MISCELLANEOUS

Epidemiological methods in diarrhoea studies—an update

Wolf-Peter Schmidt,[1]* Benjamin F Arnold,[2] Sophie Boisson,[1] Bernd Genser,[3,4] Stephen P Luby,[5] Mauricio L Barreto,[3] Thomas Clasen[1] and Sandy Cairncross[1]

[1]Department for Disease Control, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK, [2]Division of Epidemiology, University of California, Berkeley, CA, USA, [3]Instituto de Saúde Coletiva, Federal University of Bahia, Salvador, Brazil, [4]Mannheim Institute of Public Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany and [5]International Centre for Diarrhoeal Disease Research Bangladesh, Dhaka, Bangladesh

*Corresponding author. Department for Disease Control, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: wolf-peter.schmidt@lshtm.ac.uk

Accepted 14 September 2011

Background Diarrhoea remains a leading cause of morbidity and mortality but is difficult to measure in epidemiological studies. Challenges include the diagnosis based on self-reported symptoms, the logistical burden of intensive surveillance and the variability of diarrhoea in space, time and person.

Methods We review current practices in sampling procedures to measure diarrhoea, and provide guidance for diarrhoea measurement across a range of study goals. Using 14 available data sets, we estimated typical design effects for clustering at household and village/neighbourhood level, and measured the impact of adjusting for baseline variables on the precision of intervention effect estimates.

Results Incidence is the preferred outcome measure in aetiological studies, health services research and vaccine trials. Repeated prevalence measurements (longitudinal prevalence) are appropriate in high-mortality settings where malnutrition is common, although many repeat measures are rarely useful. Period prevalence is an inadequate outcome if an intervention affects illness duration. Adjusting point estimates for age or diarrhoea at baseline in randomized trials has little effect on the precision of estimates. Design effects in trials randomized at household level are usually <2 (range 1.0–3.2). Design effects for larger clusters (e.g. villages or neighbourhoods) vary greatly among different settings and study designs (range 0.1–25.8).

Conclusions Using appropriate sampling strategies and outcome measures can improve the efficiency, validity and comparability of diarrhoea studies. Allocating large clusters in cluster randomized trials is compromised by unpredictable design effects and should be carried out only if the research question requires it.

Keywords Diarrhoea, cluster randomized trial, sampling, methods

Introduction

Diarrhoeal diseases remain a leading cause of morbidity and mortality in children worldwide.[1] Reliable field data from epidemiological studies are required to study diarrhoea epidemiology and the effect of interventions,[2,3] but diarrhoea remains a condition difficult to measure.[4] Systematic reviews of diarrhoea interventions have found a great variety of approaches to measure diarrhoea.[5,6] The past decade saw a trend towards less intensive active diarrhoea surveillance,[7] the use of repeated diarrhoea prevalence measures instead of incidence as outcome measure[8] and a greater recognition of recent advances in the design of cluster randomized trials.[9,10] In this article we review current practices in conducting epidemiological studies on diarrhoeal diseases with an emphasis on randomized controlled trials (RCTs) in low-income populations, including cluster randomized trials. We discuss crucial methodological problems to be considered in the planning stage of a trial, but several issues should also be relevant for observational studies.

Literature search methods and data sets

We searched the database MEDLINE for the years 1970–2009 without language restrictions, using the search terms [diarrh(o)ea AND trial], [diarrh(o)ea AND measurement], [diarrh(o)ea AND recall] and [diarrh(o)ea AND longitudinal prevalence]. We screened the reference lists of relevant articles and contacted authors and experts in the field for further identification of relevant articles. We further used original data sets from different field sites across the world (in part described previously[11]) to address issues of design effect and adjustment for baseline variables in RCTs. These data sets came from the authors of this article or were made available to us by other researchers in the field (see 'Acknowledgements').

Reporting and recording diarrhoea symptoms

Case definitions for diarrhoea commonly are either based on reported signs and symptoms (stool frequency, presence of blood or mucus) or based on local disease perception. For example, a study in Ghana identified seven different local terms for symptoms compatible with diarrhoea.[12] Relying on local disease definitions requires extensive qualitative research and piloting,[13] but such work can provide important insights that are useful for a study as a whole. Most studies continue to use the WHO definition of diarrhoea,[14] defined as 'the passage of 3 or more loose or liquid stools per day, or more frequently than is normal for the individual'.[15,16] A stringent definition that does not depend on local disease concepts may reduce subjectivity and perhaps also the risk of bias but this has not yet been shown in practice. While not necessarily having more clinical validity, using the WHO definition facilitates comparison across sites (Box 1). Asking study participants specifically for the presence or absence of '3 or more loose or liquid stools per day' may unnecessarily force a decision by the respondent that may be prone to bias. Therefore, some diarrhoea trials record stool frequency and then apply the WHO definition post-hoc.[17,18]

Studies have shown that the longer the recall period, the greater the imprecision (especially underestimation) of prevalence estimates.[13,19–23] By assuming that reported prevalence in the last 24 h was 100% accurate, these studies may have overestimated recall error, since the higher diarrhoea prevalence closer to the day of the visit may indicate that people remember diarrhoea during the past 7 days as having occurred more recently than was actually the case. In a study from Peru, mothers reported the correct prevalence of diarrhoea but often were inaccurate in reporting the exact day when it occurred.[24] Recall error depends on the severity and duration of symptoms.[23] A decline in reported diarrhoea with time on study (independent of treatment) has been noted in diverse populations.[10,25–29] Intensive surveillance including frequent home visits can lower the reported diarrhoea prevalence,[30] perhaps due to 'reporting fatigue'. Recall can be more complete in groups of higher socio-economic status, leading to bias when comparing different populations.[31]

Recall error may not be a big problem in studies exploring disease trends or comparing diarrhoea risk between treatment arms, if it can be assumed that recall error is non-differential among the groups compared. This assumption, however, is difficult to verify in unblinded trials. There are numerous theoretical possibilities for treatment effects to be biased. For example, allocation to the control group may lead to diarrhoea episodes being remembered more acutely out of frustration of not receiving the intervention. Alternatively, allocation to a treatment group may lead participants to not report disease episodes or field staff to not record disease episodes under the expectation that the intervention is effective. Due to such biases, even a diarrhoea reduction of 50% observed in unblinded trials may be compatible with no true effect.[32]

Given the complexity of validating reported diarrhoeal disease in community-based surveys, investigators should take steps to minimize measurement error whenever possible. A 7-day recall period is commonly used in diarrhoea trials,[33] but a shorter recall period may reduce subjectivity of reporting and possibly bias in unblinded trials. Using a 2- or 3-day recall often leads to only a small to moderate loss of power compared with a 7-day recall period, especially if diarrhoea is common, if the number of measurements per individual exceeds 10 or 12 and in cluster randomized trials.[34] Instead of asking for diarrhoea in the previous 24, 48 or 72 h, one could consider asking for

whole calendar days only (Did you have diarrhoea today? Yesterday? The day before yesterday?). Such questions are usually easier to ask and to answer. Numerous studies have demonstrated that symptom recall beyond 7 days is unreliable, and we do not recommend it.[13,19–23] In any case, the final choice of the recall period should only be done after pilot-testing different approaches in a given setting.

