NBER WORKING PAPER SERIES

WATER TREATMENT AND CHILD MORTALITY: A META-ANALYSIS AND COST-EFFECTIVENESS ANALYSIS

Michael Kremer
Stephen P. Luby
Ricardo Maertens
Brandon Tan
Witold Więcek

Working Paper 30835
http://www.nber.org/papers/w30835

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2023

This research is supported by Dioraphte, Sint Antonius Stichting Foundation (Project Number 1808), and GiveWell. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Water Treatment And Child Mortality: A Meta-Analysis And Cost-effectiveness Analysis
Michael Kremer, Stephen P. Luby, Ricardo Maertens, Brandon Tan, and Witold Więcek
NBER Working Paper No. 30835
January 2023
JEL No. I1

ABSTRACT

Randomized controlled trials (RCTs) of water treatment are typically powered to detect effects on caregiver-reported diarrhea but not child mortality, as detecting mortality effects requires prohibitively large sample sizes. Consequently, water treatment is seldom included in lists of cost-effective, evidence-backed child health interventions which are prioritized in health funding decisions. To increase statistical power, we conducted a systematic review and meta-analysis. We replicated search and selection criteria from previous meta-analyses of RCTs aimed at improving water quality to prevent diarrhea in low- or middle-income countries which included children under 5 years old. We identified 52 RCTs and then obtained child mortality data from each study for which these data were collected and available, contacting authors of the study where necessary; this resulted in 15 studies.Frequentist and Bayesian methods were used to estimate the effect of water treatment on child mortality among included studies. We estimated a mean cross-study reduction in the odds of all-cause under-5 mortality of about 30% (Peto odds ratio, OR, 0.72; 95% CI 0.55 to 0.92; Bayes OR 0.70; 95% CrI 0.49 to 0.93). The results were qualitatively similar under alternative modeling and data inclusion choices. Taking into account heterogeneity across studies, the expected reduction in a new implementation is 25%. We used the results to examine the cost-effectiveness of investing in water treatment for point-of-collection chlorine dispensers or a large-scale program providing coupons for free chlorine solution. We estimate a cost per expected DALY averted due to water treatment of around USD 40 for both, accounting for delivery costs. This is approximately 45 times lower than the widely used threshold of 1x GDP per capita per DALY averted.

Michael Kremer
University of Chicago
Department of Economics
1126 E. 59th St.
Chicago, IL 60637
and NBER
kremermr@uchicago.edu

Stephen P. Luby
Division of Infectious Diseases and
Geographic Medicine
Stanford University
Palo Alto, CA 94034
sluby@stanford.edu

Ricardo Maertens
Amazon
515 Westlake Ave N
Seattle, WA 98109
ricardo.maertens@gmail.com

Brandon Tan
International Monetary Fund
btan2@imf.org

Witold Więcek,
Development Innovation
Lab University of Chicago
witold.wiecek@gmail.com

Introduction

Each year over two billion people consume drinking water contaminated with feces (1) and over 1.5 million people die from diarrheal diseases (2). Climate change and aquifer depletion threaten existing sources of clean water (3). Yet, even relatively basic and inexpensive measures to contain disease spread from fecally-contaminated water remain unimplemented in large parts of the world. Chlorination, for example, has been found to be effective in reducing the concentration of diarrheal pathogens like *E.coli* in controlled laboratory settings (4–7) and in reducing caregiver-reporter diarrhea (8, 9). However, 71% of the population in low-income countries and 40% in lower-middle-income countries do not have access to safely managed drinking water facilities (10).

In addition to municipal water systems, a variety of systems can effectively deliver chlorinated water at low cost. These include point-of-use (which provide people with the means to treat water within their household), point-of-collection (e.g. dispensers for dilute chlorine solution placed near water source, already used at scale in several countries), and in-line devices, which automatically chlorinate water.

However, water treatment is often not included in lists of cost-effective, evidence-backed child health interventions which are recommended for prioritization in health funding decisions, and health funds are typically not used to cover the cost of water treatment at scale. This may in part be due to a lack of RCT evidence on the effect of water treatment on child mortality. Because child death is a rare event, conducting adequately powered randomized controlled trials (RCTs) to measure the impact of water treatment on child mortality requires very large sample sizes and correspondingly large costs. Therefore, RCTs measuring the impact of water treatment are typically powered to detect effects on the (higher incidence) intermediate outcome of caregiver-reported child diarrhea, rather than child mortality. However, caregiver reports of child diarrhea may be subject to reporting bias (11, 12). Some have therefore recommended the need for studies which are either blinded or include as a primary outcome an objective outcome such as mortality (11).

To increase statistical power to detect child mortality impacts, we conducted a literature search aimed at combining existing RCT evidence on mortality with new evidence we obtained from authors of studies reporting other outcomes. We then used a meta-analysis to estimate the impact and cost-effectiveness of water quality interventions on child mortality.

Methods

This systematic review was registered within the American Economic Association (AEA) under the registration number AEARCTR-0005977, and can be accessed here https://www.socialscienceregistry.org/trials/5977.

We followed PRISMA 2020 guidelines (11); These are provided in the supplementary material.

Search strategy and selection criteria

We first reviewed all studies identified by previous meta-analyses examining the impact of water quality interventions on diarrhea (8, 9). Next, the search procedure and selection criteria followed by a previous meta-analysis (9) were replicated for the period not covered by the previous studies, from February 2016 to May 2020 (date of last search was April 20, 2020). The selection criteria were updated to allow for manuscripts published during the eligibility period that were updated after the period concluded. As detailed in Table S1, the search included Pubmed, Embase, Scopus and Cochrane Library using both keywords and MeSH terms to identify all studies of interventions to improve water quality. Additional papers for review were also included based on reference sections of all papers, as well as recommendations from experts.

Included studies were restricted to RCTs of interventions to improve water quality (in the microbiological sense) in low- or middle-income countries (according to the World Bank classification) which included children less than five years of age. The choice of including only RCTs was made to focus on studies that can estimate causal impacts with minimal methodological assumptions.

Data extraction and quality assessment

Two reviewers independently performed study title and abstract screening, filtering of studies in accordance with the inclusion criteria, data extraction, and quality assessment. Both author-provided and publicly available individual-level data on child (<5 years) mortality were used for the study. Data were collected through surveys, and all available data on mortality were considered. We also extracted (from the appendices available online) summary data on all studies in (9) to compare key characteristics between studies included in this meta-analysis and excluded studies (Materials and methods, section 1).

Two review authors independently assessed the risk of bias using the same Newcastle-Ottawa scale (13) as in (9) for each study included in this review on the following dimensions: sample selection, responses (blinding versus no-blinding), treatment allocation, follow-up (attrition),

degree of treatment exposure, compliance, the dimension of the assessment, and measurement of the outcome.

Data analysis

For the meta-analysis model, we used an odds ratio (OR) outcome.[1] Since death is rare (typically 1-2% annualized risk), we chose models that are appropriate for estimating treatment effects for rare events: a frequentist Peto odds model and a Bayesian logistic model with non-informative priors (14, 15). In both cases we chose a random effects model as our main specification due to heterogeneity in types of water quality interventions and study settings. We decided to estimate average effects for all studies, and the sub-sample of studies that include water chlorination. The models are described in Materials and Methods, section 2.

We also examined potential heterogeneity in treatment effects using Peto OR meta-regression models, fitting one variable at a time for the following: baseline prevalence of diarrhea, level of compliance, unit of randomization (cluster vs household), diarrhea effect estimates, and year of implementation.

We performed sensitivity analyses to understand how researcher choices on data inclusion and modeling assumptions could impact the meta-analysis estimates. Additional details explaining these analyses case-by-case, as well as detailed results, are given in Materials and Methods, section 3.

Additionally, posterior predictive distributions from Bayesian models and an alternative choice of priors were used for cost-effectiveness calculations and are described below.

We examined potential publication bias through inspection of funnel plots and the use of Egger's and Andrews and Kasy's tests (16). We made two checks: one for mortality outcome, using all papers meta-analyzed in this paper, and another for diarrhea outcome, using more studies (that measured diarrhea but not mortality) based on (9). We also checked for association between availability of mortality data and effect on diarrhea and estimated a publication bias-adjusted OR estimate following Andrews and Kasy.

All statistical analyses and visualizations were performed with R, version 4.1.

[1] From a biological viewpoint, the choice of OR seems more appropriate than modeling risk difference (RD) and, when events are rare (as is the case with child mortality), the odds ratio approaches the risk ratio (RR), and therefore we did not compare OR and RR models. Moreover, the RD model is not appropriate when differences in mortality risk across studies is very large, as is the case in our sample, due to heterogeneity in length of follow-up, which we will discuss later. Therefore we included the RD model as a sensitivity analysis only.

Cost effectiveness

Typically, policymakers' decisions about health interventions take place in two stages, starting with a regulatory decision of whether to approve a new intervention, followed by cost-effectiveness analysis to inform investment. We do not consider the first stage since water treatment has long been widely used and is widely accepted to be safe. We use the results from the meta-analysis to conduct a cost-effectiveness analysis. We consider the problem of a (risk-neutral) social planner investing on behalf of households, all given the same weight, to reduce the incidence of child death. Such a social planner will invest in water treatment if and only if the cost per expected life saved is below some threshold. We discuss the choice of such a threshold below.

We calculate the expected reductions in deaths and DALYs due to implementation of water treatment in a new setting based on the Bayesian posterior predictive distribution, which takes into account uncertainty due to heterogeneity across studies. To calculate benefits per dollar invested we then divide estimated costs by the expected reductions.[2]

We examine two water treatment approaches for cost-effectiveness analysis. First, point-of-collection dispensers of dilute chlorine solution, for which we have access to cost data from a large-scale implementation. Second, a hypothetical global program delivering free water-treatment through a coupon program, which could potentially be applied in a wide range of settings through existing health systems. Since it has not been implemented on a large scale, we consider rough cost estimates. We provide more details on the calculation in Results. Our analysis demonstrates cost effective approaches exist but it is not intended to make the case that these approaches are more cost effective than alternatives. Different settings may require different approaches and some, such as municipal water treatment programs, may generate benefits on a range of dimensions.

We compare cost-effectiveness results with two commonly used metrics. First, thresholds of 1x and 3x GDP per capita, which might be relevant if the program is funded through domestic taxes. These thresholds were first proposed by the Commission on Macroeconomics and Health and used in earlier editions of the World Health Organization's "Choosing Interventions that are Cost-Effective" (WHO-CHOICE) (henceforth 1xGDP threshold).

Second, we use cost-effectiveness brackets (e.g. $10 - $100 per DALY averted), used by the most recent edition of WHO-CHOICE publication and the World Bank's DCP-3 (17). Since this calculation does not depend on GDP, it may be relevant if the program is financed by an external actor seeking to maximize health benefits within a fixed budget.

[2] In this calculation we do not put extra weight on studies of more similar interventions, but, as we will show below, for a subgroup of chlorine studies the effects are larger in absolute magnitude.

Results

Systematic review

Figure 1 illustrates the search and the selection process. The search strategy identified 1485 studies: 1412 studies through databases and 73 studies included in (8) and (9). We screened these titles and abstracts to obtain a sample of 82 studies for full-text review. 52 studies matched the inclusion criteria and we requested child mortality data from the authors of each study. 25 authors reported that they did not collect mortality data or that the data was no longer available. The author of one study died and the authors of nine studies did not reply. Excluded studies are given in Table S2. The sample of 17 studies with mortality data is summarized in Table 1. Two studies were then excluded from the main analysis due to contamination in the control group but we conduct a sensitivity analysis with these studies included (see Materials and Methods, section 3). Raw input data for meta analysis are in Table S3.[3]

Publication bias.

Neither Egger's nor Andrews and Kasy's tests provided evidence of publication bias on diarrhea or mortality outcomes. We also did not find evidence of the magnitude of measured effect on diarrhea being associated with availability of mortality outcomes. Since the power of these tests for mortality outcome may be limited when applied to our sample of 15 studies, we also conducted post hoc simulations. We find that even if as many as 15 unpublished short studies with null effects (i.e. assuming mortality risk in both arms of 0.4%, which is one quarter of annual mortality in our data) were added to our dataset, the meta-analytic estimate of OR would still be significant. We provide more details in Materials and Methods, section 5.

Risk of bias assessment

Among the included studies, we assessed the bias attributed to the selection of studies as low. First, all included studies are randomized controlled trials. Second, although in only one out of the fifteen studies the participants were blinded, reporting bias or experimenter effects are unlikely (see Supplementary Material: Risk of Bias). The mortality status of a child who was alive at baseline can be easily verified and is far less likely to be subject to reporting bias than caregiver-reported diarrhea outcomes based on recall.

Characteristics of included studies

[3] In the sample of the included studies, six studies had Steve Luby as an author, two had Michael Kremer as an author, and one of these had Ricardo Maertens and Brandon Tan as authors. None of the authors have any financial interest in these results.

The studies included 25,300 participants. Twelve examined water chlorination, two examined water filtration, and one examined spring protection. In aggregate, 170 deaths occurred among 11,701 children in treatment arms (1.5%); in the control arms, 339 among 13,599 (2.5%). Four studies had no deaths in control and/or treatment arms. The annual risk of mortality in the pooled control group was about 1.7%. We found studies to be representative of diarrhea prevalence in LMICs (see Figure 2) and found no significant differences to the larger set of RCTs which measured diarrhea. We provide detailed characteristics of 15 included studies and details of the comparison with other RCTs in Materials and Methods, section 1.

Meta-analysis

In the full set of 15 studies, we estimated a significant average reduction in odds of all-cause child mortality of 28% (Peto OR 0.72; CI 95% 0.55, 0.92) or 30% (Bayes OR 0.70; CrI 95% 0.49, 0.93), depending on the model (see Figure 3).[4] OR confidence/credibility intervals for individual studies were typically wide, as one would expect in modeling rare event data. In fact, in only three studies the Peto or Bayesian OR 95% intervals were below 1. Restricting the analysis to studies including chlorination, the reduction was 31% (Peto OR 0.69; CI 95% 0.47, 1.01; Bayes OR 0.69; CrI 95% 0.38, 1.03).

Heterogeneity was not precisely determined; between-study SD (difference in true study means), measured on the log(OR) scale had a mean of 0.29 for Bayesian model (CrI 95% 0.01, 0.78) and 0.24 for the Peto model (CI 95% 0.00, 1.01). Relative to the mean that is 78% and 71% respectively. The I-squared (% of variation due to underlying variation in true ORs) was 29% in the Peto model (CI 95% 0%, 62%; p-value for being non-zero = 0.14) and 6% for Bayesian model (CrI 95% 0%, 26%). A leave-one-study-out cross-validation procedure for the Bayesian model suggested similar out-of-sample performance for fixed-effects and random-effects models.

