody>
</table>

**Abbreviation:** SanQoL, sanitation-related quality of life.

*respondents can choose “prefer not to answer” for any item. The recall period was 4 weeks (Supplementary Material C).
where: \( w_i \) is the weight of the \( i \)th attribute; \( N \) is the number of attributes; \( R_i \) is the mean rank of the \( i \)th attribute in the population; \( x_{ij} \) are attribute scores ranging from 0 to 3 for the \( j \)th individual, where "always" = 3 and "never" = 0; \( S_j \) is the SanQoL index value for the \( j \)th individual.

We explored differences in attribute ranks by gender, whether the respondent was elderly (aged 60 or over), and treatment group. The rationale for including gender was that women experience sanitation differently to men in that they squat for urination, manage menstrual hygiene, and are more likely to fear and experience “peeping” or assault (Tilley et al., 2013). The rationale for including elderliness was that older people are particularly impacted by poor sanitation, predominantly as a result of disabilities that occur with ageing (Groce et al., 2011). In Mozambique, an elderly person is defined in law as anyone aged 60 or older (Castel-Branco & Andrés, 2019). The rationale for including treatment group was to assess whether people who had been using higher-quality toilets for four years (intervention group) had different perceptions of the relative value of the attributes. Rank is an ordered categorical variable, so we used mixed-effects ordered logit models, clustering standard errors at the compound level. We regressed on the rank for each attribute (which ranged from 1 to 5), including as covariates gender, aged 60 or over, and treatment. We also included as covariates any variables which both (i) differed at the 5% level along any of those three lines; and (ii) plausibly influenced ranking. The two variables meeting those criteria were the wealth index (differed by treatment status) and primary education completion (differed by gender and by aged 60 or over). P-values less than 0.05 were considered statistically significant evidence of association. As a robustness check, we also used generalised linear mixed models (GLMM) to analyse rank as a continuous variable, with the same set of covariates.

2.5 | Validity and reliability assessment

We assessed internal reliability using item-total correlation and Cronbach’s alpha, for which common acceptability thresholds are 0.4 (Ware et al., 1980) and 0.7 (Nunnally, 1978) respectively. There is debate about whether these metrics are appropriate for indices, since each item measures a different dimension (Konerding, 2013), but we include them because our measure comprises a single overall score to represent a single construct. We examined distributions of frequency endorsements in aggregate and by gender (Terwee et al., 2007). We assessed test-retest reliability by re-interviewing 69 respondents (16% of sample) two weeks after the original interview (Streiner et al., 2015). We used a two-way mixed effects model (Koo & Li, 2016) to evaluate the intraclass correlation coefficient (ICC) for the SanQoL index value, against an acceptability threshold of 0.7 (Terwee et al., 2007).

To assess construct validity, we pre-specified hypotheses (Table 3) about the presence or absence and direction of associations between each attribute score and a set of user and toilet characteristics. We tested these using GLMM, because attribute scores represent points on an underlying continuous scale, with standard errors clustered at the compound level. We regressed on each attribute score, including as covariates: gender; aged 60 or over; wealth index; primary education completion; whether the toilet had a concrete or tile floor; whether the toilet had masonry or zinc sheet walls; whether the toilet had an inside lock; and, whether the enumerator smelt faeces on entering toilet. The hypotheses drew on the previous qualitative work (Ross et al., 2021) and the broader literature on motives for sanitation behaviours and mental wellbeing (Novotný et al., 2018; Sclar et al., 2018).

We assessed convergent validity by correlation (Pearson’s \( r \)) between SanQoL index values and the WHO-5 mental wellbeing index (Topp et al., 2015). We expected correlation to be positive but less than 0.5, since sanitation is unlikely to be a primary driver of mental wellbeing (Sclar et al., 2018). Finally, we investigated the convergence of SanQoL index values between respondents using the same toilet. We used inter-rater methods to test the hypothesis that responses would be positively correlated but not equal, because any two people may experience the same toilet differently. We calculated the ICC using a one-way random effects model (Koo & Li, 2016). Interpretation of this ICC is “fair” (0.40–0.59), “good” (0.60–0.74), or “excellent” (>0.75) (Cicchetti, 1994).
3 | RESULTS

