The impact of diarrhoea measurement methods for under 5s in low- and middle-income countries on estimated diarrhoea rates at the population level
Rego, Ryan; Watson, Samuel; Gill, Paramjit; Lilford, Richard

DOI:
10.1111/tmi.13739

License:
Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.
INTRODUCTION

Effective surveillance of diarrhoea in children under 5 at the population level is required to track outbreaks, allocate public health resources and evaluate water, sanitation and hygiene (WASH) interventions [1]. However, the choice of method used to ascertain whether an episode of diarrhoea has occurred has been hypothesised to impact estimated diarrhoea rates [2, 3]. If true, the results of observational and intervention studies may differ by method of measurement.
thereby obscuring the effects of the explanatory variable(s) of interest. Since diarrhoea is one of the biggest causes of death in children, this is a methodological point of considerable practical importance for surveillance, evaluation of interventions and establishing the true burden of diarrhoea-associated morbidity and mortality. The two most common methods of diarrhoea surveillance at the population level, passive and active surveillance, take different approaches. Passive surveillance relies on data collected from health facilities and therefore excludes all children with diarrhoea who do not attend a facility. Passive surveillance estimates are therefore skewed towards more severe disease and away from marginalised groups such as slum dwellers, refugees and migrants, who are less likely to visit health facilities and who are more likely to visit informal facilities than non-marginalised groups [4–9]. Passive surveillance is a useful, inexpensive tool to detect new outbreaks of severe diseases such as cholera, but is arguably less useful as an epidemiological tool for the measurement of population-level diarrhoea rates or in trials of WASH interventions.

Active surveillance, based on door-to-door surveys, provides a more complete report of diarrhoea rates than passive surveillance, but may also be subject to measurement error and bias [4]. Carers may forget events that happened in the past, particularly during longer lengths of recall [10, 11]. They may also have a poor understanding of what diarrhoea is [12]. UNICEF and the Demographics and Health Surveys (DHS) programme have recommended a method based on asking carers if their child has had three or more loose or watery stools in any 24-h period within the previous 14 days [13]. However, this method is by no means universally applied. In addition to concerns over measurement error, concern has been expressed that diarrhoea rates may be subjective to bias due to 'reactivity', where despite any true clinical difference, people report different diarrhoea rates for psychological reasons. For example, they may be the beneficiaries of an intervention and not want to appear ungrateful; not want to answer in a way that is not socially desirable or be guided subliminally or non-verbally by the surveyors (e.g. if the question is asked in a way which steers the respondents towards a certain answer) [14, 15]. As such, reactivity creates a particular concern for evaluations of WASH interventions.

In order to examine the association between the method used in diarrhoea measurement and diarrhoea rates, we conducted a systematic review of all studies published between 2000 and June 2021 that report under five diarrhoea rates in LMICs (as defined by OECD) [18]. The search strategy aimed to capture any study that estimated diarrhoea rates among under 5s, including both studies that performed direct 'head-to-head' comparisons of methods, and studies estimating diarrhoea rates using only one method that we can compare indirectly. As we are only interested in population-level diarrhoea rates, we excluded studies that were not designed to capture population-level diarrhoea rates, such as studies which measured hospital acquired infection, clinical trials in which diarrhoea was an adverse drug event and case–control studies in which diarrhoea was the case. Importantly, these exclusions do not include WASH trials which took place in the community setting. Studies were restricted to English or French (Table 1).

We searched the MEDLINE, Embase and PubMed databases for studies matching the inclusion and exclusion criteria. The search string (Appendix 1) included keywords relating to diarrhoea and population-level disease measurement. The string then restricted the results to human studies and studies in LMICs.

SW and RR independently screened each title and abstract, and any disagreements were resolved through discussion with RL. Full texts were screened in a similar manner. Where full texts were not available, we requested the article from the University of Warwick’s Article Reach Service. We

| Table 1 Inclusion and exclusion criteria |
|------------------------------------------|------------------------------------------|
| Inclusion criteria                       | Exclusion criteria                       |
| Reports diarrhoea rates among under 5s (i.e. incidence or prevalence) | Measures hospital-acquired infection |
| Takes place in LMICs                     | Clinical trials in which diarrhoea was an adverse drug event |
| Published between 2000 and June 2021     | Case–control studies in which diarrhoea is the case |
| English or French language               |                                         |
excluded unavailable studies and duplicate studies. In the event that multiple studies used the same data source, we selected a random study for inclusion.