Expected reduction in mortality odds in a new implementation, which is used by cost-effectiveness calculations and based on Bayesian posterior predictive distribution, was 25% (Bayes OR of 0.75; 95% CrI 0.29, 1.50).[5] Using the distribution of expected effect we also constructed a plot of predicted absolute mortality rates among treated for a setting with a specific control group mortality rate (Fig S10).

---

[4] Since at low event rates ORs are approximately equal to RRs, assuming under-5 mortality in settings without access to clean water is 5% (see Table S8 for details), our Bayesian OR estimate implies mean risk reduction of 29% (Bayes RR of 0.71; 95% CrI 0.50, 0.92).

[5]The effect in new implementation is different to the mean across 15 studies due to estimation being done on log ORs, which is approximately normal, the heterogeneity has an impact on both the width of the interval (which combines uncertainty in the mean with between-study variation) but also on the mean, and consequently on the expected reductions in deaths. As we take variation between studies into account, variation increases, meaning the distribution gets wider, and the asymmetry means this decreases the expected effect size. A policymaker with a strong prior that water treatment is safe could view this calculation as conservative.

We provide results for all sensitivity analyses of data and model choices that we performed in Materials and Methods, section 3. The estimates remained qualitatively similar to our main estimate, with mean OR estimates ranging from 0.64 to 0.80 for various data choices and 0.74 to 0.76 for alternative model choices.

At the beginning of this study, five RCTs were identified which reported mortality outcomes as part of their analysis (18–22). The estimates from restricting the analysis to only the five studies which published mortality outcomes were similar in magnitude to that of the full sample though insignificant at the 95% confidence level (Peto OR 0.67; CI 95% 0.41, 1.11; Bayes OR 0.74; CrI 95% 0.28, 1.50).

We conducted a simple post-hoc simulation approach (Materials and Methods, section 6) to determine whether there is sufficient power using our sample of 15 studies to find significant impacts of covariates on treatment effect. We conclude that the power to detect these relationships is low. However, the univariate meta-regression models did not find significant differences on any of the examined variables (see Figures S5-S9). For the year of implementation (Fig S9), we found an increase in log(OR) of 0.055 per year (SE = 0.029, p-value = 0.06). Given that we have not corrected for multiple hypothesis tests, finding one effect significant at the 6% level out of 5 tests is not strong evidence of heterogeneous effects, especially considering that the assumption of a linear relationship between year and logarithm of OR seems unlikely to be correct. However, more data should be collected to explore whether variables that might have changed over time (such as the overall child mortality rate, the rollout of rotavirus vaccines, or the adoption of oral rehydration therapy) influence the treatment effect.

Cost effectiveness

The cost-effectiveness calculations, based on the expected 25% reduction in the odds of mortality in a new implementation are shown in Table 2 and more details are given in Materials and Methods, section 4 and Table S8.

Point-of-access chlorine dispensers

Cost data was provided by the NGO Evidence Action, which has programs in Kenya, Uganda and Malawi. We focus on Kenya, where Evidence Action operated approximately 18,400 point-of-collection chlorine dispensers as of 2020, providing roughly 2.19 million people with access to safe water. Given an adoption rate of 52%, approximately 1.14 million people are estimated to treat their water (23). We calculated cost per DALY averted due to water treatment of USD 39 (Table 2, Column 1), far lower than Kenyan GDP per capita (about USD 1,878 in 2020), the relevant 1xGDP “highly cost-effective” threshold.

Coupons for free dilute chlorine solution

While dispensers achieve relatively high usage rates, they are only suited to certain contexts. For example, if few households share each water source, dispensers may be more costly per person reached. Evidence Action restricted the placement of dispensers to water sources used by a minimum number of households. Programs providing coupons for free dilute chlorine solution to families with young children may have wider applicability. Because such programs have so far only been conducted at a modest scale, it is difficult to assess costs for large-scale programs. A 150-milliliter bottle of dilute chlorine solution sufficient for treating one household-month of water costs USD 0.31. If for every two households targeted the program covers an additional untargeted household which already has clean water, and if the administrative costs of running a coupon program were as large as the retail price of the chlorine solution, the cost of a scaled-up program would still only be USD 2,974 per death of a child under 5 averted – or USD 38 per DALY averted (Table 2, column 2).

Lists of highly cost-effective health interventions

As discussed earlier, several multilateral organizations produce lists of the most cost-effective, evidence-backed health approaches and recommend that governments prioritize these approaches for investment. Often these lists exclude or de-prioritize water treatment, presumably in part due to insufficient evidence on child mortality impacts.

For example, the WHO-CHOICE's latest publication for maternal, newborn and child health (17) lists 39 interventions for sub-Saharan Africa with a cost of less than $100 per DALY averted, including childhood vaccination, nutritional supplementation, and malaria treatment. It lists a further eight interventions with a cost of $100 - $1,000 per healthy life year. It does not include water treatment.

The World Bank's Disease Control Priorities 3 (DCP-3) does not include water treatment in its highest priority package for Essential Universal Health Coverage. It also categorizes water treatment under "Injury and Environmental Health", de-prioritizing it in the volume on reproductive, maternal, newborn and child health (24, 25).

Discussion

Individual randomized control trials studying the impact of water treatment are typically not powered for mortality, and lack of RCT evidence on mortality has historically constrained the use of health funds for water treatment. Aggregating data from 15 studies, we estimate that water treatment reduced the odds of all-cause child mortality by about 30% on average. Taking into account heterogeneity across studies, the expected reduction in the odds of all-cause child mortality in a new implementation is 25%.

We also estimate the expected number of lives saved from water treatment relative to the cost. Our analysis suggests that water treatment is one of the most cost-effective health approaches available, and that policymakers aiming to improve child health should consider water treatment. It also suggests that even small effects on mortality would meet the conventional cost-effectiveness thresholds. For example, repeating the calculation for chlorine dispensers, we find that the threshold of 1x GDP is reached at 0.6% reduction in odds of under-5 mortality.

The WHO estimates that 2.2 billion people around the world do not have access to safely managed drinking water services (1) similar in magnitude to the global estimates from other studies. To illustrate the potential magnitude of the benefits of water treatment, in Table S7 we present a back-of-envelope calculation, which suggests that a global program that gives coupons for free water treatment solution to families with under-5 children would save around 372,000 under-five lives at a cost of approximately USD 1.1 billion each year.

As we have shown, lists of child-health interventions often do not include water treatment. Moreover, government officials responsible for water are typically based in water or public works ministries rather than health ministries and often have other priorities, such as irrigation. However, our analysis suggests that water treatment may have a very large impact on child health. It can be delivered through the health system and, unlike water access more generally, it has limited benefits beyond health benefits. Therefore it makes sense to consider it alongside other child-health interventions.

Even though our cost-effectiveness analysis suggests high expected value from water treatment, substantial uncertainty remains both about the expected effect and about how it may vary across contexts. Standard decision theory suggests that policymakers should allocate budgets so as to maximize expected benefits given current information. At the same time, additional information could overturn the conclusion that water treatment is cost-effective, or yield better information on when it's likely to be effective, or what types of treatment are more likely to be effective. That would either require larger sample sizes, multiple studies, or a combination of the two, to the extent that it's possible there is a time-trend in the impact of water treatment.

Combining information and comparison with other sources of evidence

In the future, decision makers could combine RCT evidence with other sources of evidence on water interventions, for example from a review of scientific mechanisms and the quasi-experimental literature. One way to incorporate this information is as priors in a Bayesian meta-analysis model. We present an illustrative example of how our cost-effectiveness analysis could be conducted with informative priors in Materials and Methods, section 4.

Policymakers deciding on the design and targeting of water treatment programs could also make use of Bayesian priors to incorporate context-specific information about likely drivers of

heterogeneity in treatment effects. Data on water treatment are fairly inexpensive to collect, since water can be readily tested for chlorination.

The point estimate of the mortality effect obtained in this meta-analysis is much larger than the point estimate predicted by a simple model in which diarrheal deaths are taken from the central estimate of the Global Burden of Disease (GBD) project (2), the effect of water treatment on diarrhea is taken from the central estimate in an earlier meta-analysis (8), and mortality is assumed to be linear in diarrhea cases, so that reductions in diarrhea deaths are proportional to reductions in diarrheal cases. However, this model is unlikely to be an accurate model of the relationship between water treatment and child mortality, for reasons discussed further in Materials and Methods, section 7.

Limitations

We included all studies for which authors reported that mortality data were collected and remained available, but there could be publication bias if authors were more likely to collect, preserve, and report in situations in which effect sizes were likely to be larger. We find no statistically significant evidence of publication bias (for diarrhea and for death outcomes, assessed separately), but these tests have limited power. We attempt to address this through simulations (Materials and Methods, section 5).

While including short studies does not have a major impact on this analysis (Materials and Methods, section 3), we assume that odds in each included study can be interpreted as odds of under-5 mortality. This would be an acceptable choice if treatment effect OR's are homogeneous with age, which is something we do not examine in the present analysis of aggregate data.[6] Survival models could be used to address this in the future by making use of individual-level data on age, which are available for a subset of studies (Materials and Methods, section 1).

This meta-analysis is also subject to the more general limitations of meta-analyses. The estimate of the mean effect we obtain in this study is specific to the sample of included studies, and uncertainty when generalizing to new contexts is not fully captured by the uncertainty in the mean effect. However, we incorporate heterogeneity into our cost-effectiveness assessment by using predicted mean effect, which has a higher OR (smaller effect) than the mean within the 15 studies.

Several factors could influence the effect of water treatment on child mortality: including the level of adherence, counterfactual levels of water treatment, local disease burden from diarrhea

[6] Even under a correctly specified model and unbiased estimate, treating ORs from short studies as ORs over 5 years will slightly bias the estimate in direction of no effect, due to compounding of risks. However, the bias this introduces is small, e.g. ORs will differ less than 0.01 even when comparing a 13-week to a 260-week study.

compared to other diseases, etc. We do not have sufficient power to determine the extent to which these factors influence the effect of water treatment on child mortality, as we demonstrate in simulations (Materials and Methods, section 6).

However, we find that the studies included in this meta-analysis are broadly representative of the settings in which policymakers might implement water treatment programs in terms of diarrhea prevalence and there are no significant differences from a larger sample of 73 studies (Materials and methods, section 1). There are some plausible hypotheses for the treatment effect diminishing over time due to improvements in quality of health care and availability of rotavirus vaccines (Materials and methods, section 7) and more data will be needed to test them.

Lessons for meta-analysis and pre-analysis plans

Methodologically, our results suggest that meta-analysis may be important for assessing effects which are small in absolute magnitude yet potentially large enough to be highly cost-effective. Unfortunately, multiple hypothesis testing requirements could potentially discourage authors from reporting outcomes for which power is low.

As we have shown, restricting the analysis to the five RCTs which reported mortality in publications leads to a similar, but much less precise estimate of effect. By including the additional ten studies we were able to increase statistical power (see Table S4, Fig 3). However, this necessitated a time-consuming process of contacting authors to request the data and led to the loss of some data that was once available but is no longer available.

One potential reform would be for pre-analysis plans to include a section listing outcomes for which the study is underpowered, either because the outcome is rare or noisily measured, but which will be reported for use in meta-analyses, and for individual studies to report such data, but not to be expected to conduct multiple hypotheses testing on such outcomes. Committees of scholars in the field could recommend a limited set of outcomes such as mortality for collection and incorporation in meta-analyses. Factors for inclusion could include importance and ease of data collection.

Data sharing

All data and code to replicate the results (including all figures and tables) of this meta-analysis has been made publicly available.

Acknowledgments

This research is supported by Dioraphte Foundation, Sint Antonius Stichting Foundation (Project Number 1808), and Givewell.

We thank Amrita Ahuja, Ben Arnold, Kanika Bahl, Radu Ban, Thomas Clasen, Jack Colford, Kevin Croke, Megan Higgs, Stephan Guyenet, Majdi Osman, Amy Pickering, Alix Zwane, and the participants of the Harvard School of Public Health, University of Chicago Virtual Advances with Field Experiments, and City University of New York seminars for helpful comments and suggestions.

We thank researchers Clair Null, Joseph Brown, Miles A. Kirby, Rachel Peletz, Rob Quick, Simon Collin, Sophie Boisson, and Pascaline Dupas for providing data and additional information on their work.

We thank Anbar Aizenman, Arthur Baker, Meera Nair, Alexandre Simoes Gomes, Tessa Han, Edward Jee, Dustin Marshal, Jack Marshall, Shubhalakshmi Nag, Cynthia Ji, and Nicholas Simmons for their research assistance.

Any errors are our own.

Declaration of interests

None of the authors have a conflict of interest or any financial conflict to disclose. Ricardo Maertens currently works at Amazon. He contributed to this research paper prior to joining Amazon.