3.1 | Piloting and cognitive interviews

Of the 64 piloting respondents, 48% were male and 52% female. No single category per item received greater than 50% of endorsements in piloting, well-below an 80% benchmark (Streiner et al., 2015). As is desirable, there were no floor or ceiling effects, that is, absence of high proportion of respondents with maximum/minimum scores (Streiner et al., 2015). Cognitive interviewing showed that items could be understood, and no changes to the English descriptive system were required. However, there was one change to the Portuguese. The disgust item (Table 2), originally framed as an adjective (enojado = disgusted) was reframed as a noun (nojo = disgust), considered more natural in the Portuguese spoken in the setting.

3.2 | Valuation

We interviewed 424 individuals (220 female, 204 male) each from a different household, with a response rate of 99%. Respondent characteristics are tabulated by gender in Table 4 (and by trial arm in Supplementary Material E). Respondents lived on 275 MapSan trial compounds (131 control, 144 intervention). About two thirds had completed primary education, with slightly more men (70%) than women (57%) having done so. There was near-universal access to piped water connections (98%). The vast majority (89%) of respondents shared their toilet with other households, with a mean of 12.2 people sharing each toilet stance (cubicle).

Health had the highest mean rank and shame the lowest, meaning they also had the highest and lowest attribute weights (Table 5). Using Equation 2, we derived a value set (Supplementary Material E). Across the five attributes, there was insufficient evidence for differences in ranks by gender, being elderly, or intervention status (Table 6, full output in
Supplementary Material E), which is broadly desirable as heterogeneity in public values can make decision-making more contested (Baker et al., 2021). The GLMM robustness checks support this result (Supplementary Material E).

Control group respondents used pit latrines of various levels of quality, and their SanQoL index values approximate a normal distribution (Figure 2). The distribution for intervention group respondents is unimodal but left-skewed because toilets were high-quality. The mean SanQoL index value was 0.49 in the control group sample (95% CI: 0.46–0.53) and 0.83 in the intervention group (95% CI: 0.81–0.86).
3.3 | Validity and reliability assessment

There were no missing values for SanQoL attributes. For the privacy item, 61% of women and 57% of men endorsed “always,” which is slightly elevated but not enough to cause concern in this sample (Table 7). There were only eight participants who responded “never” to all five questions, and a further five who responded “never” for all but one of the five questions. There were 75 participants who responded “always” to all five questions.

Item-total correlations ranged between 0.69 and 0.73 (Supplementary Material F). Cronbach’s alpha was 0.77 and the test-retest ICC for the index was 0.87, both above specified thresholds. Inter-item polychoric correlation coefficients averaged 0.51, with range 0.40–0.70, supporting the idea that attributes are sufficiently distinct but nonetheless measure the same construct (Supplementary Material F).

Considering construct validity, there was evidence of association for 86% (30/35) of hypothesised presence or absence associations, or 79% (11/14) considering presence only (Table 8 and Supplementary Material G). All presence associations were in the hypothesised direction. Correlation between SanQoL index values and the WHO-5 mental wellbeing index was 0.24, indicating slight convergence as hypothesised. The ICC was 0.72 for convergence of SanQoL index values between two respondents using the same toilet, indicating substantial but not complete correlation, as hypothesised.

4 | DISCUSSION

In this study, we developed a SanQoL instrument in urban Mozambique, with a five-item descriptive system and accompanying valuation. By sampling individuals from the intervention and control groups of a sanitation intervention trial, we undertook an interim valuation of the SanQoL index, and assessed its validity and reliability. Considering construct validity, there was evidence supporting 79% of hypothesised presence associations, all in the hypothesised direction.
We observed convergence between SanQoL index values and the WHO-5 mental well-being index, and between two respondents using the same toilet. The ICC for test-retest reliability was well above commonly used thresholds.