**Data extraction**

RR extracted the data, with SW duplicating data extraction for random 15% of the included studies. The random 15% extracted by SW matched completely with the extraction by RR. As a single reported study may have included several independent reports of diarrhoea rates, we treated each report as separate ‘observations’ within one study. This would not only apply to all the direct studies but also arose in indirect studies when conducted in more than one site, included multiple rounds of data collection and/or was a trial with multiple arms. For example, a two-armed trial which estimated diarrhoea rates at baseline and end-line yielded four observations.

Data extracted (Table 2) included participant demographics; study design, if the study was an observational study using primary data (including non-randomised trials), an observational study using secondary data or a randomised control trial; diarrhoea rates; measurement methods and if the study was a direct comparison of methods. We defined direct comparison studies as those which included at least two separate arms, each with a different method of diarrhoea measurement (including altering recall period or questioning frequency) that compare estimated diarrhoea rates between each arm.

**Data analysis**

We indirectly compared the diarrhoea rates from included observations, including those from studies that performed direct comparisons of measurement methods, and studies using only a single method of measurement. We summarised the key variables in each individual observation, including estimated diarrhoea rates, measurement type, year and region (as defined by UNDP). We estimated a hierarchical meta-regression model for log diarrhoea rate (incidence in episodes per child year), adjusting for region and time (in years) and interactions between time and region. We also adjusted for study design, including a categorical variable with levels: (1) observational studies which use primary data sources; (2) observational studies which use secondary data sources; (3) RCT intervention arms before the intervention, or the control arm (if reported) and (4) RCT experimental arm after the intervention. The model was estimated in StataSE Version 15 using generalised least squares with random effects at the study level, to account for within-study correlation, and weighting by the study sample size [19, 20].

We estimated two separate meta-regression models. The first iteration included the observations from both passive and active surveillance studies to estimate the effect of surveillance types (passive/active) on estimated diarrhoea rates, and as such included a dummy variable for surveillance type. We also estimated pooled temporal and regional trends from this model.

**TABLE 2** Data extracted for each observation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units (if applicable)</th>
<th>Categorisation (if categorised)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Study title</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Year of publication</td>
<td>Years</td>
<td>NA</td>
</tr>
<tr>
<td>Year of data collection</td>
<td>Years</td>
<td>NA</td>
</tr>
<tr>
<td>Mean participant age</td>
<td>Years (with standard error)</td>
<td>NA</td>
</tr>
<tr>
<td>Participant sex breakdown</td>
<td>% Female</td>
<td>NA</td>
</tr>
<tr>
<td>Geography</td>
<td>Urban, rural or mixed</td>
<td>Regions as defined by the United Nations Development Programme (UNDP) [19]</td>
</tr>
<tr>
<td>Country</td>
<td>NA</td>
<td>WASH, Health or Nutrition</td>
</tr>
<tr>
<td>Intervention (for randomised control trials [RCTs])</td>
<td>NA</td>
<td>Incidence risk ratio</td>
</tr>
<tr>
<td>Effectiveness of intervention (for RCTs)</td>
<td>Incidence risk ratio</td>
<td>NA</td>
</tr>
<tr>
<td>Diarrhoea rate (in incidence)</td>
<td>Episodes per child year (converted if necessary)</td>
<td>NA</td>
</tr>
<tr>
<td>Measurement type</td>
<td>NA</td>
<td>Passive, retrospective active, prospective (diary) active</td>
</tr>
<tr>
<td>Recall period</td>
<td>Days</td>
<td>NA</td>
</tr>
<tr>
<td>Questioning frequency</td>
<td>Days</td>
<td>Daily, weekly, monthly or annually or longer</td>
</tr>
<tr>
<td>Questioning type</td>
<td>NA</td>
<td>Pictorial, verbal or other</td>
</tr>
<tr>
<td>Sample size (observed cases for passive surveillance studies)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Study design</td>
<td>NA</td>
<td>Observational (primary), observational (secondary), RCT (broken down by arm)</td>
</tr>
<tr>
<td>Direct comparison of methods</td>
<td>NA</td>
<td>Yes/no</td>
</tr>
</tbody>
</table>
The second iteration included only observations from active surveillance studies to examine the effects of variables exclusive to active surveillance studies on estimated diarrhoea rates. These included variables for recall period (as a continuous numeric term in days), and questioning frequency and recall type (both as categorical variables). Furthermore, any other variations found between active surveillance studies, including reactivity and questioning type (e.g. verbal or pictorial), were included (Table 3).