References

1. United Nations Children's Fund (UNICEF) and World Health Organization, Progress on drinking water, sanitation and hygiene 2000–2017 (2019) (February 21, 2022).
2. , Global Burden of Disease Results Tool. *Glob. Burd. Dis. Results Tool Inst. Health Metr. Eval. IHME* (2015) (February 21, 2022).
3. A. K. Misra, Climate change and challenges of water and food security. *Int. J. Sustain. Built Environ.* **3**, 153–165 (2014).
4. L. A. McLaughlin, *et al.*, An Observational Study on the Effectiveness of Point-Of-Use Chlorination. *J. Environ. Health* **71**, 48–53 (2009).
5. J. A. Crump, *et al.*, Effect of point-of-use disinfection, flocculation and combined flocculation-disinfection on drinking water quality in western Kenya. *J. Appl. Microbiol.* **97**, 225–231 (2004).
6. K. Levy, *et al.*, Household effectiveness vs. laboratory efficacy of point-of-use chlorination. *Water Res.* **54**, 69–77 (2014).
7. L. B. Whan, I. R. Grant, H. J. Ball, R. Scott, M. T. Rowe, Bactericidal effect of chlorine on Mycobacterium paratuberculosis in drinking water. *Lett. Appl. Microbiol.* **33**, 227–231 (2001).
8. T. F. Clasen, *et al.*, Interventions to improve water quality for preventing diarrhoea. *Cochrane Database Syst. Rev.*, CD004794 (2015).
9. J. Wolf, *et al.*, Impact of drinking water, sanitation and handwashing with soap on childhood diarrhoeal disease: updated meta-analysis and meta-regression. *Trop. Med. Int. Health TM IH* **23**, 508–525 (2018).
10. H. Ritchie, M. Roser, Clean Water and Sanitation. *Our World Data* (2021) (February 21,

2022).
11. W.-P. Schmidt, S. Cairncross, Household water treatment in poor populations: is there enough evidence for scaling up now? *Environ. Sci. Technol.* **43**, 986–992 (2009).
12. A. P. Zwane, *et al.*, Being surveyed can change later behavior and related parameter estimates. *Proc. Natl. Acad. Sci.* **108**, 1821–1826 (2011).
13. G. A. Wells, *et al.*, The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses (February 21, 2022).
14. S. Balduzzi, G. Rücker, G. Schwarzer, How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Health* **22**, 153–160 (2019).
15. W. Wiecek, Meta-analysis of binary data with baggr (2022) (February 21, 2022).
16. I. Andrews, M. Kasy, Identification of and Correction for Publication Bias. *Am. Econ. Rev.* **109**, 2766–2794 (2019).
17. K. Stenberg, *et al.*, Cost-Effectiveness of Interventions to Improve Maternal, Newborn and Child Health Outcomes: A WHO-CHOICE Analysis for Eastern Sub-Saharan Africa and South-East Asia. *Int. J. Health Policy Manag.*, 1 (2021).
18. C. Null, *et al.*, Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial. *Lancet Glob. Health* **6**, e316–e329 (2018).
19. S. P. Luby, *et al.*, Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. *Trop. Med. Int. Health TM IH* **11**, 479–489 (2006).
20. S. P. Luby, *et al.*, Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. *Lancet Glob. Health* **6**, e302–e315 (2018).
21. J. A. Crump, *et al.*, Household based treatment of drinking water with flocculant-disinfectant for preventing diarrhoea in areas with turbid source water in rural western Kenya: cluster randomised controlled trial. *BMJ* **331**, 478 (2005).
22. R. Peletz, *et al.*, Assessing Water Filtration and Safe Storage in Households with Young Children of HIV-Positive Mothers: A Randomized, Controlled Trial in Zambia. *PLoS ONE* **7**, e46548 (2012).
23. , Data provided by Evidence Action.
24. , Annex 3C_Interventions in EUHC.pdf (October 31, 2022).
25. R. Black, R. Laxminarayan, M. Temmerman, N. Walker, "Disease Control Priorities, Third Edition: Volume 2. Reproductive, Maternal, Newborn, and Child Health" (World Bank, 2016) https:/doi.org/10.1596/978-1-4648-0348-2 (February 21, 2022).
26. J. C. Semenza, L. Roberts, A. Henderson, J. Bogan, C. H. Rubin, Water distribution system and diarrheal disease transmission: a case study in Uzbekistan. *Am. J. Trop. Med. Hyg.* **59**, 941–946 (1998).
27. M. E. Reller, *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.* **69**, 411–419 (2003).
28. T. Chiller, Reducing diarrhoea in Guatemalan children: randomized controlled trial of flocculant-disinfectant for drinking water. *Bull. World Health Organ.* **84**, 28–35 (2006).
29. M. Kremer, J. Leino, E. Miguel, A. P. Zwane, Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions*. *Q. J. Econ.* **126**, 145–205 (2011).
30. S. Boisson, *et al.*, Effect of Household-Based Drinking Water Chlorination on Diarrhoea

among Children under Five in Orissa, India: A Double-Blind Randomised Placebo-Controlled Trial. *PLoS Med.* **10**, e1001497 (2013).
31. J. H. Humphrey, *et al.*, Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. *Lancet Glob. Health* **7**, e132–e147 (2019).
32. M. A. Kirby, *et al.*, Effects of a large-scale distribution of water filters and natural draft rocket-style cookstoves on diarrhea and acute respiratory infection: A cluster-randomized controlled trial in Western Province, Rwanda. *PLoS Med.* **16**, e1002812 (2019).
33. J. Haushofer, M. Kremer, R. Maertens, B. J. Tan, "Water Treatment and Child Mortality: Evidence from Kenya" (National Bureau of Economic Research, 2021).
34. P. Dupas, B. Nhlema, Z. Wagner, A. Wolf, E. Wroe, Expanding Access to Clean Water for the Rural Poor: Experimental Evidence from Malawi. *Am. Econ. J. Econ. Policy* (2021) https:/doi.org/10.1257/pol.20210121 (February 21, 2022).
35. R. E. Quick, *et al.*, Diarrhoea prevention in Bolivia through point-of-use water treatment and safe storage: a promising new strategy. *Epidemiol. Infect.* **122**, 93–90 (1999).
36. S. Boisson, *et al.*, Field Assessment of a Novel Household-Based Water Filtration Device: A Randomised, Placebo-Controlled Trial in the Democratic Republic of Congo. *PLOS ONE* **5**, e12613 (2010).
37. M. du Preez, *et al.*, Randomized intervention study of solar disinfection of drinking water in the prevention of dysentery in Kenyan children aged under 5 years. *Environ. Sci. Technol.* **45**, 9315–9323 (2011).
38. The World Bank, The World by Income and Region. *World Dev. Indic.* (February 21, 2022).
39. S. Yusuf, R. Peto, J. Lewis, R. Collins, P. Sleight, Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Prog. Cardiovasc. Dis.* **27**, 335–371 (1985).
40. M. J. Bradburn, J. J. Deeks, J. A. Berlin, A. Russell Localio, Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events: META-ANALYSIS OF RARE EVENTS. *Stat. Med.* **26**, 53–77 (2007).
41. G. Schwarzer, J. R. Carpenter, G. Rücker, *Meta-Analysis with R* (Springer International Publishing, 2015) https:/doi.org/10.1007/978-3-319-21416-0 (February 22, 2022).
42. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
43. Evidence Action, Dispensers for Safe Water: Cost-effectiveness Model Guide (2019).
44. M. C. Opryszko, *et al.*, Water and hygiene interventions to reduce diarrhoea in rural Afghanistan: a randomized controlled study. *J. Water Health* **8**, 687–702 (2010).
45. P. Dupas, V. Hoffmann, M. Kremer, A. P. Zwane, Targeting health subsidies through a nonprice mechanism: A randomized controlled trial in Kenya. *Science* **353**, 889–895 (2016).
46. R. Bain, *et al.*, Global assessment of exposure to faecal contamination through drinking water based on a systematic review. *Trop. Med. Int. Health TM IH* **19**, 917–927 (2014).
47. K. Onda, J. Lobuglio, J. Bartram, Global access to safe water: accounting for water quality and the resulting impact on MDG progress. *World Health Popul.* **14**, 32–44 (2013).
48. T. Spoorenberg, "Data and methods for the production of national population estimates: An overview and analysis of available metadata" (UN Population Division, 2020).
49. R. M. Turner, D. Jackson, Y. Wei, S. G. Thompson, J. P. T. Higgins, Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian

meta-analysis. *Stat. Med.* **34**, 984–998 (2014).
50. W. T. Sedgwick, J. S. Macnutt, On the Mills-Reincke Phenomenon and Hazen§s Theorem Concerning the Decrease in Mortality from Diseases Other than Typhoid Fever Following the Purification of Public Water-Supplies. *J. Infect. Dis.* **7**, 489–564 (1910).
51. R. J. Evans, *Death in Hamburg: Society and Politics in the Cholera Years* (Penguin Books, Penguin Books) (February 21, 2022).
52. W.-P. Schmidt, S. Cairncross, M. L. Barreto, T. Clasen, B. Genser, Recent diarrhoeal illness and risk of lower respiratory infections in children under the age of 5 years. *Int. J. Epidemiol.* **38**, 766 (2009).
53. C. L. Coles, *et al.*, Nutritional Status and Diarrheal Illness as Independent Risk Factors for Alveolar Pneumonia. *Am. J. Epidemiol.* **162**, 999–1007 (2005).
54. C. L. F. Walker, J. Perin, J. Katz, J. M. Tielsch, R. E. Black, Diarrhea as a risk factor for acute lower respiratory tract infections among young children in low income settings. *J. Glob. Health* **3**, 010402 (2013).
55. S. Ashraf, M. H. Huque, E. Kenah, M. Agboatwalla, S. P. Luby, Effect of recent diarrhoeal episodes on risk of pneumonia in children under the age of 5 years in Karachi, Pakistan. *Int. J. Epidemiol.* **42**, 194–200 (2013).
56. S. Rouhani, *et al.*, Diarrhea as a Potential Cause and Consequence of Reduced Gut Microbial Diversity Among Undernourished Children in Peru. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **71**, 989–999 (2020).
57. K. L. Kotloff, *et al.*, Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet Lond. Engl.* **382**, 209–222 (2013).
58. T. J. Schuijt, *et al.*, The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. *Gut* **65**, 575–583 (2016).
59. S. Subramanian, *et al.*, Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* **510**, 417–421 (2014).
60. R. L. Guerrant, M. D. DeBoer, S. R. Moore, R. J. Scharf, A. A. M. Lima, The impoverished gut--a triple burden of diarrhoea, stunting and chronic disease. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 220–229 (2013).
61. H. Blencowe, *et al.*, Clean birth and postnatal care practices to reduce neonatal deaths from sepsis and tetanus: A systematic review and Delphi estimation of mortality effect. *BMC Public Health* **11 Suppl 3**, S11 (2011).
62. B. K. Padhi, *et al.*, Risk of Adverse Pregnancy Outcomes among Women Practicing Poor Sanitation in Rural India: A Population-Based Prospective Cohort Study. *PLOS Med.* **12**, e1001851 (2015).
63. S. L. James, *et al.*, Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1789–1858 (2018).
64. J. M. Colford Jr., Participant Blinding and Gastrointestinal Illness in a Randomized, Controlled Trial of an In-Home Drinking Water Intervention. *Emerg. Infect. Dis.* **8**, 29–36 (2002).
65. J. M. Colford, *et al.*, A Randomized, Controlled Trial of In-Home Drinking Water Intervention to Reduce Gastrointestinal Illness. *Am. J. Epidemiol.* **161**, 472–482 (2005).
66. S. Rodrigo, M. Sinclair, A. Forbes, D. Cunliffe, K. Leder, Drinking Rainwater: A

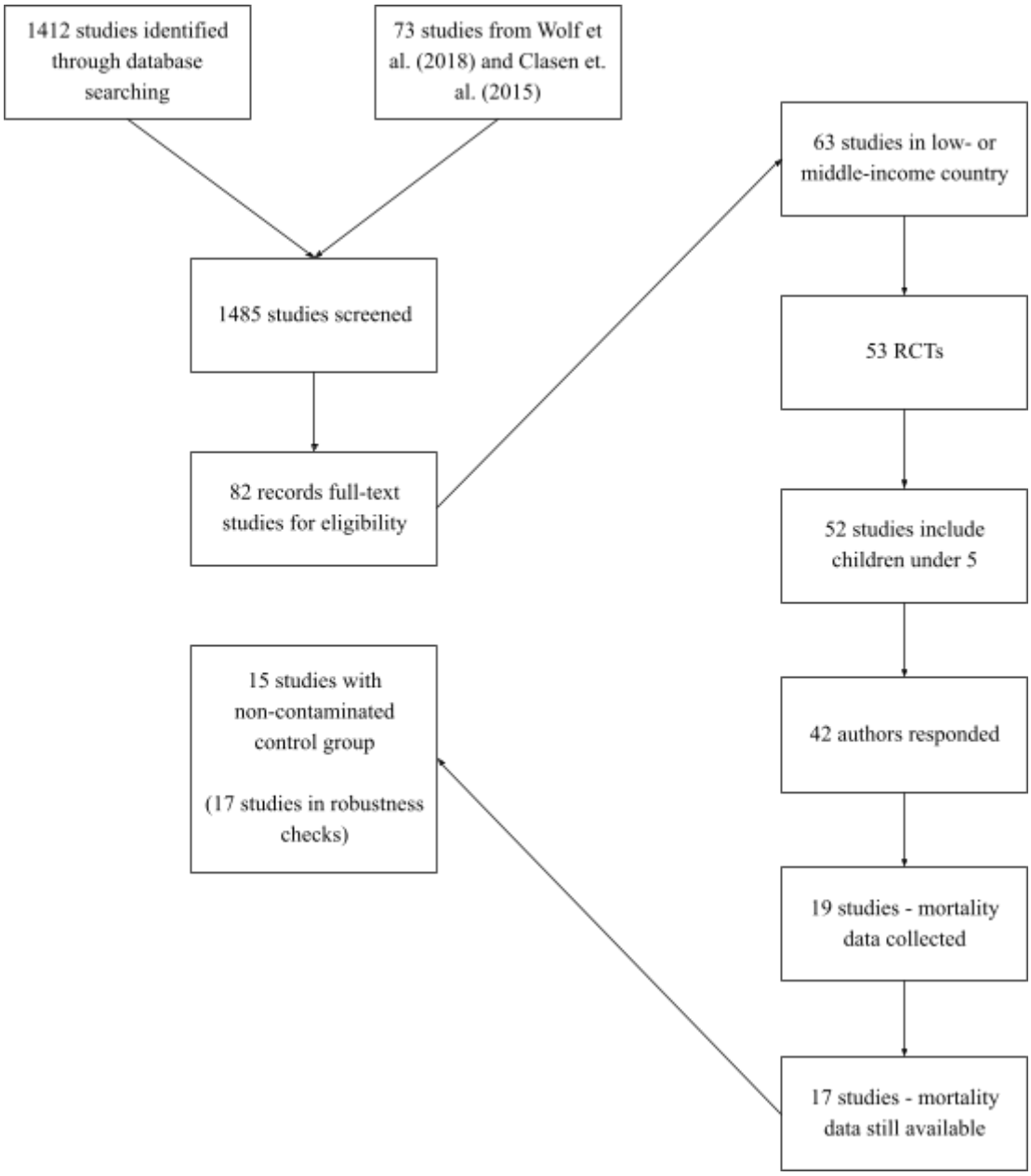
Double-Blinded, Randomized Controlled Study of Water Treatment Filters and Gastroenteritis Incidence. *Am. J. Public Health* **101**, 842–847 (2011).
67. L. V. Kirchhoff, *et al.*, Feasibility and efficacy of in-home water chlorination in rural North-eastern Brazil. *J. Hyg. (Lond.)* **94**, 173–180 (1985).
68. I. A. Alam, M. Sadiq, Metal contamination of drinking water from corrosion of distribution pipes. *Environ. Pollut.* **57**, 167–178 (1989).
69. A. A. Mahfouz, M. Abdel-Moneim, R. A. al-Erian, O. M. al-Amari, Impact of chlorination of water in domestic storage tanks on childhood diarrhoea: a community trial in the rural areas of Saudi Arabia. *J. Trop. Med. Hyg.* **98**, 126–130 (1995).
70. R. M. Conroy, M. Elmore-Meegan, T. Joyce, K. G. McGuigan, J. Barnes, Solar disinfection of drinking water and diarrhoea in Maasai children: a controlled field trial. *The Lancet* **348**, 1695–1697 (1996).
71. S. Xiao, C. Lin, K. Chen, [Evaluation of effectiveness of comprehensive control for diarrhea diseases in rural areas of east Fujian and analysis of its cost-benefit]. *Zhonghua Yu Fang Yi Xue Za Zhi* **31**, 40–41 (1997).
72. R. E. Quick, *et al.*, Diarrhea prevention through household-level water disinfection and safe storage in Zambia. *Am. J. Trop. Med. Hyg.* **66**, 584–589 (2002).
73. P. K. Jensen, *et al.*, Effect of chlorination of drinking-water on water quality and childhood diarrhoea in a village in Pakistan. *J. Health Popul. Nutr.* **21**, 26–31 (2003).
74. B. Majuru, M. Michael Mokoena, P. Jagals, P. R. Hunter, Health impact of small-community water supply reliability. *Int. J. Hyg. Environ. Health* **214**, 162–166 (2011).
75. M. Johri, M.-P. Sylvestre, G. K. Koné, D. Chandra, S. V. Subramanian, Effects of improved drinking water quality on early childhood growth in rural Uttar Pradesh, India: A propensity-score analysis. *PLOS ONE* **14**, e0209054 (2019).
76. H. Reese, *et al.*, Assessing longer-term effectiveness of a combined household-level piped water and sanitation intervention on child diarrhoea, acute respiratory infection, soil-transmitted helminth infection and nutritional status: a matched cohort study in rural Odisha, India. *Int. J. Epidemiol.* **48**, 1757–1767 (2019).
77. L. S. Abebe, *et al.*, Ceramic water filters impregnated with silver nanoparticles as a point-of-use water-treatment intervention for HIV-positive individuals in Limpopo Province, South Africa: a pilot study of technological performance and human health benefits. *J. Water Health* **12**, 288–300 (2014).
78. J. S. Gruber, *et al.*, A Stepped Wedge, Cluster-Randomized Trial of a Household UV-Disinfection and Safe Storage Drinking Water Intervention in Rural Baja California Sur, Mexico. *Am. J. Trop. Med. Hyg.* **89**, 238–245 (2013).
79. I. Günther, Y. Schipper, PUMPS, GERMS AND STORAGE: THE IMPACT OF IMPROVED WATER CONTAINERS ON WATER QUALITY AND HEALTH: PUMPS, GERMS AND STORAGE. *Health Econ.* **22**, 757–774 (2013).
80. C. K. Jain, A. Bandyopadhyay, A. Bhadra, Assessment of ground water quality for drinking purpose, District Nainital, Uttarakhand, India. *Environ. Monit. Assess.* **166**, 663–676 (2010).
81. M. K. Patel, *et al.*, Impact of a hygiene curriculum and the installation of simple handwashing and drinking water stations in rural Kenyan primary schools on student health and hygiene practices. *Am. J. Trop. Med. Hyg.* **87**, 594–601 (2012).
82. L. Roberts, *et al.*, Keeping clean water clean in a Malawi refugee camp: a randomized