The benefits of sanitation beyond infectious disease are likely to underpin household WTP for sanitation improvements, but are not captured in DALYs or QALYs. The scale of the sanitation challenge, with 2 billion people lacking a basic service and US$ 20 billion invested annually in LMICs, makes it all the more surprising there been no patient/person-reported outcome measure for use in economic evaluations (Hutton & Chase, 2016). Our measure addresses this by quantifying the QoL benefits of sanitation in a way which can be applied in economic evaluation. The aggregation into a single score weighted by the relative value of attributes elicited from the target population underlies the claim to be measuring the value of sanitation overall, not only the status of diverse attributes. Our interim ranking-based approach to valuation is straightforward for respondents to understand, but it is limited in that encourages trading off of attributes as a whole rather than trading off different levels of attributes. As a consequence, the rank sum method assumes equal intervals between response categories, which does not capture relative value to a sufficient degree (Brazier et al., 2016).

<table>
<thead>
<tr>
<th>Response</th>
<th>Overall (n = 424)</th>
<th>Female (n = 220)</th>
<th>Male (n = 204)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)</td>
</tr>
<tr>
<td>1. Disgust</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>186 (43.9)</td>
<td>98 (44.5)</td>
<td>88 (43.1)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>125 (29.5)</td>
<td>63 (28.6)</td>
<td>62 (30.4)</td>
</tr>
<tr>
<td>Rarely</td>
<td>28 (6.6)</td>
<td>15 (6.8)</td>
<td>13 (6.4)</td>
</tr>
<tr>
<td>Never</td>
<td>85 (20.0)</td>
<td>44 (20.0)</td>
<td>41 (20.1)</td>
</tr>
<tr>
<td>2. Health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>174 (41)</td>
<td>96 (43.6)</td>
<td>78 (38.2)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>127 (30.0)</td>
<td>64 (29.1)</td>
<td>63 (30.9)</td>
</tr>
<tr>
<td>Rarely</td>
<td>30 (7.1)</td>
<td>14 (6.4)</td>
<td>16 (7.8)</td>
</tr>
<tr>
<td>Never</td>
<td>93 (21.9)</td>
<td>46 (20.9)</td>
<td>47 (23.0)</td>
</tr>
<tr>
<td>3. Privacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>249 (58.7)</td>
<td>134 (60.9)</td>
<td>115 (56.4)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>94 (22.2)</td>
<td>45 (20.5)</td>
<td>49 (24.0)</td>
</tr>
<tr>
<td>Rarely</td>
<td>15 (3.5)</td>
<td>7 (3.2)</td>
<td>8 (3.9)</td>
</tr>
<tr>
<td>Never</td>
<td>66 (15.6)</td>
<td>34 (15.5)</td>
<td>32 (15.7)</td>
</tr>
<tr>
<td>4. Shame</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>198 (46.7)</td>
<td>105 (47.7)</td>
<td>93 (45.6)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>113 (26.7)</td>
<td>60 (27.3)</td>
<td>53 (26.0)</td>
</tr>
<tr>
<td>Rarely</td>
<td>28 (6.6)</td>
<td>17 (7.7)</td>
<td>11 (5.4)</td>
</tr>
<tr>
<td>Never</td>
<td>85 (20.0)</td>
<td>38 (17.3)</td>
<td>47 (23.0)</td>
</tr>
<tr>
<td>5. Safety</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>193 (45.5)</td>
<td>90 (40.9)</td>
<td>103 (50.5)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>117 (27.6)</td>
<td>63 (28.6)</td>
<td>54 (26.5)</td>
</tr>
<tr>
<td>Rarely</td>
<td>34 (8.0)</td>
<td>19 (8.6)</td>
<td>15 (7.4)</td>
</tr>
<tr>
<td>Never</td>
<td>80 (18.9)</td>
<td>48 (21.8)</td>
<td>32 (15.7)</td>
</tr>
</tbody>
</table>