RESULTS

Study identification

We identified 2040 studies in total, which was reduced to 1973 after duplicates were removed. Abstract and title screening yielded 577 studies, with a further 289 excluded after full-text review. Common reasons for exclusion included not presenting data on under 5s (n = 64), not including data on diarrhoea rates (n = 58), and not being able to obtain the full text (n = 4). Thus, 288 full-text studies (Appendix 2) were included in the final review (Figure 1).

Study characteristics

We first describe the characteristics of the described studies, such as observations per study, data collection method, geography and others. As stated, many studies included more than one observation of diarrhoea rate, arising from observations at different time points, being a trial with two or more arms, or data collection in more than one location. Appendix 3 presents information on the number of observations per study. We identified 671 separate observations of population-level diarrhoea rates, and these constitute our denominator. In total, there were 646 (96%) active surveillance observations and 25 (4%) passive surveillance observations. Of the 646 active surveillance observations, 633 (94%) were retrospective, while 13 (2%) were prospective (a diary kept by the carer). Of the observations, 354 (56%) used a 14-day recall period, as recommended by UNICEF and the Demographic and Health Survey (DHS) programme. Of the 188 observations which came from randomised control trials (RCTs), 53 (28%) used a 14-day recall period and 95 (51%) used a 7-day recall period. Furthermore, of the observations from RCTs, 21 (12%) questioned daily, 56 (30%) questioned weekly and 49 (26%) questioned biweekly.

By region (as defined by the UNDP), sub-Saharan Africa was the setting for the largest number of observations (265; 40%), followed by The Americas (140; 21%), Central and Southern Asia (180; 27%), East and South-East Asia (62; 9%), North Africa and the Middle East (23; 3%) and finally Oceania (1; 0%). Rural (212; 32%) and mixed geography (226; 34%) areas provided more observations than urban areas (116; 17%).

None of the included active surveillance observations used non-verbal methods of diarrhoea measurement (e.g. showing carers pictures of stool), and no studies made mention of a ‘gold standard’ of diarrhoea measurement.

Four of the included studies performed direct head-to-head comparisons of diarrhoea measurement methods: three examining the effect of differing recall periods on diarrhoea rates, and one examining the effect of questioning frequency on estimated diarrhoea rates. No studies were identified that analysed reactivity in diarrhoea measurement.

Differences between active and passive surveillance

As stated in the Introduction, the two main categories of disease surveillance are active surveillance (community surveying) and passive surveillance (measures of visits to health facilities). No studies performed direct ‘head-to-head’ comparisons of active and passive surveillance. After model-based adjustment to perform an indirect comparison of passive and active surveillance, passive surveillance was associated with a 97% lower estimated diarrhoea rate than active surveillance (incidence risk ratio [IRR] = 0.03, 95% CI [0.02, 0.06]) (Table 4).

Impacts of factors within active surveillance studies on estimated diarrhoea rates

Specific to active surveillance studies, questioning frequency (how often participants are surveyed), recall period of diarrhoea surveying and retrospective versus prospective questioning may differ.

Questioning frequency

One study directly examined the effect of differing questioning frequencies on estimated diarrhoea rates; Zwane et al. estimated that biweekly surveys had a 7–15% lower diarrhoea rate than six-monthly surveys when using the same recall period [21]. Our indirect comparison of active surveillance observations produced comparable results; after model-based adjustment, we found that less frequent questioning was associated with an increase in estimated diarrhoea rates. For example, one-time questioning was associated with a rate over four times higher than daily questioning (IRR = 4.22 [2.73, 6.52]) (Table 4). This, however, is not evident graphically in unadjusted crude data – however, a large amount of variance as questioning frequency increases can still be seen (Figure 2a).

Recall period

Three studies directly examined the effect of differing recall periods on estimated diarrhoea rates, though the recall
periods examined were different. Melo et al. found that diarrhoea rates were cut by a third when carers recall over 4 weeks versus 24 h [22]. Feikin et al. similarly estimated that diarrhoea rates were cut by a fifth when carers recall over 11–13 days versus 1–2 days [10]. Lee et al. estimated that diarrhoea rates were similar for carers who recalled over a 72-h period versus a 24-h period [23], but this is a much shorter range than that investigated in the other two studies.