intervention trial. *Bull. World Health Organ.* **79**, 280–287 (2001).
83. S.-S. K. Tiwari, W.-P. Schmidt, J. Darby, Z. G. Kariuki, M. W. Jenkins, Intermittent slow sand filtration for preventing diarrhoea among children in Kenyan households using unimproved water sources: randomized controlled trial: **Sand filtration to reduce childhood diarrhoea**. *Trop. Med. Int. Health* **14**, 1374–1382 (2009).
84. , Contra la morbilidad infantil: Filtros artesanales y educación (1995).
85. S. Boisson, W.-P. Schmidt, T. Berhanu, H. Gezahegn, T. Clasen, Randomized Controlled Trial in Rural Ethiopia to Assess a Portable Water Treatment Device. *Environ. Sci. Technol.* **43**, 5934–5939 (2009).
86. S. Doocy, G. Burnham, Point-of-use water treatment and diarrhoea reduction in the emergency context: an effectiveness trial in Liberia. *Trop. Med. Int. Health TM IH* **11**, 1542–1552 (2006).
87. C. E. Stauber, G. M. Ortiz, D. P. Loomis, M. D. Sobsey, A randomized controlled trial of the concrete biosand filter and its impact on diarrheal disease in Bonao, Dominican Republic. *Am. J. Trop. Med. Hyg.* **80**, 286–293 (2009).
88. C. Stauber, B. Kominek, K. Liang, M. Osman, M. Sobsey, Evaluation of the Impact of the Plastic BioSand Filter on Health and Drinking Water Quality in Rural Tamale, Ghana. *Int. J. Environ. Res. Public. Health* **9**, 3806–3823 (2012).
89. C. E. Stauber, E. R. Printy, F. A. McCarty, K. R. Liang, M. D. Sobsey, Cluster Randomized Controlled Trial of the Plastic BioSand Water Filter in Cambodia. *Environ. Sci. Technol.* **46**, 722–728 (2012).
90. E. D. Lindquist, *et al.*, A Cluster Randomized Controlled Trial to Reduce Childhood Diarrhea Using Hollow Fiber Water Filter and/or Hygiene–Sanitation Educational Interventions. *Am. J. Trop. Med. Hyg.* **91**, 190–197 (2014).
91. A. Fabiszewski de Aceituno, C. Stauber, A. Walters, R. M. Sanchez, M. Sobsey, A Randomized Controlled Trial of the Plastic-Housing BioSand Filter and its Impact on Diarrheal Disease in Copan, Honduras. *Public Health Fac. Publ.* (2012).
92. T. F. Clasen, S. Cairncross, Editorial: Household water management: refining the dominant paradigm. *Trop. Med. Int. Health* **9**, 187–191 (2004).
93. T. F. Clasen, J. Brown, S. Collin, O. Suntura, S. Cairncross, REDUCING DIARRHEA THROUGH THE USE OF HOUSEHOLD-BASED CERAMIC WATER FILTERS: A RANDOMIZED, CONTROLLED TRIAL IN RURAL BOLIVIA. *Am. J. Trop. Med. Hyg.* **70**, 651–657 (2004).
94. A. J. Pickering, *et al.*, Effect of in-line drinking water chlorination at the point of collection on child diarrhoea in urban Bangladesh: a double-blind, cluster-randomised controlled trial. *Lancet Glob. Health* **7**, e1247–e1256 (2019).
95. T. R. Handzel, The effect of improved drinking water quality on the risk of diarrheal disease in an urban slum of Dhaka, Bangladesh: A home chlorination intervention trial. *ProQuest Diss. Publ.* (February 21, 2022).
96. J. Gasana, J. Morin, A. Ndikuyeze, P. Kamoso, Impact of Water Supply and Sanitation on Diarrheal Morbidity among Young Children in the Socioeconomic and Cultural Context of Rwanda (Africa). *Environ. Res.* **90**, 76–88 (2002).
97. J. M. Brown, S. Proum, M. D. Sobsey, Escherichia coli in household drinking water and diarrheal disease risk: evidence from Cambodia. *Water Sci. Technol.* **58**, 757–763 (2008).
98. C. J. Austin, Chlorinating household water in The Gambia in *Proceedings of the 19th WEDC International Conference, Accra, Ghana*, (1993) (February 21, 2022).

99. B. Mengistie, Y. Berhane, A. Worku, Household Water Chlorination Reduces Incidence of Diarrhea among Under-Five Children in Rural Ethiopia: A Cluster Randomized Controlled Trial. *PLoS ONE* **8**, e77887 (2013).
100. D. Mäusezahl, *et al.*, Solar Drinking Water Disinfection (SODIS) to Reduce Childhood Diarrhoea in Rural Bolivia: A Cluster-Randomized, Controlled Trial. *PLoS Med.* **6**, e1000125 (2009).
101. J. R. Lule, *et al.*, Effect of home-based water chlorination and safe storage on diarrhea among persons with human immunodeficiency virus in Uganda. *Am. J. Trop. Med. Hyg.* **73**, 926–933 (2005).
102. M. du Preez, K. G. McGuigan, R. M. Conroy, Solar Disinfection of Drinking Water In the Prevention of Dysentery in South African Children Aged under 5 Years: The Role of Participant Motivation. *Environ. Sci. Technol.* **44**, 8744–8749 (2010).
103. "Kenya Demographic and Health Survey 2014" (Kenya National Bureau of Statistics, Ministry of Health/Kenya, National AIDS Control Council/Kenya, Kenya Medical Research Institute, National Council for Population and Development/Kenya, and ICF International, 2015) (February 21, 2022).
104. Child Mortality Estimates. *UN Inter-Agency Group Child Mortal. Estim.* (2020) (February 21, 2022).

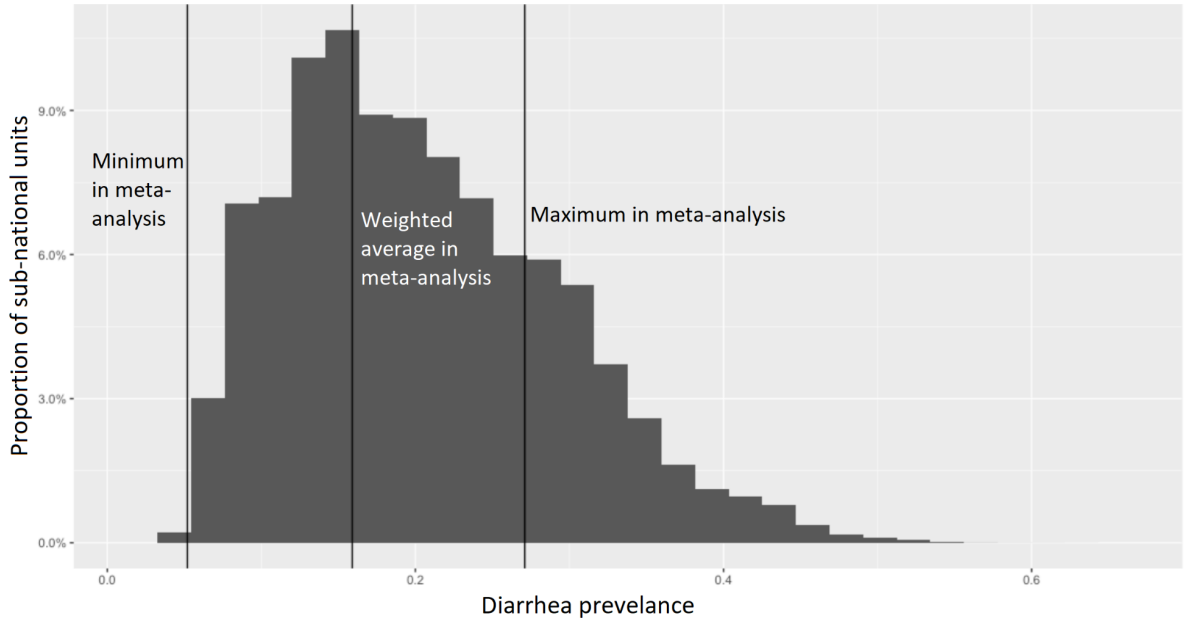
Figures and Tables

Fig. 1. Study selection



Note: This funnel chart depicts the search strategy and selection criteria for studies included in the meta-analysis. Non-contamination of control groups was added as an inclusion criteria after reviewing the studies and was not decided prior to the review of studies

Fig. 2. Distribution of diarrhea prevalence in low- and middle-income countries

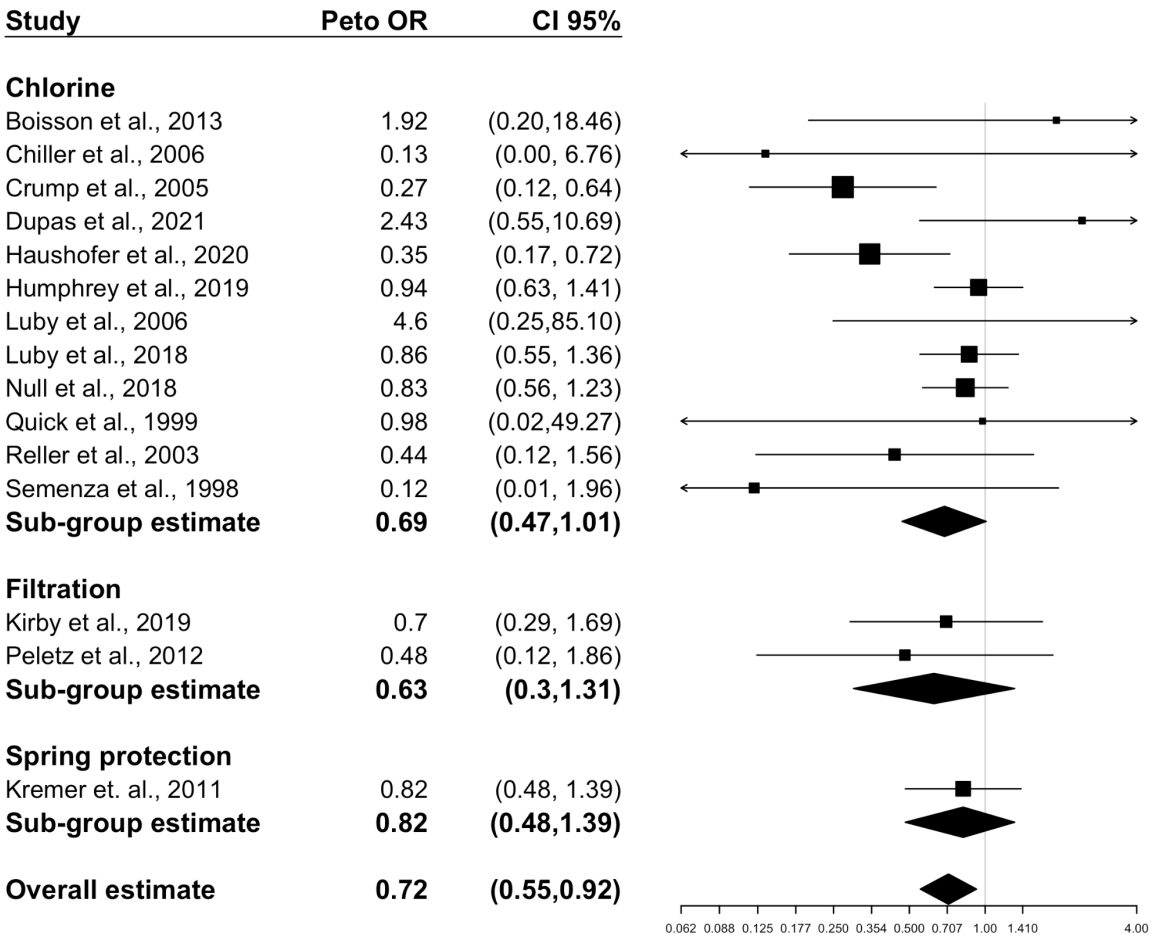


Note: This histogram shows diarrhea prevalence (%) across sub national geographic units of 94 low- and middle-income countries as of 2017. Black lines indicate the minimum, weighted average, and maximum diarrhea prevalence in studies included in the meta-analysis.