Incidence or longitudinal prevalence as outcome measure

Diarrhoea can be measured as incidence (new episodes per person-time) or prevalence (disease presence at time t, Box 1). Incidence does not account for episode duration,[8] an important risk factor for adverse outcomes.[35,36] In settings where diarrhoea is common, it can be difficult to distinguish one episode from the next. Two to three days have been suggested as the most appropriate period to separate distinct episodes,[15,37] an approach widely in use today. If diarrhoea is quite rare, it makes sense to use a longer gap (e.g. 6 days[27,38]) to separate distinct episodes, since episodes are unlikely to occur close together by chance. Such definitions are, to some extent, arbitrary and will inevitably cause some misclassification.[14,39] Methods have been developed to allow the comparison of studies using different definitions.[14] Measuring incidence especially in high-risk settings can require close disease surveillance (e.g. one to three times/week) to establish beginning and end of episodes.[18] However, a rough incidence estimate can be obtained by repeated period prevalence measurements assuming that diarrhoea preceded by a period without diarrhoea represents a new episode.[40]

Incidence is an appropriate measure if the duration of illness is not of particular interest. A new episode can be interpreted as a case of pathogen transmission to a new host, which for disease control or vaccine research can be more important than episode duration. This applies to disease surveillance in middle- and high-income settings with low risk of malnutrition and diarrhoea-related mortality. For example, a study in the UK compared the incidence of diarrhoea in the community with cases reported to surveillance agencies.[41] The duration of episodes was of little importance. Health service and vaccine researchers are often more interested in incident episodes than prevalence, focusing on the incidence of episodes with pre-defined characteristics, e.g. episodes of long duration or with blood/mucus, or watery diarrhoea for the surveillance of cholera.[36,42] Such studies often use passive case finding instead of intensive active surveillance, e.g. by measuring the incidence of hospital admissions. This approach allows obtaining detailed clinical data and causative agents assessed by health professionals, often at a higher standard compared with field data. Measuring the incidence of hospital admission biases the data towards severe episodes, which are often the episodes of highest public health interest. Since only a fraction of diarrhoea episodes are seen at hospitals, the study population receiving the intervention will have to be large. On the other hand, there is no need for repeated surveillance visits. Passive recording of the incidence of hospital admissions may be less prone to observer and responder bias than diarrhoea incidence recorded through active surveillance because, although bias cannot be excluded, study participants are less likely to decide on health-care use based on treatment allocation. If the aim of a study is to obtain detailed clinical data on all, not just severe, episodes, close active surveillance (e.g. contacting participants at least once a week) is usually required, especially if stool samples are collected.[41]

Outside clinical studies, prevalence rather than incidence is often the outcome measure of choice, especially if prevalence can be measured repeatedly in the same individual. Repeated measurements provide an estimate of an individual's proportion of time ill, also termed 'longitudinal prevalence' (LP).[8] The ideal settings for using LP as outcome are low-income, high-risk populations where preventing adverse outcomes such as death and malnutrition is important. LP is a better predictor of such complications than incidence.[8,43] Table 1 shows the results from two large RCTs conducted in Guatemala[44] (Household water treatment intervention) and Brazil[17] (Vitamin A supplementation). In the Guatemala trial, the interventions reduced the incidence of diarrhoea by 24%, whereas the mean LP (days with diarrhoea/days observed) was reduced by only 14%. This was because the intervention mostly prevented short episodes.[44] In contrast, the Brazil study achieved only a small reduction in the incidence, which, however, masks the impact of the intervention on the duration of illness, leading to an LP reduction of 12% (note the differences in the P-values for LP vs incidence). In both cases it can be argued that longitudinal prevalence is the more appropriate way to measure public health impact.

Individuals tend to differ more in the number of disease days than in the number of episodes they experience, since the variation in the duration of episodes increases the standard deviation (SD) of LP compared with incidence.[11] LP studies may require a larger sample size than incidence studies, if the

Table 1 Incidence vs longitudinal prevalence of diarrhoea: impact on study results and interpretation in two randomized trials

	Incidence reduction (%)	P	Mean LP reduction (%)	P
Guatemala ($n = 2982$)				
Water treatment	−24	0.001	−14	0.185
Brazil ($n = 1180$)				
Vitamin A	−7	0.18	−12	0.06

Table 2 Comparison between incidence and longitudinal prevalence

	Incidence	LP
Suitable setting	Low diarrhoea risk	High diarrhoea risk
	Malnutrition and case fatality uncommon	Malnutrition and case fatality a public health problem
Suitable research objectives	Disease surveillance and control	Burden of disease
	Health services research	Adverse outcomes
	Vaccine research	Nutrition studies
	Aetiological research	
Data interpretation	Disease transmission	Burden of disease
		Risk of adverse outcomes
Definition to separate episodes	Required	Not required
Sampling frequency	Usually requires frequent and regular sampling, unless passive surveillance is used	Sampling at long or irregular intervals possible and often logistically efficient
Study power	Larger than for LP if exposure or treatment has no effect on episode duration	Larger than for incidence if exposure or treatment reduces episode duration

exposure variable has no effect on episode duration.[34] If, however, the exposure variable is associated with shorter episodes (as in the Brazil Vitamin A study), using LP increases power because the effect size should be larger (Table 1). If an intervention reduces predominantly short episodes as in Guatemala (Table 1), incidence may be more powerful, but may also be less informative for public health purposes. Table 2 summarizes advantages and disadvantages of using incidence vs prevalence measurements.

At what intervals should diarrhoea prevalence be measured?

Diarrhoea prevalence can be measured at long and irregular intervals because a prevalence measurement requires no information of when an episode started. Incidence may also be estimated by infrequent or irregular sampling, e.g. by assuming that any diarrhoea occurring within the recall period is a new episode if no diarrhoea was present early in the recall period. This, however, is when recall error may be greatest, making such incidence estimates potentially unreliable.

Infrequent sampling can reduce costs[34] and may increase validity, since frequent measurements may compromise the willingness of participants to report illness. Frequent measurements may lead to a better compliance with the intervention and a lower reported prevalence of diarrhoea, at least if each visit includes procedures that are clearly related to the intervention (e.g. water testing in a household water chlorination intervention[30]).

Many repeat measurements of diarrhoea prevalence often provide little additional study power compared with fewer measurements.[34] Clustering of disease in high-risk individuals means that if an individual reports being diseased at the time of a survey, he/she is more likely to have been ill on any other day than an individual reported healthy. The more illness is clustered in individuals, the more disease absence or presence in an individual at one point in time is representative of the true disease experience. Consider as an example, a study in which weekly surveillance visits are conducted over 1 year, each time recording the daily point prevalence of diarrhoea over the past 7 days since the last visit (a 1-week recall period), an approach resulting in continuous daily diarrhoea data. It has been shown that a study in which visits are conducted every 4 weeks instead of every week (again using a 1-week recall period) only requires a 15–30% larger sample size, while reducing the number of visits by 75%.[34] In cluster randomized trials, the sample size increase in this example would even be smaller.[34] Many measurements in the same cluster (e.g. more than 12 per year) yield little additional power,[45] especially if within-cluster correlation of disease or cluster size is large.[34]

Of note, studies using diarrhoea as the 'exposure' variable require more precise estimates of an individual's burden of diarrhoea than studies with diarrhoea as 'outcome'.[46] For example, many studies have examined the effect of diarrhoea LP (the exposure variable) on mortality,[8] malnutrition[43,47–52] or the risk of other infectious diseases (the outcomes).[53,54] Imprecision in the measurement of diarrhoea as an 'exposure' variable (e.g. due to infrequent sampling) usually biases the effect estimate towards no effect ('regression dilution bias').[55] Often, more than 15 to 20 visits will be required to limit bias.[46] Also, when measuring diarrhoea as an 'exposure' variable, a short

recall period (e.g. 3 days) may be preferable to minimize bias.[46] If diarrhoea is the 'outcome' measure, imprecise diarrhoea estimates due to infrequent visits will only affect the precision of the effect estimate, not the effect size.[55]

Temporary absence of study participants and logistical constraints can cause prevalence measurements to be taken at irregular intervals. This is not a problem if irregularity occurs at random or at least similarly between comparison groups, which often should be the case. The later analysis should be weighted by the number of measurements in an individual.[56]

Point prevalence vs period prevalence

While some investigators choose to record point prevalence ('On which of the last X days did you suffer from diarrhoea?'), others collect period prevalence data ('Did you have diarrhoea at any time during the last X days?', Box 1). Period prevalence data are often used in large demographic and health surveys. Recording period prevalence may be simpler, but can reduce the difference (if expressed as a prevalence ratio) between two study groups because an individual with an episode several days long may be recorded as having the same disease experience as a person in the other group with only 1 disease day, i.e. period prevalence data bias the prevalence ratio towards no effect, especially if the disease is common (e.g. more than five episodes per person-year).[34]

Perhaps counter-intuitively, period prevalence as an imprecise outcome measure can achieve a higher study power than point prevalence even if the recall period is the same, because differences between individuals (i.e. the coefficient of variation of the mean LP) are reduced.[34] However, period prevalence data are inappropriate to capture changes in illness duration. Effect sizes will be strongly biased towards no effect if an intervention primarily works by reducing episode duration,[34] and will be exaggerated if an intervention primarily reduces short episodes (Table 1, Guatemala study).