Based on an indirect comparison of the included active surveillance observations, we estimated that recall periods and estimated diarrhoea rates were inversely associated. After model-based adjustment, we found that a doubling of recall period was associated with a 48% reduction in diarrhoea rate (IRR = 0.52 [0.46, 0.60]; Table 4). This is also evident graphically in the crude data (Figure 2b).

**Prospective versus retrospective**

No studies directly compared prospective (diary) recall designs against retrospective. We estimated through indirect

---

**TABLE 3** Summary table of the 671 observations included in the systematic review, with average estimated diarrhoea rates for key variables

<table>
<thead>
<tr>
<th>Study design</th>
<th>Passive n (% of total)</th>
<th>Prospective (diary) active n (% of total)</th>
<th>Retrospective active n (% of total)</th>
<th>Total</th>
<th>Average estimated diarrhoea rate (episodes per child year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection year</td>
<td>Total</td>
<td>25 (32.7%)</td>
<td>13 (1.9%)</td>
<td>633</td>
<td>6.75</td>
</tr>
<tr>
<td></td>
<td>2016–2018</td>
<td>1 (2.9%)</td>
<td>0 (0%)</td>
<td>33 (97.1%)</td>
<td>34 (7.08)</td>
</tr>
<tr>
<td></td>
<td>2011–2015</td>
<td>12 (6.8%)</td>
<td>2 (1.1%)</td>
<td>163 (92.1%)</td>
<td>177 (6.63)</td>
</tr>
<tr>
<td></td>
<td>2006–2010</td>
<td>6 (3.2%)</td>
<td>2 (1.1%)</td>
<td>180 (95.7%)</td>
<td>188 (6.60)</td>
</tr>
<tr>
<td></td>
<td>2001–2005</td>
<td>4 (3.4%)</td>
<td>2 (1.7%)</td>
<td>112 (94.9%)</td>
<td>118 (7.80)</td>
</tr>
<tr>
<td></td>
<td>1996–2000</td>
<td>2 (2.5%)</td>
<td>1 (1.3%)</td>
<td>77 (96.2%)</td>
<td>80 (6.15)</td>
</tr>
<tr>
<td></td>
<td>1991–1995</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>38 (100%)</td>
<td>38 (5.99)</td>
</tr>
<tr>
<td></td>
<td>1986–1990</td>
<td>0 (0%)</td>
<td>3 (42.9%)</td>
<td>4 (57.2%)</td>
<td>7 (3.54)</td>
</tr>
<tr>
<td>Region</td>
<td>Sub-Saharan Africa</td>
<td>5 (1.9%)</td>
<td>3 (1.1%)</td>
<td>257 (97.0%)</td>
<td>265 (6.32)</td>
</tr>
<tr>
<td></td>
<td>North Africa and the Middle East</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>23 (93.4%)</td>
<td>23 (6.56)</td>
</tr>
<tr>
<td></td>
<td>Central and Southern Asia</td>
<td>3 (1.7%)</td>
<td>2 (1.1%)</td>
<td>175 (97.2%)</td>
<td>180 (9.41)</td>
</tr>
<tr>
<td></td>
<td>Eastern and South-East Asia</td>
<td>13 (21.0%)</td>
<td>4 (6.4%)</td>
<td>45 (72.6%)</td>
<td>62 (3.20)</td>
</tr>
<tr>
<td></td>