Source: Institute for Health Metrics and Evaluation, 2020 (17)

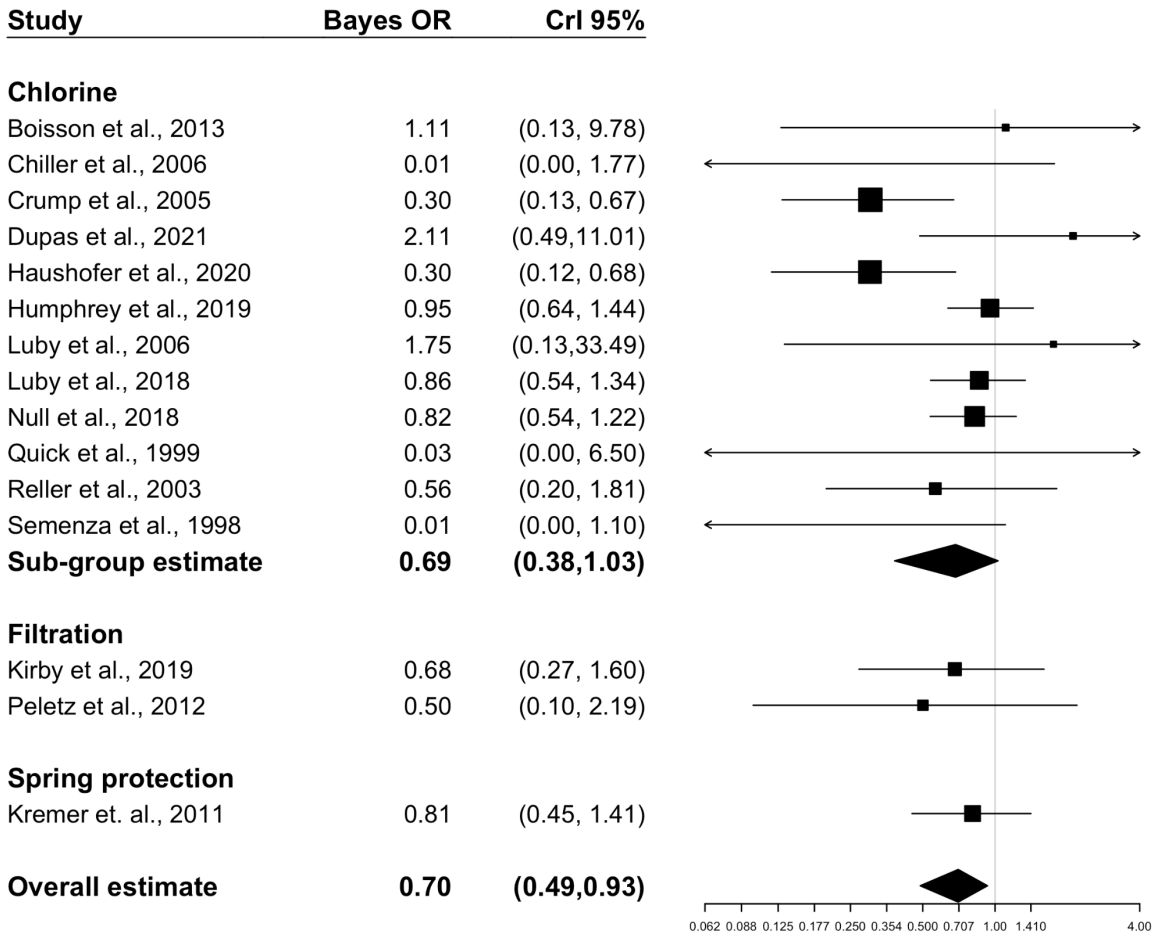
Fig. 3. Forest plots of meta-analysis results

(A) Peto Odds Ratio



Note: Dots and horizontal lines represent mean estimates and their 95% confidence intervals from individual studies. Estimates for individual studies are Peto odds ratios. The size of each dot represents the weight given to the study. Diamonds are centered around the meta-analysis estimates and their widths indicate the 95% confidence/credible interval. In addition to the overall estimate we also show estimates for subgroups of studies by intervention type.

(B) Bayesian Odds Ratio



Note: Dots and horizontal lines represent posterior means and 95% credible intervals: for individual studies they are the ORs under Bayesian no pooling model, for chlorine studies and overall estimates they are the estimates from partially pooled (random effects) model. See Materials and Methods, section 2 for details, including choice of priors. We do not report the Bayesian OR for the subset of filtration studies because we only have two filtration studies in our sample and the parameters of the Bayesian hierarchical model are not well-identified in that case.

Table 1. Summary of studies included

Study	Intervention	Sample size	Country	Population of study	Observation period	Infectious environment indicators		Compliance rate**
						Contamination levels	Diarrhea prevalence	
Semenza et al., 1998 (26)	(1) Chlorination	(1) 68 hh (C) 58 hh	Uzbekistan	Households with a <5y old child	9.5 weeks	54 TTC/100 ml pre-treatment	12.77% in control children	73%
Reller et al., 2003 (27)	(1) Flocculant-disinfectant (2) Flocculant-disinfectant + vessel (3) Chlorination (4) Chlorination + vessel	(1) 102 hh (2) 97 hh (3) 97 hh (4) 100 hh (C) 96 hh	Guatemala	Households with a ≤11m old or pregnant woman in third trimester	1 year	Concentration of E. coli per 100ml: 63	13.2% in control children (≤12m)	(1) 27% (2) 34% (3) 36% (4) 44%
Crump et al., 2005 (21)	(1) Chlorination (2) Flocculant-disinfectant	(1) 203 hh (2) 201 hh (C) 201 hh	Kenya	Family compounds with at least one child <2y old	20 weeks	Concentration of E. coli per 100ml: 98 (mean at baseline); Share of households meeting WHO water quality standard: 14% in control	9.6% in control children; 2.7% in control group (all ages)	52.5%
Luby et al., 2006 (19)	(1) Chlorination (2) Flocculant‐disinfectant water treatment	(1) 265 hh (2) 262 hh (C) 282 hh	Pakistan	Households with a <5y old child	37 weeks	Diarrhea is a leading cause of death and the environment is heavily contaminated with sewage	8.62% in control group	Unavailable
Chiller et al., 2006 (28)	(1) Chlorination	(1) 1702 ind. (C) 1699 ind.	Guatemala	Households with a <1y old child	13 weeks	98% drinking water sources contaminated with E. coli at beginning of study	6% in control group	85%

Kremer et. al., 2011 (29)	(1) Spring protection	(1) 749 hh (C) 685 hh	Kenya	Households which use selected springs	2 years	Concentration of E. coli per 100ml: 44.3	20% in control children	69%
Peletz et al., 2012 (22)	(1) Filtration	(1) 61 hh (C) 59 hh	Zambia	Households with a 6m-1y old at enrollment and with HIV+ mothers (100 HIV+ and 20 HIV-)	1 year	181 TTC/100ml for control (endline)	13.6% in control children (<2 y)	87%
Boisson et al., 2013 (30)	(1) Chlorination	(1) 1080 hh (C) 1083 hh	India	Households with a <5y old child	1 year	122 TTC/100 ml in control over the course of the study	5.2% at baseline for children (<5 y)	32.0%
Null et al., 2018 (18)	(1) Chlorination	(1) 904 hh (C) 1913 hh	Kenya	Newborns	2 years	>75% of household collected water from improved water sources at baseline	27.1% in (active) control group*	30%
Luby et al., 2018 (20)	(1) Chlorination + vessel	(1) 698 hh (C) 1382 hh	Bangladesh	Newborns and their siblings under 36m old	2 years	74% collected drinking water from shallow tube wells at baseline	5.7% in control group	81%
Humphrey et al., 2019 (31)	(1) Chlorination + sanitation + hand washing + play space + hygiene counseling + construction of improved pit latrines (WASH)	(1) 918 children. (C) 884 children.	Zimbabwe	Households with a <18m old child	1.5 years	63% of household collected water from improved water sources at baseline	9.5% in control	79%
Kirby et al., 2019 (32)	(1) Filtration + Cookstoves	(1) 87 vill. (C) 87 vill.	Rwanda	Households with a <5y old child	1.25 years	>100 TTC/100 ml for 38% of households in control	12.9% in control	69.9%
Haushofer et al. 2021 (33)	(1) Chlorination (follow-up to Null et al., 2018)	(1) 65 vill. (C) 67 vill.	Kenya	Children <5y	4-6 years	>75% of household collected water from improved water sources at baseline	27.1% in (active) control group*	31%

Dupas et al. 2021 (34)	(1) Coupons for chlorination (subsidy) (2) Coupons + free delivery + WASH promotion (3) Coupons + WASH (4) WASH	(1) 441 hh (2) 458 hh (3) 468 hh (4) 468 hh (C) 460 hh	Malawi	Households with a <6y old child	61 weeks	70.7% of household collected water from a protected water source at baseline	12.4% in control group	40%
Quick et al. 1999 (35)	(1) Chlorination + safe storage of treated water + community education	(1) 400 ind. (C) 391 ind.	Bolivia	All households in study communities	34 weeks	Median E. coli colony count for well water (baseline): 34/100 ml and for stored water (baseline): 44/ 100 ml	38.0% in control group	71%
Studies included for robustness checks								
Boisson et al., 2010 (36)	(1) Filtration	(1) 546 ind. (C) 598 ind.	Democratic Republic of Congo	All households in selected communities	1 year	75% of source water samples had >1,000 TTC/100 ml	8.96% in control children (<5 y)	68%
du Preez et al., 2011 (37)	(1) Solar disinfection	(1) 579 children. (C) 554 children.	Kenya	Children 6m-5yo	1.5 years	Most households collected water from standpipes (with treated water)	5.2% of dysentery in control	68%

Notes: * In Null et al., 2018 (18) there was an active control group which received enumerator visits and a passive control group with no visits. **Compliance rate was defined in a way that was specific to each study; we provide these definitions in Materials and Methods, section 1.

For each study, the corresponding meta-analysis input data for each study - i.e. the number of events (deaths) and non-events in each study - are reported in Table S3. Abbreviations: hh: “households;” ind.: “individuals;” child.: “children;” (C): “(Control).”.

Table 2. Cost–effectiveness analysis

	Chlorine Dispensers in Western Kenya	Global Coupon Program
Estimated mean OR effect of water treatment on child mortality, mean (95% CrI)	0.70 (0.49, 0.93)	0.70 (0.49, 0.93)
Posterior predictive estimate (RR) of effect, mean	0.75	0.75
Expected deaths averted per person	0.015	0.007
Expected DALYs averted per person	1.16	0.53
Cost per expected death averted (USD)	3,104	2,974
Cost per expected DALY averted (USD)	39	38

This cost-effectiveness calculation is based on the Bayesian logit model and incorporates uncertainty in predicting effects to a new setting.Details of calculation and assumptions of costs are given in Table S8.

Supplementary Materials: Water Treatment and Child Mortality: A Systematic Review and Meta-Analysis

Michael Kremer, Stephen Luby, Ricardo Maertens, Brandon Tan, Witold Więcek

Materials and Methods

Figs. S1 to S11

Tables S1 to S8

- Table S4. The child mortality impact of water quality interventions – studies which report child mortality effects
- Table S5. Sensitivity of main results to dropping each study
- Table S6. Additional sensitivity checks
- Table S7. Total lives saved and costs: preliminary calculations from a global Coupon Program
- Table S8: Cost–effectiveness analysis

Materials and Methods

Section 1 presents a summary of included studies and a comparison between studies included and excluded from our analysis. Section 2 presents the meta-analysis models. Section 3 presents sensitivity analyses. Section 4 presents additional considerations for cost-effectiveness analysis. Section 5 presents an assessment of publication bias (for both diarrhea and mortality outcomes) and exploratory simulation of small study publication bias. Section 6 provides an exploratory assessment of power to detect heterogeneous effects. Section 7 compares meta-analysis estimates with model predictions.

1. Details of included studies and comparison with other RCTs

Summary of 15 included studies

The median follow-up length for mortality was 52 weeks, with the longest follow-up being 4-6 years (33). Ten studies were conducted in lower-middle income countries, and five were conducted in low-income countries, according to the World Bank classification at the time of the study (38). The age at which children were enrolled, as well as the periods for which they were followed, varied across studies; see below.

Out of the 15 studies, 13 were conducted in rural areas, one was conducted in both rural and urban areas, and one was conducted in a peri-urban setting. The compliance rate (see Table 1 for definition) in the sample ranged from a low of 27% to a high of 87%, with a median of 69%.

In all included studies, the primary outcomes were intermediate outcomes such as diarrhea, while mortality data was collected as a secondary outcome, as part of internal respondent tracking systems, or for IRB reporting purposes by the authors. Only five studies explicitly report mortality outcomes in the published manuscript, highlighting the importance of following up with authors to request the mortality data.

Baseline contamination levels of water in studies are also reported in Table 1. Contamination level measures are not consistently reported across studies. Four studies report 54 to 181 TTC/100 ml (thermotolerant coliforms, which include *E. Coli* and three other bacteria species). Another 4 studies report *E.Coli* concentration from 34 to 98 per 100 ml.

Estimates of diarrhea prevalence among the 15 included studies are representative of prevalence in low- and middle-income countries. Household surveys across 94 low- and middle-income countries found diarrhea prevalence in 2017 ranging from 3.2% to 66.4% across sub national units, with a median of 19.2% (18). Diarrhea prevalence rates (at baseline or, if baseline not available, in the control group) in our sample of studies range from 5.2% to 27.1%, with a weighted mean (using weights from the Peto model) of 15.6%; this corresponds to the 35th percentile of the distribution of sub national diarrhea estimates (see Figure 2).

Choice of treatment and control groups

When relevant, multiple treatment or control arms were combined so as to maximize power and to avoid introducing the correlation between treatment effect estimators that would arise if different treatments were compared to the same control group. For two studies (18, 33) which report the impact of closely related interventions on different samples, we report sensitivity to combining these studies. For 13 out of 15 studies, water treatment was compared to a pure control group which received no intervention. In two of these, several experimental arms with different kinds of water treatment were combined. These included some combination of water chlorination, flocculant-disinfection, and safe storage vessels (19, 27). In two cases, water treatment was combined with another intervention, cookstoves (32) or other sanitation and hygiene interventions (31).