To conclude, investigators need to balance the advantages of using period prevalence data (easy to collect, slightly more powerful in many situations) with the risk of bias, which depends on the effect of the factor under study on illness duration. The collection of daily point prevalence with a limited recall period provides flexibility to use either outcome measure, but investigators should specify in advance which is to serve as the main study outcome to protect against selectively choosing a measure that provides the result most aligned with the investigators' pre-conception.

Adjusting for baseline diarrhoea and age

In many trials, investigators measure diarrhoea at baseline (before randomization). In general, baseline measurements in trials may serve to (i) verify randomization success, (ii) adjust the final analysis for imbalances and (iii) increase precision of the

treatment effect by including the baseline measure as a covariate in an adjusted analysis. The latter two uses require that the baseline measure be strongly associated with the later outcome to be effective.[57,58]

Concerns over imbalances in diarrhoea prevalence at baseline have in the past prevented or severely delayed publication of trials.[59] However, caution is warranted in interpreting baseline diarrhoea data, specifically when used to verify the success of randomization. Most demographic variables commonly assessed at recruitment, such as date of birth, gender, family size or socio-economic status, do not change rapidly (if at all) and may later be used to adjust for imbalances. In contrast, diarrhoea prevalence is highly variable over time.[60] If an individual has diarrhoea at baseline, it indicates that they may be more prone to diarrhoea during the follow-up period, but this depends on the within-person clustering of disease in a given setting. Typically, diarrhoea trials are designed to detect a certain difference between trial arms given a pre-specified number of repeat measurements (often more than 10), assuming a chance of false positivity of, say 0.05. A 'single' measurement at baseline in the same number of people has a considerable chance of suggesting a relevant imbalance where there may be none. It has been suggested that multiple baseline measurements collected during a run-in period could improve the efficiency of studies with both continuous and incidence rate outcomes,[61] but this may not necessarily apply to diarrhoea. For example, Figure 1 plots the village-level diarrhoea incidence in 11 control villages from a randomized trial of solar water disinfection in Bolivia. The baseline diarrhoea measurement included 6 weeks of surveillance (six measurements per individual) that were collected 6 months before the intervention. As Figure 1 illustrates, baseline incidence bears no relation to incidence during the year-long intervention period. Several factors may have contributed to this, such as the long gap between baseline measurement and the actual trial and, in particular, the high spatial and temporal variability of diarrhoea often observed in the field.[60] This contrasts with strong associations between baseline village HIV prevalence and subsequent incidence,[62] or between baseline height-for-age Z-scores and subsequent height measurements.[63]

Table 3 shows the effect of adjusting for baseline diarrhoea (single measurement) or age on the effect estimate and standard error (SE) in studies available to us (age usually is a strong predictor of diarrhoea). In some cases, adjusting for baseline diarrhoea or age can have a relevant effect on the effect estimates (e.g. Kenya and Colombia), and often reduces the SE. However, adjusting for covariates in RCTs by using statistical models in general can lead to bias, and should be conducted with caution.[64] The protocol for adjusted analyses in randomized trials to gain study power or reduce bias should be pre-specified[58] and reserved for large studies, where statistical models

may be less biased.[64] The age adjustments shown in Table 3, however, do not suggest a great gain in study power in large studies. In cluster randomized trials the gain in power due to baseline adjustment may be even lower than in individually randomized trials, especially if the between-cluster variation is high.[57] Based on these results and Figure 1, we infer that baseline diarrhoea would make a poor matching or stratification variable in a trial's design.

To conclude, a single baseline measurement of diarrhoea should primarily be useful to confirm trial procedures and familiarize study participants and field
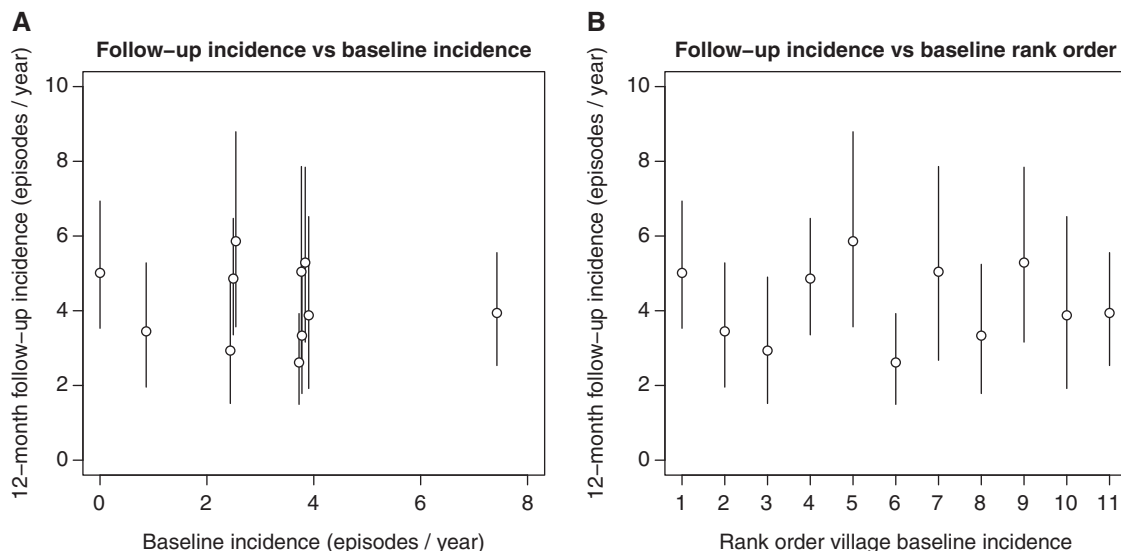

Figure 1 Village-level diarrhoea incidence during a 12-month follow-up period in 11 control villages that participated in an intervention trial of solar water disinfection.[10] Vertical lines mark bootstrapped 95% confidence intervals. The follow-up incidence is plotted against baseline incidence measured over a 6-week period (A), and against the village rank in baseline incidence over that same period (B)

Table 3 Effect of adjusting for baseline diarrhoea or age on point estimate and SE

References	Country	Age range (years)	N	Crude analysis			Adjusted analysis			SE change (%)
				PR	SE	P	PR	SE	P	
Adjustment for baseline diarrhoea										
Clasen et al.[65]	Bolivia	0–80	317	0.55	0.16	0.042	0.52	0.16	0.038	+1
Boisson et al.[66]	Congo	0–84	1144	0.85	0.15	0.336	0.88	0.15	0.447	+1
Colford et al.[67]	USA	55–95	770	0.90	0.06	0.119	0.90	0.06	0.123	+0.1
Clasen et al.[68]	Colombia	0–82	684	0.54	0.14	0.017	0.54	0.13	0.015	−1
Boisson et al.[69]	Ethiopia	0–91	1516	0.75	0.09	0.011	0.74	0.08	0.007	−3
Trotta[59]	Peru	0.5–1.5	483	0.98	0.20	0.902	1.03	0.18	0.850	−9
Tiwari et al.[70]	Kenya	<15	216	0.37	0.13	0.004	0.31	0.12	0.002	−9
Adjustment for age										
Tiwari et al.[70]	Kenya	<15	216	0.37	0.13	0.004	0.37	0.13	0.005	+1
Colford et al.[67]	USA	55–95	770	0.90	0.06	0.119	0.91	0.06	0.129	+0.3
VAST[12]	Ghana	0–5	1918	0.99	0.01	0.316	1.01	0.01	0.289	−1
Reller et al.[44]	Guatemala	0–80	2980	0.86	0.08	0.106	0.86	0.08	0.112	−2
Boisson et al.[69]	Ethiopia	0–91	1516	0.75	0.09	0.011	0.75	0.08	0.009	−3
Boisson et al.[66]	Congo	0–84	1144	0.85	0.15	0.336	0.88	0.14	0.434	−3
Clasen et al.[68]	Colombia	0–82	684	0.54	0.14	0.017	0.43	0.13	0.010	−6
Clasen et al.[65]	Bolivia	0–80	317	0.55	0.16	0.042	0.48	0.12	0.005	−23

Age adjustment was made with age as categorical variable (<1 year, 1 to <2 years, 2 to <3 years, 3 to <5 years, 5 to <10 years, 10 to <15 years, ⩾15 years), except for the US elderly population (55–64, 65–74, 75–84 and 85–95 years); PR, prevalence ratio.

staff with measurement procedures. Occasionally, it has been observed that the first or a single measurement in a trial may provide implausibly high estimates compared with follow-up visits.[66,71,72] Participants concerned about potentially not being included in a trial may over report the disease at first visit. A baseline measurement that would not be included in a later analysis may limit the impact of this possible effect.