<td>Oceania</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (100.0%)</td>
<td>1 (4.02)</td>
</tr>
<tr>
<td></td>
<td>The Americas</td>
<td>5 (3.6%)</td>
<td>3 (2.1%)</td>
<td>132 (94.3%)</td>
<td>140 (5.74)</td>
</tr>
<tr>
<td></td>
<td>Observational (primary)</td>
<td>19 (6.5%)</td>
<td>11 (3.7%)</td>
<td>264 (88.9%)</td>
<td>294 (8.21)</td>
</tr>
<tr>
<td></td>
<td>Observational (secondary)</td>
<td>6 (3.2%)</td>
<td>0 (0%)</td>
<td>183 (96.8%)</td>
<td>189 (3.87)</td>
</tr>
<tr>
<td></td>
<td>RCT (baseline or control)</td>
<td>0 (0%)</td>
<td>1 (1.1%)</td>
<td>93 (98.9%)</td>
<td>94 (8.37)</td>
</tr>
<tr>
<td></td>
<td>RCT (experimental post-intervention)</td>
<td>0 (0%)</td>
<td>1 (1.1%)</td>
<td>93 (98.9%)</td>
<td>94 (6.32)</td>
</tr>
<tr>
<td>Geography</td>
<td>Mixed</td>
<td>14 (5.8%)</td>
<td>2 (0.8%)</td>
<td>226 (93.4%)</td>
<td>242 (4.02)</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>4 (1.8%)</td>
<td>4 (1.8%)</td>
<td>212 (96.3%)</td>
<td>220 (10.55)</td>
</tr>
<tr>
<td></td>
<td>Urban</td>
<td>4 (3.1%)</td>
<td>7 (5.5%)</td>
<td>116 (91.3%)</td>
<td>127 (5.15)</td>
</tr>
<tr>
<td></td>
<td>Missing data</td>
<td>3 (3.7%)</td>
<td>0 (0%)</td>
<td>79 (96.4%)</td>
<td>82 (6.99)</td>
</tr>
<tr>
<td>Questioning frequency</td>
<td>Daily</td>
<td>NA</td>
<td>0 (0%)</td>
<td>55 (100%)</td>
<td>55 (4.68)</td>
</tr>
<tr>
<td></td>
<td>Weekly</td>
<td>NA</td>
<td>9 (9.5%)</td>
<td>86 (90.5%)</td>
<td>95 (4.49)</td>
</tr>
<tr>
<td></td>
<td>Monthly</td>
<td>NA</td>
<td>1 (1.3%)</td>
<td>77 (98.8%)</td>
<td>78 (13.47)</td>
</tr>
<tr>
<td></td>
<td>Annual+</td>
<td>NA</td>
<td>0 (0%)</td>
<td>30 (100%)</td>
<td>30 (5.43)</td>
</tr>
<tr>
<td></td>
<td>One off</td>
<td>NA</td>
<td>3 (0.8%)</td>
<td>385 (99.2%)</td>
<td>388 (6.76)</td>
</tr>
<tr>
<td>Recall period</td>
<td>1–3 days</td>
<td>NA</td>
<td>1 (1.6%)</td>
<td>60 (98.4%)</td>
<td>61 (25.88)</td>
</tr>
<tr>
<td></td>
<td>4–14 days</td>
<td>NA</td>
<td>11 (2.1%)</td>
<td>520 (97.9%)</td>
<td>531 (5.22)</td>
</tr>
<tr>
<td></td>
<td>15–30 days</td>
<td>NA</td>
<td>1 (2.6%)</td>
<td>37 (97.4%)</td>
<td>38 (3.94)</td>
</tr>
<tr>
<td></td>
<td>31–90 days</td>
<td>NA</td>
<td>0 (0%)</td>
<td>13 (100%)</td>
<td>13 (1.00)</td>
</tr>
<tr>
<td></td>
<td>91+ days</td>
<td>NA</td>
<td>0 (0%)</td>
<td>3 (12.5%)</td>
<td>3 (1.83)</td>
</tr>
</tbody>
</table>
comparison that retrospective recall observations were associated with a lower rate than prospective (diary) observations, but the effect size was relatively uncertain (IRR = 0.85 [0.53, 1.38]; Table 4).