Definition of compliance variable

Our definition of compliance for each study depends on the type of treatment and the available data. For studies involving chlorination, compliance was defined as the percentage of stored water samples (one per household) with detectable free chlorine above 0.1 ppm (Chiller et al. 2006; Reller et al. 2003; Luby et al. 2018; Haushofer et al. 2020; Crump et al. 2005; Humphrey et al. 2019) or the percentage of samples with any detectable chlorine (Semenza et al. 1998; Boisson et al. 2013; Dupas et al. 2021; Quick et al. 1999; Null et al. 2018). Data from unannounced visits was used whenever it was available. In Crump et al. (2005), compliance was recorded as an average across two treatment groups, and in Null et al. (2018), compliance was measured from the one-year follow-up. For the two studies on water filtration, compliance was defined as the percentage of households which had a filter and reported using it in the last three days (Kirby et al. 2019; Peletz et al. 2012). Peletz et al. (2012) additionally required that reportedly-treated stored water with a low measured bacteria concentration was present in the household. Finally, for the study involving spring protection, compliance was measured as the increase in the fraction of trips to protected springs in the treatment group (Kremer et al. 2011).

Definitions for each study:

- Semenza et al: percentage with detectable chlorine residuals in the water at the time of visit
- Reller et al: Proportion of households drinking water with detectable free chlorine > 0.1 mg/L
- Crump et al: Average effect across two treatment groups
- Chiller et al: Residual free chlorine concentration > 0.1 ppm (scheduled visits)
- Kremer et al. (2011): increase in fraction of trips to protected springs (units = percentage points, not percentage)
- Peletz et al. (2012): percentage of households satisfying:

    - The water filter was observed in household at the time of visit
    - The storage vessel contained water reported to be treated at the time of visit
    - The respondent reported using the filter on the day of or day prior to the day of visit.
    - There was at least a 1 log10 TTC improvement in stored household water over their unfiltered water, or stored water quality was 10 TTC/100 mL
- Boisson et al: Presence of residual chlorine in child's drinking water
- Null et al. (2018): Just used detectable free chlorine measured in one-year follow-up
- Luby et al. (2018): Stored drinking water has detectable free chlorine (>0·1 mg/L) at 2-year follow-up
- Humphrey et al.: Percentage of households with detectable free chlorine above 0.1 ppm
- Kirby et al: "Filter observed and reports last filled since day before yesterday"
- Haushofer et al: Household uses chlorine dispensers (unannounced visit)
- Dupas et al. (2021): Positive chlorine test
- Quick et al.: Proportion of stored water samples with detectable levels of total chlorine

Comparison of characteristics between included and excluded studies

We additionally compare some key characteristics of the water treatment studies included with those excluded from the analysis, but included in (9). There were 73 studies in (9), yielding 80 observations. Some studies had multiple observations on account of multiple study locations, and hence yielded multiple effect estimates. 7 of these studies were included in our meta-analysis, resulting in 73 observations excluded from our meta-analysis.

The distribution of effect estimates of water treatment on diarrhea and compliance rates are similar across included and excluded data (see Fig. S4).

47 out of 73 observations (64%) are conducted in a rural setting, with 15.1% and 20.5% being conducted in mixed and urban settings respectively. Similar to this, among the included studies, 73.3% (11 out of 15 studies) of the studies are set in rural areas and the proportion of the studies conducted in mixed and urban settings is 13.3% for both.

In terms of the water source, the primary source of water at baseline (or in the control group) was an unimproved water source in 49 out of 73 observations (67%.) This is comparable to 86.6% (13 out of 15 studies) among the included studies.

A t-test of mean difference between included and excluded studies yield insignificant differences for the diarrhea effect size (p-val=0.34), compliance rate (p-val=0.23), setting (binary variable indicating whether the setting was rural, p-val=0.51), and water source type (binary variable indicating if the primary water source was unimproved, p-val=0.13).

Age characteristics of included children

Two studies excluded some of under 5 year olds at enrollment:

- Luby et al., 2018 did not collect data for children over 3 years or under three months old at the time of enrollment
- Null et al., 2018 did not collect data for children older than 2 years

Where available, we used individual-level data to characterize age composition of samples in 11 studies where we had access to age data.

Below is a summary of person-years at each age, per study. We also calculated mean age at follow-up, varying from 1.0 (Null et al., 2018) to 3.5 (Kremer et al., 2011). In two studies – Null et al., 2018 and Haushofer et al., 2020 – most of data collected came from children under the age of 2, while Kremer et al. 2011, Boisson et al. 2013, Dupas et al., 2021 included much more information on children aged over 2. (To illustrate this person-years calculation, if a child was followed up between ages of 1.5 and 3 we counted 0.5 person-years in the "Age 1" column and 1 year in "Age 2".)

Because in some cases the ages and follow-up times were not precisely recorded (e.g. rounded up to a year and in four studies no age information was given), we report ranges rather than means. We italicize studies with no age information.

	Person-years						
Study	**<1 W**	**1 W-1 M**	**1 M-1 Y**	**1-2 Ys**	**2-3 Ys**	**3-4 Ys**	**4-5 Ys**
Boisson et al., 2013	6	22	445	601	574	562	543
Chiller et al., 2006	0	0-1	102-111	32-42	41-51	50-60	35-45
Crump et al., 2005	3	11	80-85	160-165	121-126	111-116	66-71
Dupas et al., 2021	0	1	241	712	689	578	429
Haushofer et al., 2020	38	126-127	1603-1605	1202-1203	742-743	325	24
Humphrey et al., 2019	0-38	0-125	0-1791	0-977	0	0	0
Kirby et al., 2019	5	19	414-418	657-661	682-684	758-761	610-614
Kremer et. al., 2011	0	2	496-509	1178-1188	1332-1338	1030-1033	469
Luby et al., 2006	0	0	204	168	207	253	274
Luby et al., 2018	0-38	0-126	0-1798	0-1962	0-1962	0-1962	0-1962
Null et al., 2018	62-69	189-217	2492-2895	579-993	47-454	0-1	0
Peletz et al., 2012	0	0	41	54	0	0	0
Quick et al., 1999	0-15	0-51	0-725	0-791	0-791	0-791	0-791
Reller et al., 2003	1-2	5-6	210-224	283-298	115-130	184-200	82-98
Semenza et al., 1998	0-3	0-11	0-154	0	0	0	0
Total	116-218	374-717	6328-11247	5625-9814	4549-7748	3851-6643	2532-5320
Total/Interval Length	**6040-11347**	**5836-11185**	**6903-12269**	**5625-9814**	**4549-7748**	**3851-6643**	**2532-5320**

Note: The above table displays person-years of data within each study and age group. In cases where age or follow-up time are not known precisely, the minimum and maximum numbers of person-years are shown. The final row adjusts for time interval length (~0.02 years for column 1, 0.06 years for column 2, 0.92 years for column 3, and 1 year for all other columns). Studies with no age information are in italics.

As discussed in the main text, understanding possible heterogeneity of treatment effects with age would require us to use survival analysis, which we plan to do in the future.

To validate individual-level data we also summed person-years within each study and compared those values to the number of person-years that was implied by multiplying the publication-stated follow-up length by number of subjects. With the exception of Haushofer et al., 2020 (5-year follow-up), we found values agreed with what was stated in publications. For Haushofer et al., 2020 the actual time between start of treatment and last follow-up was about 2 years. This is because that study (in its main model specification, which we follow in this paper) only included children born after the year of the intervention, thereby reducing the average follow-up length in the observed data.

Study	**N**	**Length (weeks)**	**Total person-years in microdata**	**Average follow-up length (weeks), microdata**
Chiller et al., 2006	1093	13	22	12.4
Crump et al., 2005	1538	20	46	18.7
Haushofer et al., 2020	1981	260	339	106.7
Kirby et al., 2019	2470	65	263	66.6
Kremer et. al., 2011	2221	104	378	106.3
Null et al., 2018	3699	104	511	86.2
Peletz et al., 2012	121	52	8	40.8
Reller et al., 2003	926	52	73	49.5
Boisson et al., 2013	2991	52	229	47.9
Dupas et al., 2021	2616	61	221	52.7
Luby et al., 2006	1548	37	92	37.2
Semenza et al., 1998	168	10	NA	NA
Humphrey et al., 2019	1954	78	NA	NA
Luby et al., 2018	1962	104	NA	NA
Quick et al., 1999	791	34	NA	NA

2. Meta-analysis models

Comparing odds ratios and risk ratios for rare events

We chose modeling of odds ratios (ORs) instead of risk ratios (RRs). The standard odds ratio estimator is given by $\frac{a/c}{b/d}$ where a is the observed number of events (deaths) in the treatment group, and c is the number of non-events in the treatment group. Similarly, b and d are the number of events and non-events in the control group respectively. A normal approximation of the logarithm OR is typically used to meta-analyze odds ratios. Under this notation, risk ratios are given by $\frac{a/c}{(b+a)/(d+c)}$. As we can see, when a and c are small in relation to b and d, respectively, relative odds will be close to relative risks.

The following table illustrates how OR changes as a function of event rate in the controls and RR, for a plausible range of values for mortality:

	RR = 0.9	**RR = 0.8**	**RR = 0.7**
c / (c+d) = 1%	OR = 0.899	OR = 0.798	OR = 0.698
c / (c+d) = 2%	OR = 0.898	OR = 0.797	OR = 0.696
c / (c+d) = 5%	OR = 0.895	OR = 0.792	OR = 0.689

Peto odds ratio model

The main frequentist specification is the Peto one-step OR method. However, the sample OR and/or its variance estimates are undefined when there are zero deaths in either the control or the treatment group, meaning that the standard meta-analysis approach requires dropping three studies (19, 26, 28).[7] Instead, the Peto one-step method computes an approximation of the log odds ratio which allows for zero deaths in one of either the control or treatment group (39), but is not defined when both control and treatment arms have no events.

The Peto odds ratio under the assumption of fixed effects is estimated as follows:

[7] Chiller et al., 2006 (28), Semenza et al., 1998 (26) report zero deaths in the treatment group, and Luby et al., 2006 (19) has zero deaths in the control group.

$$\psi = exp\left(\frac{\sum_{k=1}^{m}(O_k - E_k)}{\sum_{k=1}^{m} V_k}\right)$$

where

$$O = a,$$

$$E = \frac{(a+b)(a+c)}{n},$$

$$\text{and } V = \frac{(a+b)(c+d)(a+c)(b+d)}{n^2(n-1)},$$

where *a* is the number of treatment group participants who died, *b* is the number of treatment group participants who did not, *c* is the number of control group participants who died, and *d* is the number of control group participants who did not; n = a+b+c+d is the total number of participants in a given study (in the notation above we drop subscripts for simplicity); *k* indexes each study, and *m* is the total number of studies.

The canonical Peto OR specification (39) uses fixed effects. A random-effects specification may be preferable since treatment effect heterogeneity is expected due to differences across studies in ages of children, baseline child mortality rates, baseline water contamination, treatment compliance, and water treatment technologies.[8] We therefore fit a random-effect model of Peto log odd ratios by using the default Restricted Maximum Likelihood (REML) estimator as implemented in R package meta (41). We use a typical continuity correction of adding 0.5 events to each of the cells (*a,b,c,* and *d*) in the study where *a=0* and *c=0*.

Bayesian logistic meta-analysis model

We also estimate the effect under a hierarchical Bayesian logistic regression model. This model is particularly suitable for the setting as it is able to handle zero death events and also model heterogeneity. The model accounts for both sampling variation and heterogeneity across studies by applying a logit model of individual-level data (which can be generated from aggregate data on numbers of events and non-events in each study), as follows:

$$y_i \mid \mu_{k(i),}\tau_{k(i)},\ T_i \sim Bernoulli\left(logit^{-1}\left[\mu_{k(i)} + \tau_{k(i)}T_i\right]\right)$$

where,

[8] However, meta-analysis simulation studies show that heterogeneity may have a minor impact on estimates when rare events are considered (28, 40).

$$\tau_k \sim N\left(\tau, \sigma_\tau^2\right) \quad \forall k,$$

$$and\ \mu_k \sim N\left(\mu, \sigma_\mu^2\right) \quad \forall k,$$

where y_i is an indicator variable for child i being dead, and T_i is an indicator for the treatment group; μ_k corresponds to study-specific control group probabilities of event and τ_k are the estimated study-specific effects. Under this formulation, the mean (also referred to as hypermean) log odds ratios of death between treatment and non-treatment in the population of included studies are given by τ and σ_τ^2 (hypervariance) reflects the true variation in mean effects across settings. Rates of events in the control arms are also partially pooled, i.e. assigned a hierarchical distribution.

For the main specification, we use mildly informative priors on the hyper-parameters, similar to (15). For τ, we set a normal distribution with mean 0 and standard deviation of 10. This prior encodes the belief that causal effects should not be thought of as large unless data contains evidence to the contrary. For μ, we use a standard distribution with SD of 10, but centered at -4.59, to encode our knowledge that child mortality is a rare event (approximately log(0.01) = -4.59). For σ_τ and σ_μ we use a zero-centered standard distribution with SD of 10, which allows for very large heterogeneity. The discussion of the Bayesian OR estimates throughout the paper refers to 95% posterior credible intervals (CrI) from Bayesian inference, which may not be symmetric.

Figure 3B uses a no pooling model, i.e. one where σ_τ^2 is infinite (assuming that individual studies do not influence each other). Following the literature (15), in the sensitivity analysis section (see below) model fit is compared across full pooling (fixed effects) and partial pooling (random effects) specifications using cross validation. Full pooling model is one where $\sigma_\tau^2=0$, i.e., there are no differences between studies. All other priors are unchanged across no, partial, and full models. For each specification, 15 Bayesian hierarchical models are fitted to the data, leaving out one study at a time and then calculating expected log predictive density (ELPD) for each study (42). This measures the out-of-sample predictive performance of the model for each study, automatically penalizing the model for the number of parameters. The ELPD averaged over the fifteen models is used as the cross-validation information criterion. A value closer to zero implies a better fit.

For the Bayesian model the weight of study k, w_k, is determined by the estimated between-study variance of effects, σ_τ^2, and the sampling variance of study k, se_k, as follows:

$$w_k = \frac{(se_k^2 + \sigma_\tau^2)^{-1}}{\sum_k (se_k^2 + \sigma_\tau^2)^{-1}}.$$

We report the meta-analysis weights in Table S5.

3. Sensitivity analysis

List of sensitivity analyses

To understand the impact of model choice on treatment effect estimates, we fitted: (i) a fixed effects (Bayesian logit) model instead of random effects, (ii) an inverse variance model instead of Peto OR model, and (iii) a risk difference model instead of OR.