Group-level clustering and design effect

Many diarrhoea studies need to consider clustering of diarrhoea in households or villages/neighbourhoods, e.g. if an intervention is randomized at group level. The effect of clustering can be expressed as the design effect DEFF, the factor by which the sample size needs to be increased to account for clustering:[73]

$$\text{DEFF} = 1 + (m - 1) \times \text{ICC}$$

where m is the number of individuals per cluster, and ICC is the intra-cluster correlation coefficient.[9] Estimating ICC and DEFF is one of the most challenging aspects in complex diarrhoea trials. Both depend on factors such as (i) mean number of persons per cluster, (ii) mean number of measurements per person, (iii) within-person correlation of diarrhoea (which strongly depends on the age range included) and (iv) the differences in diarrhoea risk between clusters (i.e. the between-cluster variability). In areas where a substantial proportion of diarrhoea occurs as localized epidemics shifting from place to place, between-cluster variability (i.e. ICC and DEFF) will be high because some areas may be experiencing an outbreak at the time of study, whereas others are not. In addition, the DEFF increases if cluster size and number of measurements per individual vary,[74] which is usually the case in field studies.

Calculating the DEFF for diarrhoea as a binary outcome based on an ICC estimate is not straightforward, perhaps best highlighted by the many different methods available.[75,76] Estimating the ICC treating diarrhoea LP as a continuous outcome is problematic since follow-up time usually differs between individuals.

Alternatively, the DEFF can be estimated directly from the SEs of the log prevalence ratio or log rate ratio resulting from clustered and unclustered analyses:[9]

$$\text{DEFF} = \frac{\text{SE}^2_{\text{clustered}}}{\text{SE}^2_{\text{unclustered}}}$$

where $\text{SE}_{\text{clustered}}$ is the standard error from an analysis accounting for clustering, and $\text{SE}_{\text{unclustered}}$ is the standard error from an analysis ignoring clustering. We calculated DEFFs from the data of several randomized trials available to us using this formula (for details, see footnote of Table 4). We calculated DEFFs separately for within-person and within-cluster correlation of disease to show the effect of

group-level clustering in addition to the design effect due to within-person correlation.

DEFFs for 'household' clustering are quite similar across studies, ranging from one to approximately three regardless of the study design (Table 4). In contrast, we found very different design effects of up to 22 if the unit of clustering was large (villages or neighbourhoods). In one case (urban Brazil), the design effect was much smaller for the analysis accounting for neighbourhood clustering compared with the analysis accounting for within-person clustering only. In this setting an individually randomized trial may require a larger sample size than a cluster randomized trial, because children in the same cluster had very different diarrhoea risks, whereas the cluster-level diarrhoea risks were similar. For six studies with continuous diarrhoea records we did the same calculation for incidence of new episodes [Table 4, DEFFs in brackets], mostly resulting in much lower within-person DEFFs and slightly higher household DEFFs compared with prevalence data. The DEFFs for incidence vs prevalence due to village/neighbourhood clustering were quite different in three of the six studies (rural Bolivia, rural Pakistan and urban Brazil).

Overall, DEFFs in trials randomizing large clusters are difficult to predict unless previous data from the same site are available. Randomization of large clusters should perhaps be 'avoided like the plague' unless the research question requires it.[77]

The DEFFs due to 'within-person' correlation very strongly depend on the number of measurements (Table 4), showing again that many repeated measurements contribute little to study power. Continuous surveillance of daily point prevalence generally results in extremely large within-person DEFFs because of high day-to-day correlation. DEFFs are much reduced if measurements are either reduced to period prevalence, or separated by intervals between measurements. Repeat measures add to the complexity of sample size calculations for cluster randomized diarrhoea trials, since the number of measurements also affects 'group level' ICC and hence DEFF.[34] Sample size calculations for diarrhoea trials may have to be pragmatic and, even more so than for diseases that do not recur, undergo an iterative process testing different sampling intervals and cluster sizes. Several approaches are available.[9] If diarrhoea is common, it can make sense to treat diarrhoea as a continuous variable (e.g. LP or number of episodes per person) and remove one level of complexity. This requires knowledge of the mean LP or number of episodes and the SD given a specified number of measurements (examples have been published elsewhere[34]). The sample size resulting from simple formulae for the comparison of two means[73] can be multiplied by a group-level DEFF deemed appropriate (Table 4). Note that the presence of several levels of clustering (e.g. person, household and area) does

Table 4 Design effects in diarrhoea studies

Authors	Country	Age range (years)	Individuals (N)	Follow-up method	Weeks follow-up per person, mean (SD)	Outcome measure	Mean LP (% days or weeks ill)	Persons per cluster, mean (SD)	DEFF overall	DEFF due to clustering by person	DEFF due to clustering by household or village
Household clustering											
Clasen et al.[65]	Bolivia (rural)	0–80	317	Monthly visits with 7-day recall	4.8 (0.5)	Weekly period prevalence	3.7	5.4 (2.4)	1.3	1.3	1.0
Colford et al.[67]	USA (rural)	55–95	770	Weekly visits with 7-day recall	45.1 (13.6)	Weekly period prevalence	7.4	1.4 (0.5)	2.4 (1.4)[a]	2.3 (1.3)[a]	1.0 (1.0)[a]
Reller et al.[79]	Guatemala (rural)	0–80	2980	Weekly visits with 7-day recall	52 (0)	Daily point prevalence	2.5	6.3 (2.4)	46.8 (7.1)[a]	44.0 (5.9)[a]	1.1 (1.2)[a]
Arnold et al.[80]	India (rural)	0–5	1184	Monthly visits with 7-day recall	11.1 (2.5)	Weekly period prevalence	1.8	1.4 (0.6)	1.4	1.3	1.1
Luby et al.[81]	Bangladesh (rural)	0–5	1278	Monthly visits with 2-day recall	4.7 (1.2)	2-days period prevalence	11.0	1.3 (0.5)	3.1	2.5	1.2
Boisson et al.[69]	Ethiopia (rural)	0–91	1516	Monthly visits with 7-day recall	10 (1.2)	Weekly period prevalence	3.5	4.8 (2.4)	1.4	1.2	1.2
Tiwari et al.[70]	Kenya (rural)	0–15	216	Monthly visits with 7-day recall	5.3 (1.4)	Daily point prevalence	3.6	3.3 (1.4)	7.4	5.7	1.3
Mausezahl et al.[10]	Bolivia (rural)	0–5	725	Weekly visits with 7-day recall	32.9 (9.9)	Weekly period prevalence	9.4	1.7 (0.7)	14.2 (6.1)[a]	10.1 (4.0)[a]	1.4 (1.5)[a]
Boisson et al.[66]	DRC (rural)	0–84	1144	Monthly visits with 7-day recall	10.6 (1.2)	Weekly period prevalence	3.5	4.8 (2.4)	2.5	1.8	1.4
Luby et al.[82]	Pakistan (urban)	0–15	4691	Twice-weekly visit with 3–4 day recall	44.8 (11.0)	Weekly period prevalence	3.4	5.2 (2.2)	6.2 (4.5)[a]	4.3 (2.8)[a]	1.4 (1.6)[a]
Van der Hoek et al.[83]	Pakistan (rural)	0–80	1500	Weekly visits with 7-day recall	58.4 (10.0)	Daily point prevalence	1.5	6.7 (2.8)	57.5 (7.5)[a]	37.5 (4.8)[a]	1.5 (1.6)[a]
Luby et al.[84]	Pakistan (urban)	0–66	8949	Twice-weekly visit with 3–4 day recall	34.2 (7.6)	Weekly period prevalence	5.0	6.7 (2.6)	5.0 (3.7)[a]	3.3 (2.0)[a]	1.5 (1.9)[a]
Clasen et al.[68]	Colombia (rural)	0–82	684	Monthly visits with 7-day recall	3.8 (0.5)	Weekly period prevalence	7.1	5.0 (2.5)	3.2	1.5	2.1
Clasen et al.[85]	Bolivia (rural)	1–86	278	Monthly visits with 7-day recall	3.9 (0.3)	Weekly period prevalence	16.7	5.6 (1.8)	4.6	1.4	3.2
Village/neighbourhood clustering											
Barreto et al.[86]	Brazil (urban)	0–5	1880	Thrice-weekly visit with 2–3 day recall	37.6 (17.9)	Daily point prevalence	2.9	76.4 (24.5)	1.0	10.2	0.1
Reller et al.[79]	Guatemala (rural)	0–80	2980	Weekly visits with 7-day recall	52 (0)	Daily point prevalence	2.5	248.5 (159.7)	50.2 (6.1)[a]	44.0 (4.7)[a]	1.1 (1.3)[a]
Trotta[59]	Peru (urban)	0–2	483	Weekly visits with 7-day recall	6.7 (0.6)	Daily point prevalence	9.6	24.2 (1.0)	6.6	5.5	1.2
Arnold et al.[80]	India (rural)	0–5	1184	Monthly visits with 7-day recall	11.1 (2.5)	Weekly period prevalence	1.8	51.4 (17.4)	1.6	1.3	1.3
Mausezahl et al.[10]	Bolivia (rural)	0–5	725	Weekly visits with 7-day recall	32.9 (9.9)	Weekly period prevalence	9.4	33.0 (7.6)	19.2 (7.4)[a]	10.1 (4.0)[a]	1.9 (1.8)[a]
Van der Hoek et al.[83]	Pakistan (rural)	0–80	1500	Weekly visits with 7-day recall	58.4 (10.0)	Daily point prevalence	1.5	146.8 (66.7)	115.1 (38.2)[a]	37.5 (4.8)[a]	3.1 (7.9)[a]
Luby et al.[81]	Bangladesh (rural)	0–5	1278	Monthly visits with 2-day recall	4.7 (1.2)	2-days period prevalence	11.0	12.4 (2.7)	14.8	2.7	5.5
Luby et al.[82]	Pakistan (urban)	0–15	4691	Twice-weekly visit with 3–4 day recall	44.8 (11.0)	Weekly period prevalence	3.4	130.3 (34.2)	56.7 (46.8)[a]	4.3 (2.8)[a]	13.2 (16.6)[a]
Luby et al.[84]	Pakistan (urban)	0–66	8949	Twice-weekly visit with 2–3 day recall	34.2 (7.6)	Weekly period prevalence	5.0	190 (42.7)	82.0 (51.5)[a]	3.2 (2.0)[a]	25.8 (25.4)[a]