DISCUSSION

Main findings

We provide evidence that estimated under 5 diarrhoea rates are sensitive to the methods used in their measurement. This includes variance introduced by the choice of passive or active surveillance, as well as factors specific to active surveillance.

Passive surveillance methods were associated with 97% lower diarrhoea rates than active surveillance methods. The most probable explanation is that carers do not seek health care for the majority of cases where, if asked, they would report diarrhoea. While not shown in our results, several studies on access to health care among infants in LMICs show that the propensity of carers to seek care for their under 5s with diarrhoea is influenced by diarrhoea severity, socioeconomic or legal status and other demographic characteristics [5–9].

Regarding active surveillance methods, we found that different questioning frequencies influence estimated diarrhoea rates. There was a trend to lower estimated diarrhoea rates given higher questioning frequencies: one off questioning was associated with an over four times higher estimated diarrhoea rate than daily questioning. We also found that differing recall periods were associated with a change in estimated diarrhoea rates: a doubling of recall period was associated with a halving of estimated diarrhoea rate.

Factors that result in subjective measurements during active surveillance

As the distinction between and recall of a diarrhoeal or non-diarrhoeal stool by carers is largely subjective, several cognitive factors can affect measurement. These include respondent fatigue (becoming tired of answering questions), recall bias (forgetting events that have occurred in the past), perception bias (not understanding the question being asked) and reactivity (answering differently due to experiencing an intervention).

Respondent fatigue

Declining diarrhoea rates with increasing questioning frequency (but the same recall period) suggest respondent
fatigue – participants may be inclined to pay attention to their bowel movements at first, but lose motivation with further rounds of questioning.

Recall bias

Recall bias is the effect of forgetting: participants are more likely to recall recent than older events. We would expect a lower reported number of diarrhoea episodes with longer recall periods and this was borne out by our analysis including two of the three head-to-head comparisons – the exception examined a much smaller gap between questions than the other two [10, 22]. This finding was corroborated by a more recent head-to-head comparison where daily recall was associated with a 30 percentage point higher estimated diarrhoea rate than fortnightly recall during a text message survey of under 5 diarrhoea in urban Tanzania [24]. While not examined in our review, it has also been reported that the effect of recall bias is more apparent for moderate diarrhoea compared to severe diarrhoea. Zafar et al., for example, found that moderate diarrhoea is reported at half the rate of severe diarrhoea during longer recall periods [11].

Other factors affecting diarrhoea measurement

Other factors outside the scope of this review can further influence estimated diarrhoea rates. For example, poor caregiver perception of diarrhoea (understanding what is or is not diarrhoea) can result in error in diarrhoea measurement. Voskuil et al. determined that carers of children under 5 were only able to identify 56–75% of loose or watery stool and 80% healthy stools [12].
Another relevant phenomenon is ‘reactivity’, whereby participants adjust their answers to a survey according to how they believe they ought to respond, regardless of any true underlying difference. We did not identify any studies of reactivity in this review, but it has been discussed as a potential explanatory factor in previous trials. Luby et al. mentioned courtesy bias as a potential source of reactivity, stating ‘people who received the intervention might have been grateful and, out of courtesy, reported less diarrhoea’ [14]. Wood et al. further found evidence for reactivity in clinical trials for various diseases, reporting that inadequate concealment of interventions is associated with improved treatment performance in trials, particularly for subjective outcomes [25]. It was not possible to examine reactivity through courtesy bias or inadequate concealment in our review as WASH trials by nature are unblinded.

CONCLUSION

The magnitude of the variation in diarrhoea rates, even among active methods of surveillance, suggests the need for standardisation of diarrhoea measurement methods to facilitate comparisons between studies. Despite the 14-day UNICEF and DHS standard, there does seem to be a trend towards using a 7-day recall period. It was the most frequently used recall period (51%) among RCTs in our review. Three of the three recent large integrated WASH trials (the SHINE trials in Bangladesh and Kenya, and the WASH Benefits trial in Zimbabwe) also used a 7-day retrospective recall to measure diarrhoea [26–29], in contravention of the UNICEF and DHS guidelines. However, the above three trials differed among themselves with respect to question frequency: the SHINE trials questioned carers annually, while the WASH Benefits trial questioned mothers every ‘2 to 6’ months [27–29].

It could be argued that lack of standardisation simply introduces measurement error in trials that can be counteracted by increasing sample size. However, this is only likely to be true if it is assumed that the different methods affect only the propensity of someone with a true case of diarrhoea (or indeed enteric infection) to report a case of diarrhoea (the ‘sensitivity’ of the method) [30]. If, however, there is a loss of ‘specificity’ – the propensity of someone who did not have diarrhoea (or enteric infection) to report a case of diarrhoea – then intervention effects will also be biased across studies using different methods. It is also likely that any measurement errors will bias results towards the null, rather than towards reports of intervention effectiveness [30]. It is therefore possible that the choice of methodology is at least partially responsible for the widely varying, and often disappointing, results of evaluations on WASH interventions [14, 27, 29, 31, 32].