For sensitivity to choice of data, we considered the following: (i) exclusion of any particular study from the analysis, (ii) combining two studies that measure impacts of a similar program on different populations (18, 33), (iii) the inclusion of studies with contaminated control groups (12, 31), (iv) the use of an alternative control group in a study with active and passive control arms (16), (v) use of an alternative treatment group in a spring protection study (36), (vi) restricting to studies with long monitoring durations, and (vii) dropping studies where water treatment was combined with another intervention (31, 32).

Summary of result

Over a set of the following 50 sensitivity analysis models (8 Peto OR models looking at lengths of follow-up, additional inverse variance model, Bayes and Peto models repeated for 15 choices of dropping one study and, 6 other choices of studies to include/exclude) the study estimates remain qualitatively similar to our main estimate. In this set of sensitivity analyses, the mean OR estimates range from 0.67 to 0.80.

Case-by-case details of sensitivity analyses

The exclusion of any single particular study. The Peto and Bayesian odds ratio estimates are given in Table S5. For the Bayesian model, the mean OR ranged from 0.67 to 0.80, with the lowest 95% CrI lower bound of 0.49 and the highest 95% CrI upper bound of 0.97. For Peto OR, the means ranged from 0.67 to 0.78, with the lowest lower bound of 0.42 and the highest upper bound of 1.01.

Combining studies that cover related programs. One of the studies (33) relies on a continuation[9] of a program from another study (18). The two studies cover different populations, time-periods, and interventions. In Null et. al (18), households have access to both chlorine

[9] In (33), the study sample includes 132 villages from two of the three counties (65 treatment villages and 67 passive control villages) of the original WASH-B study (18). The 65 treatment villages include villages which received free sodium hypochlorite dispensers for point-of-collection water treatment (which was continued by the NGO Evidence Action after the end of the WASH-B study) and dilute chlorine solution. Villages where water treatment was combined with sanitation, handwashing, or nutrition interventions in (18) were excluded from the sample. (33) uses data collected by John & Orkin (2018) four to five years after the rollout of the water treatment intervention on a sample of children born to mothers not enrolled in (18), over twice as large as that analyzed in the original study.

dispensers, and home delivery of water treatment solution. One study (33) uses data on non-treated individuals collected four to five years after the initial roll-out of the program studied in (18). In (33), home delivery of water treatment was discontinued, some chlorine dispensers had closed, and others had opened. We do not have data on how compliance varied between the two studies (33) measures the combined effect of the roll-out and continuation of the program. As a sensitivity test, we combine these into one study. The meta-analysis estimates remain quantitatively similar and significant with a mean reduction in mortality odds of 29-31%, depending on the model (see Table S6).

Including studies with a contaminated control group.

Two studies were not included in the main analysis due to contaminated control groups. In the blinded filtration study (36), the placebo filter removed more than 90% of the source water bacterial contaminant. Participants in the solar disinfection trial (37) were temporarily displaced due to political violence and following the displacement, most gathered water from standpipes with treated water—largely reducing the likelihood of source water contamination. Moreover, displacement could have affected adherence to solar disinfection practices. We report meta-analysis estimates including these two studies in a Table S6. Adding the solar disinfection trial (37) to the meta-analysis results in a mean reduction in mortality odds of 28-29%; adding the blinded filtration study (19) results in a mean reduction of 26-27% (see Table S6).

Alternate control group in study with active and passive arms. In one study (19), the experimental design included two control arms: an active one, receiving monthly visits by enumerators, and a passive one with no such visits. While in the original publication the authors restricted their analysis to treatment vs. active control comparisons, the present analysis combines data from the active and passive controls into a single control group to increase statistical power. Ignoring data from the passive control group (18) for the meta-analysis, leads to a mean reduction in mortality odds of 28-30% (see Table S6).

Alternate treatment group in spring protection study. In one study (29), the treatment effect from the water intervention was estimated using data from the study's treatment group, who received spring protection in Year 1, and the control group, who received spring protection in years 3 and 4. When those who receive spring protection in year 2 are included in the treatment group (29) for the meta-analysis, the estimated mean reduction in mortality ranges from 28-30% (see Table S6).

Dropping studies which combine water treatment with other interventions. Dropping studies where the water treatment intervention was combined with the provision of cookstoves (32) or other hygiene and sanitation interventions (31) leads to significant Peto OR and Bayesian OR estimates, with a mean reduction in mortality odds around 35% (Peto OR 0.66, Bayes OR 0.64), see Table S6.

Restricting to studies with longer monitoring periods. The studies included in the meta-analysis have differing lengths of follow-up, ranging from 9.5 to 260 weeks. Meta-analysis models of event data may overweight the contribution ("effort") of shorter studies. However, the weights in both the Peto OR and Bayesian logit model assigned to short studies are low, as seen in Table S5. For the Peto OR model, estimates are expected to be imprecise for studies with

shorter monitoring periods, owing to the shorter period over which events can occur. In our Bayesian logit model, when mortality event rates are low, this is reflected in the model by estimation of imprecise baseline risk and, hence, imprecise estimation of treatment effects, which ultimately leads to a lower weight in the meta-analysis estimate. As a sensitivity check, we repeat Peto OR analysis by excluding studies that are shorter than any given follow-up length in our dataset (104, 78, 65, 52, 37, 20, 13 and 9.5 weeks).

The results are plotted in Figure S2. For Peto OR, the mean reduction in mortality odds ranges from 19% (Peto OR 0.81: CI 95% 0.66, 0.98) to 28% (Peto OR 0.72: CI 95% 0.55, 0.92), and all estimates are significant.

We conducted an additional check of whether short studies may be unduly impacting the model. We started from 10 studies in the dataset that include one year or more of follow-up data and fit the Peto OR model. Then, we considered a hypothetical short study of 13 weeks (3 months), where the death risk is supposed to (crudely) approximate event rates in the dataset, 0.4%, and the size of the control arm is same as average size of control in the dataset, 1189. We assumed 1:1 randomization and that the true OR is the same as in the model of 10 long studies (0.80). We then simulated a growing number of short studies, 1, 2, 3, …, 10, in each case conducting 100 replications. We examined the behavior of mean and 95% intervals. Predictably, the mean was not affected and the intervals shrank only slightly: in the model of only 10 long studies the 95% interval was 66.0% to 97.2%. In the model with 10 long and 10 simulated short studies the 95% interval was 66.9% to 95.6% (averaged over 100 replications). This suggests that including short studies has a negligible impact on precision of the estimate, unless they have high event rates.

Inverse variance estimation. The inverse variance method assigns to each study a weight proportional to the inverse of the variance of the effect estimate. As a result, larger studies are given more weight than smaller studies, which have larger standard errors. To perform an inverse variance random effects estimation, we use a normal approximation of the log odds ratios and drop studies with zero deaths in either the treatment or control group: Chiller et al., 2006 (28) and Semenza et al., 1998 (26) report zero deaths in the treatment group; Luby et al., 2006 (19) reports zero deaths in the control group; Quick et al. (35) reports zero deaths overall. The results are presented in Figure S1. Inverse variance random effects estimation implies an average reduction in all-cause odds of child mortality of 26% with random effects (OR 0.74, 95% CI 0.59 to 0.93).

Below, we report results using only studies with published mortality outcomes to highlight the importance of collecting data from studies that did not report mortality outcomes as part of their analysis.

Contribution of studies with published mortality outcomes only. At the beginning of the study, five randomized controlled trials (RCTs) were identified which reported mortality outcomes as part of their analysis (18–22).[10] Two of the five studies did not pre-specify mortality as an outcome, yet reported large effects on mortality in their published manuscript (14, 27). By

[10] Papers which reported mortality in Clasen et al., 2015 (1) and other studies that we were aware of.

including studies which did not report mortality outcomes, we are able to increase statistical power to detect significant effects in our main results.[11] Another key reason is to avoid bias if those with positive point estimates are more likely to publish.

The estimated reduction in all-cause odds of child mortality with the Peto OR model was 33%. However, the result was not significant at the 95% confidence level (Peto OR 0.67: CI 95% 0.41, 1.11). The Bayesian logistic odds ratio estimate is similar in point estimate, and the uncertainty interval includes 1 (Bayes OR 0.74: CI 95% 0.28, 1.50). Dropping the study with zero deaths in its treatment arm (21) and using inverse variance OR results in significant estimates for only the fixed effect specification (random OR 0.65: CI 95% 0.40, 1.05; fixed OR 0.73: CI 95% 0.55, 0.97), see Table S4.

Risk difference (RD) model. As discussed, RD specification is not appropriate when there are very large differences in baseline risks, as is the case in this meta-analysis: one study had no events, and another had a 10% event rate in controls. A RD model is also not appropriate when probabilities are low, as this would imply that expected event rates for some studies are negative. This is also the case in our meta-analysis, since several studies had close to 0 events. However, we include results using an RD model for transparency. Fitting a Bayesian RD model we found a non-significant reduction in mortality risk of 0.2 percentage points (d = -0.002, 95% CrI -0.007, 0.001).

Fixed and random effects model. We used both full (fixed effect) and partial (random effect) pooling specifications for the Bayesian logit model. Under a fixed effect Bayesian logit model the reduction in odds was 24% (OR 0.76, 95% CrI 0.63, 0.91), compared to 30% under the random effects model. Using a leave-one-study-out cross-validation (LOO CV) procedure, the expected log predictive density (ELPD) for the partial pooling model was -956 (with SE of 268) and for the full pooling model -944(SE of 266). This suggests no significant differences in the out-of-sample performance of both models, with slight preference for the full pooling model.

[11] In January 2021, Waddington and Cairncross released a protocol for a meta-analysis of the effect of WASH interventions overall on mortality. They are planning to rely on mortality estimates in published manuscripts, similar to our analysis in this section.

4. Cost-effectiveness analysis

The cost-effectiveness analysis in this paper does not seek to determine the most cost-effective approach to water treatment, which may vary by context, but to simply use a few illustrative examples to argue to that there is likely to be tremendous potential to cost effectively reduce child mortality in a wide variety of settings in low-and middle income countries that do not already have access to safe water. To the extent that other delivery technologies can do so more effectively, benefits will be even greater. The present estimates imply that free provision of water treatment is a very cost-effective way to reduce child mortality.

We limit ourselves to a cost-effectiveness analysis and do not consider an earlier process that is used to determine whether regulatory approval should be given to water treatment, since water treatment has been widely used and has been generally accepted to be safe and effective against multiple pathogens.

Cost effectiveness results are given in the main text and Table 2 and more detailed calculations are given in Table S8 of this supplement.

In this section we provide more detail for two cost-effectiveness calculations presented in the paper and then discuss how decision maker's priors would impact the analysis.

Chlorine dispensers in western Kenya

The cost effectiveness of chlorine dispensers for point-of-collection water disinfection in western Kenya is calculated using data from Evidence Action, which operates 18,405 dispensers with 1,138, 964 people using the dispensers in western Kenya (43). Only benefits of reduced child mortality risk are included, while possible health gains through reduced child morbidity and health gains for people over the age of 5 years - such as those with suppressed immune systems (e.g., HIV+) - are ignored. The estimated cost of installing and maintaining chlorine dispensers at scale in western Kenya is about USD 9.13 per child under five served, per year (see Table S8, row 7).[12] Thus, the cost of operations is USD 3,104 per death of a child under 5 averted. Assuming that a death within the first 5 years of life leads to 81.25 disability-adjusted life years (DALYs)[13], the cost of chlorine dispensers for point-of-collection water disinfection is USD 39 per DALY averted (see Table S8, row 9). This cost is far lower than the Kenyan gross domestic product (GDP) per capita (about USD 1,878 in 2020), which is the threshold suggested by the Commission on Macroeconomics and Health to determine if interventions are "highly cost-effective" and, of course, even lower than three times the GDP per capita which is the threshold to determine if interventions are "cost-effective" (44).

Coupons for water treatment solution

[12] This is calculated as the ratio of the total cost of the program (serving all community members) and the number of children under 5 served by dispensers.

[13] As recommended by the World Health Organization.

Programs providing coupons for free water treatment solution to families with young children have so far only been conducted at modest scales (34, 45), but back of the envelope calculations suggest coupon programs would also be highly cost effective (see Table S8 Column 4). These calculations are based on rates of usage and coupon redemption in Kenya (45) and Malawi (34). Approximately 32% of all households who receive coupons treat their water with 37% of all coupons being redeemed.[14] The under-5 mortality rate among populations without access to safe drinking water is estimated as 5.02% using data from the UN Interagency Group for Child Mortality Estimation – slightly lower than that for rural Kenya used in the calculation above.[15] Based on the expected effect of water treatment on child mortality in a new implementation (Bayes OR 0.75; 95% CrI 0.29 to 1.46) and adjusting for usage rates, it is estimated that the program would reduce under-5 mortality per year by 0.7 p.p. (see Table S8, row 5). Each 150-milliliter bottle of WaterGuard (a brand of dilute chlorine solution) costs USD 0.31 and is enough for roughly one month's supply of treated water (for drinking and cooking) for a household. The studies that evaluated this intervention (34, 45) focused on environments where people did not have access to clean water. In scaling up such a coupon program, it may be difficult to exclude areas where much of the population already has access to clean water and it is possible that people in these areas might also treat their water. However, even considering that for every two households targeted the program covers an additional untargeted household which already has clean water, and that the administrative costs of running a coupon program were as large as the retail price of the chlorine solution, the cost of a scaled-up program would still only be USD 2,974 per death of a child under 5 averted – or USD 38 per DALY averted (see Table S8, rows 9).

Coupon programs could potentially be operated almost everywhere in the world, and rough calculations suggest that a global chlorine coupon program for all families with under-5 children without access to safe drinking water could avert up to half a million under-5 deaths each year (see Table S7). 2.2 billion people around the world do not have access to safely managed drinking water services (1), see row 1.[16] This number is similar in magnitude to the global estimates from other studies (46, 47). Of this population approximately 273 million are children under the age of 5 (2, 48).[17] The under-5 mortality rate among populations without access to safe drinking water implies 2.74 million deaths per year in absence of water treatment. Based on the expected effect of water treatment on child mortality in a new implementation (Bayes OR 0.75)

[14] Average across Dupas et al., 2016 (45) and Dupas et al., 2021 (34) who find water treatment rates of 34.5% and 30.0% and coupon redemption rates of 41.1% and 33.4%, respectively.

[15] The mean under 5 mortality rate is calculated across countries, weighted by the population without access to safe drinking water.

[16] Safely managed drinking water services are defined as improved sources of drinking water accessible on premises, available when needed and free from contamination. The "free from contamination" component of the indicator relies on data from household surveys and administrative data to estimate what proportion of users of improved sources drink water which does not contain fecal indicator bacteria (E. coli or thermotolerant coliform) and, where data is available, arsenic or fluoride.