Design effects were calculated as DEFF = $SE_{clustered}^2/SE_{unclustered}^2$; diarrhoea was treated as a binary variable. Prevalence of diarrhoea (days or weeks with diarrhoea over days or weeks observed) was compared between treatment and control using log-binomial regression (family = binomial, link = log). Clustering was accounted for by using robust SE. $SE_{unclustered}$ was calculated ignoring any within-person or within-group correlation of diarrhoea. $SE_{clustered}$ to account for clustering by person was estimated with person identification (ID) as the cluster variable. $SE_{clustered}$ to account for clustering at group level used the group ID as cluster variable (the unit of randomization). If the original data provided no treatment variable, clusters were allocated post-hoc at random to a simulated treatment and control arm. DEFF due to clustering at group level was calculated as $DEFF_{overall}/ DEFF_{person}$.
[a]DEFFs show the same analyses for incidence (new episodes/days or weeks observed), calculated analogously to prevalence, using Poisson regression without ($SE_{unclustered}$) and with robust SE to account for clustering ($SE_{clustered}$).

Table 5 Examples of different epidemiological studies and suggested sampling strategy

	Study example	Suggested sampling strategy
1	Context: RCT of household level food hygiene promotion to reduce the burden of diarrhoea, delivered by community health workers to mothers of young children Study population: children aged <5 years Logistics: adequate budget, trained staff and large eligible population available	Outcome measure: LP Surveillance duration: 1 year Sampling frequency: every 6–8 weeks (∼6–9 contacts) Recall period: 3 days Data type: point prevalence Comment: Incidence is not suitable as the treatment aims to lower disease burden, for which LP is likely to be a better measure Sampling at intervals (with a corresponding increase in the overall sample size) is chosen to decrease survey effects and bias. 3-day recall is chosen to minimize recall error. The study is done over 1 year to study potential seasonal effects in food contamination. For the sample size within-household clustering can be ignored as the average number of young children per household is usually small (less than two).
2	Context: RCT of household level food hygiene promotion to reduce the burden of diarrhoea, delivered by community health workers to mothers of young children Study population: children aged <5 years Logistics: tight budget, trained staff scarce, large eligible population available	Outcome measure: LP Surveillance duration: 5 months Sampling frequency: every 4 weeks (∼6 contacts) Recall period: 7 days Data type: period prevalence Comment: Incidence is not suitable as the treatment aims to lower disease burden, for which LP is likely to be a better measure. Sampling at intervals (with a corresponding increase in the overall sample size) is chosen to decrease survey effects and bias. 7-day recall (period prevalence) is chosen to maximize power. The study is restricted to 5 months because of the tight budget, focussing on the hot season where food contamination may be most common. For the sample size within-household clustering can be ignored as the number of young children per household is small
3	Context: RCT of household level food hygiene promotion to reduce the burden of diarrhoea, delivered by community health workers to mothers of young children Study population: children aged <5 years Logistics: tight budget, trained staff scarce, eligible population small (e.g. refugee camp)	Outcome measure: LP Surveillance duration: 5 months Sampling frequency: every 2 weeks (∼12 contacts) Recall period: 3 days Data type: point prevalence Comment: Incidence is not suitable as the treatment aims to lower disease burden, for which LP is likely to be a better measure. Frequent sampling is chosen to make the most of the small sample size. Short recall (point prevalence) is chosen to minimize recall error. Because of the short visit intervals, longer recall periods do not add much power.[56] The study is restricted to 5 months because of the tight budget, focussing on the hot season where food contamination may be most common. For the sample size within-household clustering can be ignored as the average number of young children per household is usually small (less than two).

(continued)

Table 5 Continued

	Study example	Suggested sampling strategy
4	Context: RCT of a new vaccine against a pathogen causing severe diarrhoea Study population: children aged <5 years Logistics: adequate budget, trained staff and large eligible population available	Outcome measure: incidence Surveillance duration: 12 months Sampling approach: passive surveillance of hospital admissions Recall period: Not applicable (NA) Data type: incidence of severe episodes
		Comment: Incidence is suitable as the treatment aims to lower disease transmission of a specific pathogen. Passive surveillance is chosen because a vaccine can be delivered relatively easily to a large study population, focussing on episodes of particular clinical interest. Because hospital admissions do not allow estimating the effect of the vaccine on LP (a better marker for adverse effects on nutritional status), one could consider adding a substudy with active surveillance similar to Example 1, as was done in a Vitamin A trial in Ghana.[12]
5	Context: cluster RCT of a large rural sanitation programme delivered at village level Study population: all ages Logistics: tight budget, trained staff scarce, large eligible population available	Outcome measure: LP Surveillance duration: 1 year Sampling frequency: every 6–8 weeks (~6–9 contacts) Recall period: 7 days Data type: period prevalence
		Comment: Incidence is not suitable as the treatment aims to lower disease burden, for which LP is likely to be a better measure. Sampling at long intervals (with a corresponding increase in the number of included villages) is chosen to limit the number of surveillance teams and transport costs. The sampling procedure aims to measure the outcome in one village per day per team. In a cluster randomized trial, more frequent surveillance rounds add relatively little power. 7-day recall (period prevalence) is chosen to maximize power. Data on 3-day point prevalence can be obtained in addition as a secondary outcome. The effect of the intervention in children aged <5 years can be a secondary outcome. Because of the great uncertainties in study power due to the cluster-design, it is preferable to include all household members to maximize power. This is specifically the case if there is little reason to assume the intervention will affect young and older ages differently.
6	Context: observational study with recurrent infections as exposure (e.g. to study association between diarrhoea and reduction in weight-for-age Z-score) Study population: children aged <5 years Logistics: adequate budget, trained staff and large eligible population available	Outcome measure: LP Surveillance duration: 1 year Sampling frequency: every 2 weeks (~20–25 visits) Recall period: 3 days Data type: point prevalence
		Comment: Frequent sampling is chosen to minimize bias towards no effect. Efforts should be made to keep study participants happy and interested. Bias is not a great concern in an observational study without differential treatment of study participants. Short recall period is chosen to minimize bias that could exaggerate the effect size.