While a widely accepted standard would facilitate comparisons across different observational and experimental studies, this raises the question of what the optimal standard may be that would also produce reliable reports of diarrhoea rates. There is no ‘gold standard’ method for measurement of diarrhoea rates. In part, this is because of the difficulty in defining the underlying construct and providing a culturally and linguistically consistent definition of a case or episode of ‘diarrhoea’. Direct observation by an expert might constitute a gold standard against which other methods could be compared, as has been described above in the study by Voskuilj and colleagues [12]. However, judgements among experts may not be universal. Moreover, the collection of every stool and use of experts to classify them quickly becomes impractical at larger scales.

We propose two policies to mitigate the problem. First, an agreed consensus method for the measurement of diarrhoea rates in surveys. Second, triangulation of diarrhoea rates with other observations that reflect on gastrointestinal health when interventions are evaluated. Many WASH studies already include anthropometric measurements as outcomes alongside diarrhoea. Furthermore, direct measurement of environmental contamination and pathogen levels in stool samples should complement diarrhoea rates in clinical studies. This would also allow for determination of how much diarrhoea is attributable to infection (and which can be reduced by WASH interventions), rather than

---

**FIGURE 2**  
(a) Box plot of estimated diarrhoea rate against questioning frequency for active surveillance studies. (b) Scatter plot and trendline of estimated diarrhoea rates against recall period for active surveillance studies
non-infective reasons (which would likely not be impacted by WASH interventions) [32]. Investigation of the link between interventions, environmental contamination and the profiles of pathogen carriage in childhood stools is an important topic for scientific research.

REFERENCES


How to cite this article: Rego R, Watson S, Gill P, Lilford R. The impact of diarrhoea measurement methods for under 5s in low- and middle-income countries on estimated diarrhoea rates at the population level: A systematic review and meta-analysis of methodological and primary empirical studies. Trop Med Int Health. 2022;00:1–22. https://doi.org/10.1111/tmh.13739
APPENDIX 1

THE NUMBER OF OBSERVATIONS CONTRIBUTED PER STUDY

<table>
<thead>
<tr>
<th>Number of observations per study</th>
<th>Number of studies (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>175 (60.14)</td>
</tr>
<tr>
<td>2</td>
<td>62 (21.31)</td>
</tr>
<tr>
<td>3</td>
<td>14 (4.81)</td>
</tr>
<tr>
<td>4</td>
<td>19 (6.53)</td>
</tr>
<tr>
<td>5</td>
<td>6 (2.06)</td>
</tr>
<tr>
<td>6</td>
<td>3 (1.03)</td>
</tr>
<tr>
<td>7</td>
<td>2 (0.69)</td>
</tr>
<tr>
<td>8</td>
<td>1 (0.34)</td>
</tr>
<tr>
<td>10</td>
<td>1 (0.34)</td>
</tr>
<tr>
<td>12</td>
<td>1 (0.34)</td>
</tr>
<tr>
<td>15</td>
<td>1 (0.34)</td>
</tr>
<tr>
<td>16</td>
<td>2 (0.69)</td>
</tr>
<tr>
<td>17</td>
<td>1 (0.34)</td>
</tr>
<tr>
<td>21</td>
<td>1 (0.34)</td>
</tr>
<tr>
<td>34</td>
<td>1 (0.34)</td>
</tr>
</tbody>
</table>