Note: No respondents selected "prefer not to answer" for any attribute.
Abbreviation: SanQoL, sanitation-related quality of life.
In combination with a measure of time using a sanitation service, the index values arising from our instrument can quantify incremental benefits in a sanitation-focused cost-effectiveness analysis (CEA). For example, analogous to the QALY, a year of sanitation service experienced at full sanitation-related capability (i.e., SanQoL = 1) might be used in CEA. The “length” dimension would be person years of service rather than life years. However, since CBA is most commonly used to evaluate sanitation interventions (Hutton & Chase, 2016), it would be important to estimate the monetary value of a such a weighted person year, based on WTP as with QALYs (Ryen & Svensson, 2015). After monetary valuation, the broader QoL benefits could be summed with health gains and other benefits such as time savings. It is through CBA that SanQoL is likely to have relevance to decision-makers with broader resource allocation remits (e.g., Ministry of Finance, Health, or Education) through comparing sanitation to non-sanitation interventions. See for example the Copenhagen Consensus (2020) studies using CBA to prioritise amongst 80+ interventions across multiple sectors in countries including Ghana, Haiti, Bangladesh, and India (Ghana Priorities, 2020). In theory, techniques for mapping to preference-based HRQoL measures could be used to estimate SanQoL-related QALYs gained from sanitation interventions, in addition to QALY gains from avoided infectious disease (Chuang & Whitehead, 2012). However, this may be conceptually problematic as the scope of SanQoL is so different to HRQoL measures, and would not be a research priority until QALYs are regularly used to evaluate sanitation interventions.

Measuring objective toilet characteristics has often been the focus of efforts to assess sanitation quality (Schelbert et al., 2020; Tidwell et al., 2018). Instead, SanQoL focuses on individuals’ self-assessed capabilities. This is important because sanitation interventions might improve toilets but with heterogeneous impacts on QoL, particularly if appropriate attention is not given to gender and social norms (O’Reilly, 2016). The only previous attempt to measure aspects of these outcomes was a 60-item profile measure of women’s sanitation insecurity (WSI) developed in rural India (Caruso et al., 2017). It is not appropriate for use in economic evaluation due to its lack of an overall score, use of equal weighting, and focus on women only. Its length also makes it impractical for routine use. Finally, while the WSI profile is conceptualised and applied as a measure of exposure (Caruso et al., 2018; Delea et al., 2019), SanQoL focuses on capabilities as QoL outcomes.

A short-form approach, with one item per attribute, was necessary for permitting future valuation through preference elicitation. In support of content validity, we note that our SanQoL attributes are broadly consistent with studies of insecurity, stress and motives related to sanitation in both rural and urban areas (Novotný et al., 2018; Sclar et al., 2018). However, there are exceptions. For example, convenience is commonly identified as important in settings where open defecation (OD) is prevalent (Novotný et al., 2018). However, OD is uncommon in our setting and most urban settings in general (WHO & UNICEF, 2021). In other settings where use of public toilets is common, convenience may also be an important benefit of switching to a household toilet. However, most economic evaluations of sanitation interventions already capture a major aspect of convenience by including the value of time savings (Hutton & Chase, 2016). Since the SanQoL attributes align with those identified in studies across diverse settings, the measure is likely to be appropriate for

### Table 8

<table>
<thead>
<tr>
<th>Variable</th>
<th>Disgust</th>
<th>Health</th>
<th>Shame</th>
<th>Safety</th>
<th>Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.698</td>
<td>0.461</td>
<td>0.250</td>
<td>&lt;0.001***</td>
<td>0.781</td>
</tr>
<tr>
<td>Aged 60+</td>
<td>0.594</td>
<td>0.757</td>
<td>0.336</td>
<td>0.283</td>
<td>0.795</td>
</tr>
<tr>
<td>Wealth index</td>
<td>0.028**</td>
<td>0.441</td>
<td>0.855</td>
<td>0.198</td>
<td>0.816</td>
</tr>
<tr>
<td>Toilet has concrete or tile floor</td>
<td>0.271</td>
<td>0.003***</td>
<td>0.006***</td>
<td>&lt;0.001***</td>
<td>0.003***</td>
</tr>
<tr>
<td>Toilet has masonry or complete zinc sheet walls</td>
<td>0.155</td>
<td>0.184</td>
<td>0.841</td>
<td>0.014**</td>
<td>0.030**</td>
</tr>
<tr>
<td>Toilet has inside lock</td>
<td>0.118</td>
<td>0.981</td>
<td>0.080*</td>
<td>0.005***</td>
<td>0.003***</td>
</tr>
<tr>
<td>Enumerator smells faeces on entering toilet</td>
<td>0.005***</td>
<td>&lt;0.001***</td>
<td>&lt;0.001***</td>
<td>0.171</td>
<td>0.084*</td>
</tr>
</tbody>
</table>

Note: *, **, *** indicate significance at the 10%, 5% and 1% level. All presence associations at the 5% level or below were in the hypothesised direction. Full regression output is in Supplementary Material G.