(continued)

Table 5 Continued

	Study example	Suggested sampling strategy
7	Context: observational study >1 year, aimed at detailed exploration of clinical features of individual episodes (e.g. illness duration, severity, clinical signs and symptoms, stool testing for pathogens) Study population: children aged <5 years Logistics: adequate budget, trained staff and large eligible population available	Outcome measure: incidence Surveillance duration: 1 year Sampling frequency: once a week (~50 contacts) Recall period: 7 days Data type: point prevalence data from which incidence can be calculated Comment: Frequent sampling is chosen to accurately establish the beginning and end of episodes, and to record clinical signs and symptoms in detail. Efforts should be made to keep study participants happy and interested. Bias is not a great concern in an observational study without treatment allocation of study participants. Continuous disease records may be needed, but depending on the budget, the surveillance period can be cut into blocks of, e.g., 6–8 weeks where surveillance is intense. This could allow capturing different seasons where different pathogens may circulate (dry cold season, wet season, hot season).
8	Context: demographic and health survey (DHS). The aim of the survey is to gain information on a range of topics, but the investigator also wishes to explore risk factors for diarrhoea (e.g. water, sanitation, socio-economic status) Study population: all ages Logistics: adequate budget, trained staff and large eligible population available	Outcome measure: LP Sampling frequency: one visit Recall period: 2–3 days Data type: point prevalence Comment: A short recall period is preferred to minimize recall error. A DHS usually aims to estimate prevalence as an absolute figure, not primarily to compare two groups, and therefore requires accurate data. Given the large sample size of most DHS surveys, loss of power due to a short recall period is normally not a big issue. Point prevalence data may often be easier to interpret and compare with, than period prevalence data, since diarrhoea definitions used in most DHS and epidemiological studies are based on disease experience during one day.

not necessarily require accounting for all of them in a later analysis. In cluster-randomized trials it is often sufficient to incorporate clustering at the level of the unit of randomization, i.e. the level of independence.[78] This is because lower-level correlation of disease should increase the between-cluster variation at higher levels, which increases the SE accordingly.[78]

Conclusion

When planning a study that measures diarrhoea, investigators must jointly consider the interdependent methodological points we have discussed in this article, which include recall periods, measures of disease occurrence (incidence vs prevalence), sampling frequencies and design effects. For example, the sampling frequency and the choice of the measure of disease occurrence can both influence the design effect. Conversely, the design effect can influence the choice of the sampling frequency or the recall period, because a strong design effect limits the study power gained from frequent sampling and long recall periods. Further, study settings differ from one another, especially in their logistics, which in turn has great implications for the study design. In some places it is difficult to recruit and train many field workers; in others it may be difficult to recruit study participants. As a consequence, it is difficult to develop universally applicable guidelines or a simple algorithm to identify the best way to measure diarrhoea in a specific study. In Table 5 we list examples of diarrhoea studies and suggest approaches to measure diarrhoea. None of our suggestions is meant to be absolute. As already suggested by Table 2, investigators must consider the

Box 1 Definitions

Diarrhoea day	Since diarrhoea symptoms occur intermittently, diarrhoea case definitions in epidemiology are usually based on the nature and frequency of symptoms experienced during one day (or 24 h). A diarrhoea case is therefore equivalent to a 'diarrhoea day'. For example, the WHO definition requires the occurrence '3 or more loose or liquid stools per day'.
Diarrhoea episode	One or more diarrhoea days occurring closely in time, presumably caused by a single agent or the interaction of multiple causative agents (e.g. as super-infection). Defining a diarrhoea episode requires deciding on how many diarrhoea-free days separate independent episodes. This decision is necessarily pragmatic especially in high-risk settings, as it is usually difficult to know whether diarrhoea days occurring closely in time belong to the same episode or not.
Diarrhoea incidence	The number of diarrhoea episodes per person-time (incidence density) or over a defined period of time (cumulative incidence).
Diarhoea point prevalence	The proportion of the population experiencing a diarrhoea day at the time of interest, e.g. the day of a surveillance visit or the day before.
Diarrhoea period prevalence	The proportion of the population experiencing at least 1 day with diarrhoea over a pre-defined time window (recall period) prior to a given point in time, e.g. a surveillance visit by the study team.
Recall period	The period of time over which the occurrence of diarrhoea is assessed at each contact with a study participant (e.g. phone call or home visit). To measure point prevalence, the recall period is treated as individual days (for example: 'on which of the last 7 days did you have diarrhoea?'). To measure period prevalence, the recall period is treated as a single time window (e.g. 'did you have diarrhoea at any day during the last 7 days?'). Thus, when using a 7-day recall period, a single surveillance visit yields 7-point prevalence datapoints, but only one period prevalence datapoint.
Longitudinal prevalence	The proportion of time an individual has diarrhoea. This can either be the proportion of days with diarrhoea (for point prevalence), or the proportion of time windows with at least 1 diarrhoea day (for period prevalence). For example, a person reporting diarrhoea on 10% of days has a longitudinal point prevalence of diarrhoea of 10%. A person reporting diarrhoea at any time in the last week, in 10% of weeks of surveillance has a longitudinal period prevalence of 10%. Note that while prevalence is a population measure of disease occurrence, LP is an individual measure. A person can have an LP of 10%, but not a prevalence of 10%. At population level, LP is best described by the mean and SD of individual LP values.

research question first, as many critical decisions depend on it. For example, incidence of diarrhoea (such as hospital admissions) could be the preferred measure in vaccine trials. Point or period prevalence measured at long intervals could be ideal for large environmental health interventions in high-risk populations where many villages and individuals need to be surveyed over a long time. A high-risk study population here means a setting where malnutrition and case fatality are a public health problem. Some studies (such as Demographic and Health Surveys) require obtaining precise absolute prevalence figures, for which collecting point prevalence data with a short recall period is most suitable.

We did not describe a number of important methodological challenges in diarrhoea trials that have been discussed elsewhere, such as the clinical definition of disease severity,[35,36,42,87,88] or objective proxy markers for diarrhoea in trials of interventions that cannot be blinded.[89] We also did not discuss recent advances in diagnostic tools for pathogen identification currently in use in some population-based studies.[90]

Diarrhoea continues to be a major global health problem, and there is an ongoing debate over identifying research priorities and the development of cheap and effective interventions, given the limited funding.[2,3,91–94] Whereas standard clinical trial procedures are often adequate to assess the effect of a vaccine or drug on diarrhoea in individuals, environmental interventions aiming at diarrhoea control are often much more complex, and more difficult to evaluate with randomized trials. Efficient methods to measure diarrhoea should allow more valid and generalizable results from research to be conducted with the same resources, especially in settings where resources are scarce.

KEY MESSAGES

- The design of epidemiological studies on diarrhoea requires specifying recall periods, sampling frequencies and outcome measures that are most suitable to answer the research question in a given setting.

- Sample size calculations often need to be done based on scarce data. This article outlines how the validity and logistical efficiency of diarrhoea studies can be improved by careful consideration of these factors.

References

[1] Black RE, Morris SS, Bryce J. Where and why are 10 million children dying every year? *Lancet* 2003;**361:** 2226–34.

[2] Kosek M, Lanata CF, Black RE *et al*. Directing diarrhoeal disease research towards disease-burden reduction. *J Health Popul Nutr* 2009;**27:**319–31.

[3] Schmidt WP. Setting priorities in diarrhoeal disease research: merits and pitfalls of expert opinion. *J Health Popul Nutr* 2009;**27:**313–15.

[4] Blum D, Feachem RG. Measuring the impact of water supply and sanitation investments on diarrhoeal diseases: problems of methodology. *Int J Epidemiol* 1983; **12:**357–65.

[5] Ejemot RI, Ehiri JE, Meremikwu MM, Critchley JA. Hand washing for preventing diarrhoea. *Cochrane Database Syst Rev* 2008;**23:**CD004265.

[6] Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *BMJ* 2007;**334:**782.

[7] Morris SS, Santos CA, Barreto ML *et al*. Measuring the burden of common morbidities: sampling disease experience versus continuous surveillance. *Am J Epidemiol* 1998; **147:**1087–92.

[8] Morris SS, Cousens SN, Kirkwood BR, Arthur P, Ross DA. Is prevalence of diarrhea a better predictor of subsequent mortality and weight gain than diarrhea incidence? *Am J Epidemiol* 1996;**144:**582–88.

[9] Hayes RJ, Moulton LH. *Cluster Randomised Trials*. Boca Raton: Chapman & Hall/CRC, 2009.