APPENDIX 2

SEARCH STRING FOR THE SYSTEMATIC REVIEW

diarrhea? or AWD or loose stool or watery stool or dysentery or ABD or fe?cal
   Developing Country.sh.
   (Africa or Asia or Caribbean or West Indies or South America or Latin America or Central America).hw,ti,ab,cp.
   (Afghanistan or Albania or Algeria or Angola or Antigua or Barbuda or Argentina or Armenia or Armenian or Aruba or Azerbaijan or Bahrain or Bangladesh or Barbados or Benin or Byelorussian or Belarus or Byelorussian or Belorussia or Belize or Bhutan or Bolivia or Bosnia or Herzegovina or Hercegovina or Botswana or Brazil or Brazil or Bulgaria or Burkina Faso or Burkina Faso or Upper Volta or Burundi or Cameroun or Cameroon or Cameroons or Cameroon or Comoros or Cape Verde or Central African Republic or Chad or Chile or China or Colombia or Comoros or Comoro Islands or Comores or Mayotte or Congo or Zaire or Costa Rica or Cote d’Ivoire or Ivory Coast or Croatia or Cuba or Cyprus or Czechoslovakia or Czech Republic or Slovakia or Slovak Republic or Djibouti or French Somaliland or Dominica or Dominican Republic or East Timor or East Timur or Timor Leste or Ecuador or Egypt or United Arab Republic or El Salvador or Eritrea or Estonia or Ethiopia or Fiji or Gabon or Gabonese Republic or Gambia or Gaza or Georgia Republic or Georgian Republic or Ghana or Gold Coast or Greece or Grenada or Guatemala or Guinea or Guam or Guiana or Guyana or Haiti or Honduras or Hungary or India or Maldives or Indonesia or Iran or Iraq or Isle of Man or Jamaica or Jordan or Kazakhstan or Kazakh or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirghizia or Kyrgyz Republic or Kirghiz or Kirgizistan or Lao PDR or Laos or Latvia or Lebanon or Lesotho or Basutoland or Liberia or Libya or Lithuania or Macedonia or Madagascar or Malagasy Republic or Malaysia or Malaya or Malay or Sabah or Sarawak or Malawi or Nyasaland or Mali or Malta or Marshall Islands or Mauritania or Mauritius or Agalega Islands or Mexico or Micronesia or Middle East or Moldova or Moldovia or Moldovan or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Burma or Namibia or Nepal or Netherlands Antilles or New Caledonia or Nicaragua or Niger or Nigeria or Northern Mariana Islands or Oman or Muscat or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Philipines or Philippine or Poland or Portugal or Puerto Rico or Romania or Rumania or Roumania or Russia or Russian or Rwanda or Ruanda or Saint Kitts or St Kitts or Nevis or Saint Lucia or St Lucia or Saint Vincent or St Vincent or Grenadines or Samoa or Samoan Islands or Navigator Island or Navigator Islands or Sao Tome or Saudi Arabia or Senegal or Serbia or Montenegro or Seychelles or Sierra Leone or Slovenia or Sri Lanka or Ceylon or Solomon Islands or Somalia or South Africa or Sudan or Suriname or Surinam or Swaziland or Syria or Tajikistan or Tadjikistan or Tadjikistan or Tadzhikistan or Tadjikistan or Tadzhikistan or Tanzania or Thailand or Togo or Togolese Republic or Tonga or Trinidad or Tobago or Tunisia or Turkey or Turkmenistan or Turkmenistan or Uzbekistan or Ukraine or Uruguay or USSR or Soviet Union or Union of Soviet Socialist Republics or Uzbekistan or Uzbek or Vanuatu or New Hebrides or Venezuela or Vietnam or Viet Nam or West Bank or Yemen or Yugoslavia or Zambia or Zimbabwe or Rhodesia).hw,ti,ab,cp.
   ((developing or less* developed or under developed or underdeveloped or middle income or low* income or under served or under served or deprived or poor*) adj (countr* or nation? or population? or world)).ti,ab.
   ((developing or less* developed or under developed or underdeveloped or middle income or low* income) adj (economy or economies)).ti,ab.
   (low* adj (gdp or gnp or gross domestic or gross national)).ti,ab.
   (low adj3 middle adj3 countr*).ti,ab.
   (lmic or lmics or third world or lami countr*).ti,ab.
   transitional countr*.ti,ab.
or/2-10 measurement or burden or survey or questionnaire or stool sample or collection or (Bristol adj2 scale) or (Amsterdam adj2 scale) or scale or pathogen or microbiological or biological or protein or olfaction or prevalence or incidence or odds or risk
   exp animals/ not humans.sh.
   1 and 11 and 12
   14 not 13
   Limit to ((english or french) and yr = "1993 -Current")
APPENDIX 3

STUDIES INCLUDED IN THE SYSTEMATIC REVIEW


53. Boisson S, Schmidt W-P, Berhanu T, Gezahgen H, Clasen T. Randomized controlled trial in rural Ethiopia to assess


79. Desmennu AT, Oluwasamu MM, John-Akinola YO, Oladunni O, Adebowale AS. Maternal Education and...


111. Geresomo NC, Mbuthia EK, Matofari JW, Mwangwela AM. Targeting caregivers with context specific behavior change training increased uptake of recommended hygiene practices during food preparation and complementary feeding in Dedza district of Central Malawi. Ecology of food and nutrition. 2018;57(4):301–313.


275. Watts H, Gregson S, Saito S, Lopman B, Beasley M, Monasch R. Poorer health and nutritional outcomes in orphans and vulnerable young children not explained...