Abbreviations: GLMM, generalised linear mixed models; SanQoL, sanitation-related quality of life.
widespread use. However, in any setting, further assessment of validity through cognitive interviews is recommended, as recently undertaken for SanQoL in urban Ghana (Tidwell et al., 2020). In settings with substantially different sanitation practices, for example, rural areas where open defecation is common, further qualitative work may be required. Respondents not using toilets may need questions reformulated in terms of “Can you practise open defecation...” rather than “Can you use the toilet...”

Across the five SanQoL items, only 6% of responses (range: 4%–8%) endorsed the “rarely” option (Table 7). This compares to 11% in piloting. Alternative response options could be investigated in future research in a broader population, to establish the effect of response labels on endorsement distributions. However, our distributions align with the fact that for the WSI profile, also on a four-level frequency scale, only 5% of respondents endorsed its “often” category in a study with three times the sample size (Caruso et al., 2017). It may be that respondents prefer to scale these concepts on three levels. However, we think it is more likely that “sometimes” (used in both SanQoL and the WSI profile) comprises a very broad category, such that it might dominate the other middle category in any four-level scale. This could be investigated using think-aloud methods or item response theory in a larger sample of a population using a broad range of toilet types.

Use of the intervention and control groups of a trial as the study population was pragmatic, and contributed to the objectives of our broader body of research. However, it did provide only a narrow set of sanitation service types and living arrangements—for example, almost all respondents shared their toilet with other households in a compound setting. In addition, our study’s sample size was relatively small. Further research on the validity and measurement properties of SanQoL should take place in other settings, and a larger random sample of a target population that uses a broader range of sanitation service levels. In addition, our sample provides evidence of validity only for those aged 18 years or older, and further research would be required for use in younger populations who may experience sanitation in different ways (Nallari, 2015; Sahoo et al., 2015).

5 | CONCLUSION

The SanQoL instrument provides a measure of sanitation-related quality of life, and we have provided evidence for its validity and reliability in this setting. It enables measurement and valuation of the attributes of sanitation which users consider important, weighted by the relative value they place on those attributes. Such a measure can support future cost-effectiveness and cost-benefit analyses, as well as being appropriate for impact evaluation or routine program monitoring to assess the performance of interventions. The results of validity and reliability investigations reported here provide evidence that this measure can be used for assessing SanQoL in urban settings of LMICs, but further testing in other settings is required.

ACKNOWLEDGMENTS

We greatly appreciated the cooperation of survey respondents. We also appreciated the efforts of the fieldworkers: Euclimia Titosse, Carla Panguene, João-Pedro Guambe and Faustino Benzane. The authors gratefully acknowledge the technical and logistical support received from Vasco Parente and his team at WSUP Mozambique. We also benefitted from discussions with Sarah Smith, Julia Lohmann, Catherine Pitt and seminar participants at the Global Health Economics Centre at LSHTM. The inputs of the expert reviewers were helpful in improving items. This work was supported by the Economic and Social Research Council through a PhD studentship. The fieldwork was funded by the Bill & Melinda Gates Foundation. The funders had no role in the identification, design, conduct, or reporting of the analysis.

CONFLICT OF INTERESTS

The time or research costs of the following researchers on this study was funded in part by the study’s funders, the Bill and Melinda Gates Foundation: Ian Ross, Oliver Cumming, Zaida Adriano, Rassul Nala, Joe Brown, Robert Dreibelbis.

DATA AVAILABILITY STATEMENT

Deidentified individual participant data, codebook, and replication code are available on LSHTM data repository: https://doi.org/10.17037/DATA.00002442.

ORCID

Ian Ross © https://orcid.org/0000-0002-2218-5400
REFERENCES