[10] Mausezahl D, Christen A, Pacheco GD *et al*. Solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia: a cluster-randomized, controlled trial. *PLoS Med* 2009;**6:**e1000125.

[11] Schmidt WP, Genser B, Chalabi Z. A simulation model for diarrhoea and other common recurrent infections: a tool for exploring epidemiological methods. *Epidemiol Infect* 2009;**137:**644–53.

[12] Vitamin A. supplementation in northern Ghana: effects on clinic attendances, hospital admissions, and child mortality. Ghana VAST Study Team. *Lancet* 1993;**342:**7–12.

[13] Melo MC, Taddei JA, Diniz-Santos DR, May DS, Carneiro NB, Silva LR. Incidence of diarrhea: poor parental recall ability. *Braz J Infect Dis* 2007;**11:**571–79.

[14] Wright JA, Gundry SW, Conroy R *et al*. Defining episodes of diarrhoea: results from a three-country study in Sub-Saharan Africa. *J Health Popul Nutr* 2006;**24:**8–16.

[15] Baqui AH, Black RE, Yunus M, Hoque AR, Chowdhury HR, Sack RB. Methodological issues in diarrhoeal diseases epidemiology: definition of diarrhoeal episodes. *Int J Epidemiol* 1991;**20:**1057–63.

[16] WHO. http://www.who.int/topics/diarrhoea/en/. 2009 (5 January 2011, date last accessed).

[17] Barreto ML, Santos LM, Assis AM *et al*. Effect of vitamin A supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in Brazil. *Lancet* 1994;**344:**228–31.

[18] Strina A, Cairncross S, Prado MS, Teles CA, Barreto ML. Childhood diarrhoea symptoms, management and duration: observations from a longitudinal community study. *Trans R Soc Trop Med Hyg* 2005;**99:**407–16.

[19] Alam N, Henry FJ, Rahaman MM. Reporting errors in one-week diarrhoea recall surveys: experience from a prospective study in rural Bangladesh. *Int J Epidemiol* 1989; **18:**697–700.

[20] Boerma JT, Black RE, Sommerfelt AE, Rutstein SO, Bicego GT. Accuracy and completeness of mothers' recall of diarrhoea occurrence in pre-school children in demographic and health surveys. *Int J Epidemiol* 1991; **20:**1073–80.

[21] Feikin DR, Audi A, Olack B *et al*. Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. *Int J Epidemiol* 2010;**39:**450–58.

[22] Ramakrishnan R, Venkatarao T, Koya PK, Kamaraj P. Influence of recall period on estimates of diarrhoea morbidity in infants in rural Tamilnadu. *Indian J Public Health* 1999;**43:**136–39.

23 Zafar SN, Luby SP, Mendoza C. Recall errors in a weekly survey of diarrhoea in Guatemala: determining the optimal length of recall. *Epidemiol Infect* 2010;**138:**264–69.

24 Lee G, Cama V, Gilman RH, Cabrera L, Saito M, Checkley W. Comparison of two types of epidemiological surveys aimed at collecting daily clinical symptoms in community-based longitudinal studies. *Ann Epidemiol* 2010;**20:**151–58.

25 Carabin H, Gyorkos TW, Soto JC, Joseph L, Payment P, Collet JP. Effectiveness of a training program in reducing infections in toddlers attending day care centers. *Epidemiology* 1999;**10:**219–27.

26 Colford JM Jr, Wade TJ, Sandhu SK *et al*. A randomized, controlled trial of in-home drinking water intervention to reduce gastrointestinal illness. *Am J Epidemiol* 2005;**161:** 472–82.

27 Colford JM Jr, Hilton JF, Wright CC *et al*. The Sonoma water evaluation trial: a randomized drinking water intervention trial to reduce gastrointestinal illness in older adults. *Am J Public Health* 2009;**99:**1988–95.

28 Genser B, Strina A, Teles CA, Prado MS, Barreto ML. Risk factors for childhood diarrhea incidence: dynamic analysis of a longitudinal study. *Epidemiology* 2006;**17:** 658–67.

29 Haggerty PA, Muladi K, Kirkwood BR, Ashworth A, Manunebo M. Community-based hygiene education to reduce diarrhoeal disease in rural Zaire: impact of the intervention on diarrhoeal morbidity. *Int J Epidemiol* 1994;**23:**1050–59.

30 Zwane AP, Zinman J, Van Dusen E *et al*. Being surveyed can change later behavior and related parameter estimates. *Proc Natl Acad Sci USA* 2011;**108:**1821–26.

31 Manesh AO, Sheldon TA, Pickett KE, Carr-Hill R. Accuracy of child morbidity data in demographic and health surveys. *Int J Epidemiol* 2008;**37:**194–200.

32 Schmidt WP, Cairncross S. Household water treatment in poor populations: is there enough evidence for scaling up now? *Environ Sci Technol* 2009;**43:**986–92.

33 Byass P, Hanlon PW. Daily morbidity records: recall and reliability. *Int J Epidemiol* 1994;**23:**757–63.

34 Schmidt WP, Genser B, Barreto ML, Clasen T, Luby SP, Cairncross S *et al*. Sampling strategies to measure the prevalence of common recurrent infections in longitudinal studies. *Emerg Themes Epidemiol* 2010;**7:**5.

35 Baqui AH, Black RE, Sack RB, Yunus MD, Siddique AK, Chowdhury HR. Epidemiological and clinical characteristics of acute and persistent diarrhoea in rural Bangladeshi children. *Acta Paediatr Suppl* 1992;**381:**15–21.

36 Victora CG, Huttly SR, Fuchs SC *et al*. International differences in clinical patterns of diarrhoeal deaths: a comparison of children from Brazil, Senegal, Bangladesh, and India. *J Diarrhoeal Dis Res* 1993;**11:**25–29.

37 Morris SS, Cousens SN, Lanata CF, Kirkwood BR. Diarrhoea–defining the episode. *Int J Epidemiol* 1994;**23:** 617–23.

38 Payment P, Siemiatycki J, Richardson L, Renaud G, Franco E, Prevost M. A prospective epidemiological study of gastrointestinal health effects due to the consumption of drinking water. *Int J Environ Health Res* 1997;**7:**5–31.

39 Pickering H, Hayes RJ, Tomkins AM, Carson D, Dunn DT. Alternative measures of diarrhoeal morbidity and their association with social and environmental factors in urban children in The Gambia. *Trans R Soc Trop Med Hyg* 1987;**81:**853–509.

40 Luby SP, Agboatwalla M, Feikin DR *et al*. Effect of handwashing on child health: a randomised controlled trial. *Lancet* 2005;**366:**225–33.

41 Sethi D, Wheeler J, Rodrigues LC, Fox S, Roderick P. Investigation of under-ascertainment in epidemiological studies based in general practice. *Int J Epidemiol* 1999; **28:**106–12.

42 Armah GE, Sow SO, Breiman RF *et al*. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in subSaharan Africa: a randomised, double-blind, placebocontrolled trial. *Lancet* 2010;**376:**606–14.

43 Torres AM, Peterson KE, de Souza AC, Orav EJ, Hughes M, Chen LC. Association of diarrhoea and upper respiratory infections with weight and height gains in Bangladeshi children aged 5 to 11 years. *Bull World Health Organ* 2000;**78:**1316–23.

44 Reller ME, Mendoza CE, Lopez MB *et al*. A randomized controlled trial of household-based flocculant-disinfectant drinking water treatment for diarrhea prevention in rural Guatemala. *Am J Trop Med Hyg* 2003;**69:**411–19.

45 Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *Int J Epidemiol* 1999;**28:**319–26.

46 Schmidt WP, Genser B, Luby SP, Chalabi Z. Estimating the effect of recurrent infectious diseases on nutritional status: sampling frequency, sample size and bias. *J Health Popul Nutr* 2011;**29:**317–26.

47 Black RE, Brown KH, Becker S. Effects of diarrhea associated with specific enteropathogens on the growth of children in rural Bangladesh. *Pediatrics* 1984;**73:**799–805.

48 Checkley W, Buckley G, Gilman RH *et al*. Multi-country analysis of the effects of diarrhoea on childhood stunting. *Int J Epidemiol* 2008;**37:**816–30.

49 Cole TJ, Parkin JM. Infection and its effect on the growth of young children: a comparison of the Gambia and Uganda. *Trans R Soc Trop Med Hyg* 1977;**71:**196–98.

50 Condon-Paoloni D, Cravioto J, Johnston FE, De Licardie ER, Scholl TO. Morbidity and growth of infants and young children in a rural Mexican village. *Am J Public Health* 1977;**67:**651–56.

51 Lutter CK, Mora JO, Habicht JP *et al*. Nutritional supplementation: effects on child stunting because of diarrhea. *Am J Clin Nutr* 1989;**50:**1–8.

52 Rowland MG, Rowland SG, Cole TJ. Impact of infection on the growth of children from 0 to 2 years in an urban West African community. *Am J Clin Nutr* 1988;**47:**134–38.

53 Fenn B, Morris SS, Black RE. Comorbidity in childhood in northern Ghana: magnitude, associated factors, and impact on mortality. *Int J Epidemiol* 2005;**34:**368–75.

54 Schmidt WP, Cairncross S, Barreto ML, Clasen T, Genser B. Recent diarrhoeal illness and risk of lower respiratory infections in children under the age of 5 years. *Int J Epidemiol* 2009;**38:**766–72.

55 Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ* 2010;**340:** c2289.

56 Schmidt WP, Luby SP, Genser B, Barreto ML, Clasen T. Estimating the longitudinal prevalence of diarrhea and other episodic diseases: continuous versus intermittent surveillance. *Epidemiology* 2007;**18:**537–43.

57 Gail MH, Mark SD, Carroll RJ, Green SB, Pee D. On design considerations and randomization-based inference for community intervention trials. *Stat Med* 1996;**15:** 1069–92.

58 Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002;**21:**2917–30.

59 Trotta A. Statistical analysis and lessons learned from a hygiene intervention carried out 20 years ago in Lima, Peru. MSc Thesis. London School of Hygiene and Tropical Medicine, 2010.

60 Luby SP, Agboatwalla M, Hoekstra RM. The variability of childhood diarrhea in karachi, pakistan, 2002-2006. *Am J Trop Med Hyg* 2011;**84:**870–77.

61 Frost C, Kenward MG, Fox NC. Optimizing the design of clinical trials where the outcome is a rate. Can estimating a baseline rate in a run-in period increase efficiency? *Stat Med* 2008;**27:**3717–31.

62 Hayes RJ, Alexander N, Bennett S, Cousens SN. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Stat Methods Med Res* 2000;**9:**95–116.

63 Bruhn M, McKenzie D. In pursuit of balance: randomization in practice in development field experiments. *Policy Research Working Paper 4752.* The World Bank, Washington, 2008.

64 Freedman DA. On regression adjustments to experimental data. *Adv Appl Math* 2008;**40:**180–93.

65 Clasen TF, Brown J, Collin S, Suntura O, Cairncross S. Reducing diarrhea through the use of household-based ceramic water filters: a randomized, controlled trial in rural Bolivia. *Am J Trop Med Hyg* 2004;**70:**651–57.

66 Boisson S, Kiyombo M, Sthreshley L, Tumba S, Makambo J, Clasen T. Field assessment of a novel household-based water filtration device: a randomised, placebo-controlled trial in the Democratic Republic of Congo. *PLoS One* 2010;**5:**e12613.

67 Colford JM Jr, Wade TJ, Sandhu SK *et al.* A randomized, controlled trial of in-home drinking water intervention to reduce gastrointestinal illness. *Am J Epidemiol* 2005;**161:** 472–82.

68 Clasen T, Garcia PG, Boisson S, Collin S. Household-based ceramic water filters for the prevention of diarrhea: a randomized, controlled trial of a pilot program in Colombia. *Am J Trop Med Hyg* 2005;**73:**790–95.

69 Boisson S, Schmidt WP, Berhanu T, Gezahegn H, Clasen T. Randomized controlled trial in rural Ethiopia to assess a portable water treatment device. *Environ Sci Technol* 2009;**43:**5934–39.

70 Tiwari SS, Schmidt WP, Darby J, Kariuki ZG, Jenkins MW. Intermittent slow sand filtration for preventing diarrhoea among children in Kenyan households using unimproved water sources: randomized controlled trial. *Trop Med Int Health* 2009;**14:**1374–82.

71 Boisson S, Schmidt WP, Berhanu T, Gezahegn H, Clasen T. Randomized controlled trial in rural Ethiopia to assess a portable water treatment device. *Environ Sci Technol* 2009;**43:**5934–39.

72 Ross DA, Huttly SR, Dollimore N, Binka FN. Measurement of the frequency and severity of childhood acute respiratory infections through household surveys in northern Ghana. *Int J Epidemiol* 1994;**23:**608–16.

73 Kirkwood B, Sterne J. *Medical Statistics.* Malden: Blackwell Science, 2003.

74 Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol* 2006; **35:**1292–300.

75 Katz J, Carey VJ, Zeger SL, Sommer A. Estimation of design effects and diarrhea clustering within households and villages. *Am J Epidemiol* 1993;**138:**994–1006.

76 Ridout MS, Demetrio CG, Firth D. Estimating intraclass correlation for binary data. *Biometrics* 1999;**55:**137–48.

77 Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ* 2001;**322:**355–57.

78 Hayes RJ, Moulton LH. *Repeated Measures During Follow-up. Cluster Randomised Trials.* Boca Raton: Chapman & Hall/ CRC, 2009.

79 Reller ME, Mendoza CE, Lopez MB *et al.* A randomized controlled trial of household-based flocculant-disinfectant drinking water treatment for diarrhea prevention in rural Guatemala. *Am J Trop Med Hyg* 2003;**69:**411–19.

80 Arnold BF, Khush RS, Ramaswamy P *et al.* Causal inference methods to study nonrandomized, preexisting development interventions. *Proc Natl Acad Sci USA* 2010;**107:** 22605–10.

81 Luby SP, Halder AK, Tronchet C, Akhter S, Bhuiya A, Johnston RB. Household characteristics associated with handwashing with soap in rural Bangladesh. *Am J Trop Med Hyg* 2009;**81:**882–87.

82 Luby SP, Agboatwalla M, Feikin DR *et al.* Effect of handwashing on child health: a randomised controlled trial. *Lancet* 2005;**366:**225–33.

83 van der Hoek W, Feenstra SG, Konradsen F. Availability of irrigation water for domestic use in Pakistan: its impact on prevalence of diarrhoea and nutritional status of children. *J Health Popul Nutr* 2002;**20:**77–84.

84 Luby SP, Agboatwalla M, Painter J *et al.* Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. *Trop Med Int Health* 2006;**11:**479–89.

85 Clasen TF, Brown J, Collin SM. Preventing diarrhoea with household ceramic water filters: assessment of a pilot project in Bolivia. *Int J Environ Health Res* 2006;**16:**231–39.

86 Barreto ML, Genser B, Strina A *et al.* Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast Brazil: assessment by two cohort studies. *Lancet* 2007;**370:**1622–28.

87 Johnston BC, Shamseer L, da Costa BR, Tsuyuki RT, Vohra S. Measurement issues in trials of pediatric acute diarrheal diseases: a systematic review. *Pediatrics* 2010; **126:**e222–31.

88 Lima AA, Moore SR, Barboza MS Jr *et al.* Persistent diarrhea signals a critical period of increased diarrhea burdens and nutritional shortfalls: a prospective cohort study among children in northeastern Brazil. *J Infect Dis* 2000; **181:**1643–51.

89 Schmidt WP, Boisson S, Genser B *et al.* Weight-for-age z-score as a proxy marker for diarrhoea in epidemiological studies. *J Epidemiol Community Health* 2009;**64:**1074–79.

90 GEMS. *Global Enteric Multi-Center Study.* http://medschool .umaryland.edu/GEMS/default.asp. 2010.

91 Bartram J, Lewis K, Lenton R, Wright A. Focusing on improved water and sanitation for health. *Lancet* 2005; **365:**810–12.

92 Cairncross S. Water supply and sanitation: an agenda for research. *J Trop Med Hyg* 1989;**92:**301–14.

93 Fontaine O, Kosek M, Bhatnagar S *et al.* Setting research priorities to reduce global mortality from childhood diarrhoea by 2015. *PLoS Med* 2009;**6:**e41.

94 Rudan I, El Arifeen S, Black RE, Campbell H. Childhood pneumonia and diarrhoea: setting our priorities right. *Lancet Infect Dis* 2007;**7:**56–61.