[17] The mean under 5 population share is computed across countries weighted by population without access to safe drinking water.

and adjusting for coupon usage rates, it is estimated that a program that targets this population would save approximately 372,000 under-5 lives at a cost of approximately 1.1 billion USD each year.

Using alternative estimates that 1.8 billion people globally use a source of drinking water which suffers from fecal contamination (46), the total cost of the program would be USD 892.8 million and 304,000 under-5 lives would be saved annually. Alternatively, a coupon program targeting the 1.1 billion people estimated as only having access to drinking water that is of at least 'moderate' risk (>10 E. coli or triphenyltetrazolium chloride (TTC) per 100 ml) (46) would save approximately 186,000 lives at an annual cost of around a half billion dollars. If coverage of the program was restricted to the twenty countries[18] where Population Services International already markets hypochlorite solution at scale, approximately 95,000 deaths would be averted annually at a cost of around a quarter billion USD.

As noted, we include these estimates not to recommend these particular approaches to water treatment, as other approaches may be better suited to particular contexts. However, since this could be achievable virtually anywhere, it serves as a lower bound. Additionally, as the developing world becomes increasingly urban, our estimates potentially can be applied to improving access to clean water through piped water systems.

Alternative prior choices and detailed calculations

The main case we present (Table 2) uses the Bayesian model with diffuse prior. However, a decision maker may want to use informative priors (see "Cost-effectiveness" in Methods). As we discuss in the main text and in section 7 of this supplement, the effect we estimate is higher than what extrapolating from evidence on reductions in diarrhea would suggest, so informative expert-derived priors for effect are likely to be centered at a value lower than the effect estimated in this meta-analysis. Here, we do not attempt to define such informative priors, but rather provide two illustrative examples.

First, we define a prior centered on a 10% reduction in odds of mortality. We set SD on the mean effect (Bayesian hypermean) so that a roughly 30% reduction is two SDs away from the mean, i.e. log(OR) hypermean distributed according to Normal(log(0.9), log(0.9 - 0.7)/2). That means a priori there is a 2.5% chance for the reduction in odds of mortality to exceed 30% (and 20% chance that the odds ratio is greater than 1).

For heterogeneity we use an informative prior proposed by Turner et al. for binary outcomes, based on review of the Cochrane Database of Systematic Reviews (49), with hyperSD distributed as lognormal(-2.56/2, 1.74/2).

However, the normal prior on log(OR) is very restrictive in terms of tail behaviour: under this prior, the probability of reduction in odds exceeding 50% is 1 in a million and of 35% —less than 0.5%. A decision maker may want to instead use a strongly informative but less restrictive

[18] Zambia, Madagascar, Tanzania, Rwanda, Malawi, Kenya, Afghanistan, Burkina Faso, India, Uzbekistan, Myanmar, Mozambique, Nigeria, Uganda, Nepal, Vietnam, Ethiopia, Burundi, Guinea, and Cameroon.

prior for hypermean, such as a generalized Student T distribution with few degrees of freedom, which will have heavy tails. For illustration purposes we also fit a model where we replace Gaussian distribution with generalized Student's T with 1 degree of freedom, that is T(1)*log(0.9 - 0.7)/2 + log(0.9). This prior resembles the Gaussian distribution around the mean but has heavier tails: the probability of reduction in odds exceeding 30% is 15% and the probability of exceeding 50% reduction is 6.6%.

The detailed calculation and results for three different models (diffuse prior, informative prior with Gaussian hypermean, informative prior with Student T hypermean) are given in Table S8. We find that replacing diffuse prior with a strongly informative priors of 10% reduction in odds which lead to cost per DALY of $50 to $59 (depending on model and intervention), compared to $38-39 under diffuse prior. However, as discussed in the main text, even at $60 per DALY the intervention is highly cost-effective.

Our simple example also highlights that the model is sensitive not only to choice of location and scale parameters but also choice of family of distributions, with the choice of Student T instead of Gaussian leading to over 10% lower estimate of costs. Any decision maker looking to incorporate Bayesian priors must therefore be careful in deciding on how to incorporate various sources of uncertainty.

5. Publication bias

Publication bias on diarrhea outcome

Using the same dataset of 80 diarrhea outcomes as in section 1, but with diarrhea outcomes from several studies included in this meta-analysis added to the original dataset, for a total of 86 observations (see Materials and Methods section 1 just above for details). For simplicity, we assumed multiple observations from the same publication are independent. The outcome variable was risk ratio for diarrhea in children under-5, same as in (9).

We created funnel plots (Figure S11) and estimated Egger's tests for funnel plot asymmetry for all studies and studies that include chlorination interventions only. We rejected the hypothesis of asymmetry in both cases (p-value=0.782, p-value=0.341 respectively), which matches the result in Wolf et al., who also found no evidence of funnel plot asymmetry across water interventions (p-value=0.8).

We also used Andrews and Kasy's publication bias correction technique (16) on the joint dataset. The table below shows estimates obtained by assuming symmetric publication bias cut-off around z = |1.96| and the meta-study replication method. For this result we included all data points. The distribution of intervention effects, adjusted for publication bias and assumed to be normal, has a (hyper)mean of -0.40 (SE = 0.06) and (hyper)SD of 0.32 (SE = 0.05). The relative probability of publication (between studies with |z| less/more than 1.96), beta_p, is 1.009. That is, insignificant results are 0.9% more likely to be published relative to significant results with a standard error of 0.379. Repeating the same method for the subset of studies that considered

chlorination (lower panel of table below), the results are noisier. We cannot reject the hypothesis that relative publication probability is equal to 1.

Funnel plots, Egger's test, and Andrews and Kasy's test all seem to indicate there is little evidence of publication bias in papers investigating interventions targeting diarrhea.

Results for the Andrews and Kasy publication bias adjustment method:

	hypermean (theta)	hyper-SD (tau)	beta_p
All Interventions			
mean	-0.404	0.315	1.009
SE	(0.060)	(0.047)	(0.379)
Chlorination Studies Only			
mean	-0.472	0.331	2.249
SE	(0.116)	(0.058)	(1.557)

Relationship between mortality and diarrhea

As mentioned, point estimates and z-values for diarrhea outcome are available for 86 observations, including 14 out of 15 studies with mortality data. We tested (using a logistic model) if availability of mortality data in a given study depends on (1) point estimate of diarrhea effect in that study, (2) absolute z-value of the diarrhea effect exceeding 1.96. In both cases the diarrhea effect was measured as log(RR). We found no evidence of either association, with p-values of 0.34 and 0.19 respectively.

Publication bias on mortality outcome

We report a funnel plot (Figure S3) and conduct Egger's and Begg's tests for the funnel plot asymmetry. In both the tests we are not able to reject the null hypothesis of symmetrical funnel (p-value=0.45 and p-value=0.78 respectively).

We also use meta-study publication bias adjustment methods from Andrews and Kasy, 2019 (same as in the case of diarrhea outcome above). We used Peto log odds ratios as inputs into the random effects model, assuming that probability of publication changes when $|z| > 1.96$. As discussed in the main text, this ignores potential bias in diarrhea outcomes and treats mortality results as the sole factor on which publication decisions are made. However, since there are only

two studies outside of the funnel (see Figure S3), the relative probability of publication (of insignificant to significant results) is not precisely estimated (mean of 2.3 with SE = 2.1). The mean of the true treatment effect across 15 studies (adjusted for publication bias) is not meaningfully different from our estimate, OR = 0.63 (logOR = -0.46, SE = 0.26). Mean hyper-SD (heterogeneity) parameter across studies is 0.3 (SE = 0.2). (16)

Exploratory simulation of small-study publication bias

As an exploratory assessment, we simulated additional unpublished studies to better understand the potential impact of publication bias. Since studies with few events might be less likely to report on mortality, we simulated studies with a low mortality risk of 0.4% (equivalent to 3 months of follow-up on average in our dataset). For simplicity we assumed that all simulated studies had true OR of 1 (a strong assumption, given our strong prior of non-negative effects based on water treatment literature), a per-arm sample size of 1189 (the average across 15 studies included in our dataset). We added the simulated unpublished studies to the original dataset of 15 studies and fit all data using the default Peto OR model. We calculated averages over 250 replications.

With 5 additional studies the estimated reduction in odds was 24%. With 15 additional unpublished simulated studies with a true OR of 1 (i.e. 15 real studies with OR of 0.72 + 15 simulated studies with OR of 1), the meta-analysis estimate had a mean of 0.81, with 95% interval of 0.67 to 0.99. Given our search strategy, which included directly contacting researchers, we find it unlikely that so many studies could be missed. We also find it unlikely that the effect of publication bias is so strong that all missed studies would have an OR of 1. However, this assessment does not cover the scenario where studies with large numbers of deaths were missed.

In the future version of the paper we will conduct additional simulations, based on a more realistic sample of studies which measured diarrhea but not mortality outcomes. This will allow us to obtain realistic sample sizes and follow-up durations.

6. Exploratory assessment of power to detect heterogeneous effects

As shown in Fig S5-S9, univariate meta-regressions do not find statistically significant linear relationships between five predictors and treatment effect estimated using the frequentist model. However, given small sample size and uncertain estimates in individual studies, a meta-regression model that would typically be used in such situations may not have sufficient power to detect linear relationships between the predictors and the treatment effects. To assess this, we conducted a simple post-hoc exploratory analysis of whether a meta-regression model would have sufficient power to detect the relationship between treatment effects and three continuous predictors: prevalence of diarrhea, compliance, and year of implementation. The linear relationship between the first two is easiest to hypothesize, since at x=0 (no compliance, no prevalence), we would expect the true effect to be 0; we also investigate year of implementation as it is of practical importance to policy makers.

Let us assume there is a strict linear relationship between y = log(OR) and x, which will denote compliance, prevalence, or year of implementation. We parameterise these such that the expected average effect in the population corresponds to the estimated mean OR. That is, in the case of prevalence we set slope ($y=ax$) to a = -0.335*x/0.183, where -0.335 = log(0.7149), i.e. the logarithm of the OR estimated by the Peto OR model; 0.183 is population-weighted prevalence across 15 studies. In the case of compliance we set it to a = -0.335*x/0.462, where 0.462 is the population-weighted compliance across 14 studies (see Table 1); one study did not report compliance. In the case of year of implementation we set y = (0.0335/2)*(x-2010) - 0.335, that is, we assume that in 2010 (weighted average of year of implementation in 15 studies) the mean effect was log of 0.7149 and it decreased linearly to the point where by 2020 half of the effect disappears, which we would consider to be a very strong effect.

We simulate new datasets with some noise ($y = ax + e$), using observed compliance/prevalence values for each x (for compliance we impute the one missing value as mean) and for noise e using posterior SDs for 15 studies from the main Bayesian model, to obtain a crude but realistic estimate of variation in each study. For each simulated dataset we fit a univariate linear regression model and check if the coefficient is significant. We repeat this 10,000 times. Simulated power is the fraction of coefficients that were significant. We find it to be 43% for compliance, 51% for diarrhea prevalence, and 29% for year of implementation. While the simulated power results will vary a lot depending on assumptions, our calculation should already be treated as optimistic with regards to power, since we assumed no confounding and a strictly linear relationship. This suggests our data are insufficient to detect the relationship between compliance, year of implementation, or prevalence and mortality, even under the assumption of strong effects.

7. Comparing meta-analysis estimates with model predictions

The point estimate of the mortality effect from the meta-analysis is much larger than the point estimate predicted by a simple model in which diarrheal deaths are taken from the central estimate of the Global Burden of Disease (GBD) project (2), the effect of water treatment on diarrhea is taken from the central estimate in the Clasen meta-analysis (8), and mortality is assumed to be linear in diarrhea cases, so reductions in diarrhea deaths are proportional to reductions in diarrheal cases. However, the differences are small enough that they could fairly easily be accounted for by known epidemiological factors not captured by this simple linear model and sampling variation

A recent Cochrane review (8) found a reduction in under-five diarrhea due to water quality interventions of 39% (CI 95% 25%, 51%).[19] Under a simple model in which deaths are

[19] This confidence interval reflects sampling variation only, but estimated effects of water treatment on caregiver-reported diarrhea may also be subject to reporting bias (8, 9).

approximated as linear in cases and cases are estimated as linear in treatment rates, multiplying the central GBD estimate of the proportion of under-5 deaths attributable to diarrhea of 9.9% (CI 95% 8.2%, 11.6%) times the central Cochrane estimate of 39% gives a predicted 3.9% mean reduction in child mortality from water treatment. If we interpret the two CIs above as Bayesian intervals, the 95% interval on this estimate is 2.6% to 5.5%. In contrast, the meta-analysis in this paper gives a central estimate of an approximately 30% reduction in the odds of all-cause child mortality.

One likely reason for differences between the predictions of a simple linear model and the meta-analysis findings is that several scientifically plausible pathways through which water treatment could reduce mortality are not captured by the linear model.

First, water treatment could reduce both the mortality rate and the incidence of diseases other than diarrhea (Mills-Reincke phenomenon) (50, 51). Epidemiological studies lend support to this hypothesis, showing that diarrheal episodes are followed by increased risk of acute lower respiratory tract infection among children in Ghana, Nepal, India, Pakistan and Israel (52–55). Continued exposure to diarrheal pathogens alters the gut microbiome, increasing susceptibility to infection (56). Such subclinical or clinical episodes of infection can induce impairments in gut function and undernutrition phenotypes leading to increased mortality (57, 58) . Relatedly, diarrhea can lead to malnutrition (59, 60), which in turn can put a child at risk for higher mortality from a range of illnesses, or simply death from malnutrition itself. The Global Burden of Disease uses a “one death one cause” methodology, which allows it to estimate all causes of deaths without double counting. However, it could under-estimate the mortality effect of addressing a given disease in scenarios like this, where morbidity from multiple diseases combines to cause a death.

Second, water treatment could prevent diseases which can cause life-threatening illness in the absence of diarrhea. It could reduce worm loads; kill enteroviruses, *Salmonella Typhi* and *Salmonella Paratyphi*, and prevent hepatitis A and hepatitis E. Water treatment could also prevent deaths from sepsis among infants by facilitating cleaner births and postnatal care practices (61). Poor water quality and exposure to a more pathogenic environment is associated with preterm birth and low birth weight (62).

Third, water treatment could potentially have a larger effect on severe diarrhea than on overall diarrhea. This is the case for some other interventions. For example there is evidence that RRV-TV rotavirus vaccines lead to greater reductions in severe diarrhea episodes than in mild ones (10, 11). Another example is COVID-19 vaccines, many of which have been far more effective against hospitalization and death than infection.

Fourth, the GBD estimates of the diarrheal death rate are limited by data availability, requiring modeling to fill data gaps, and according to the authors, many datasets have biases or errors,

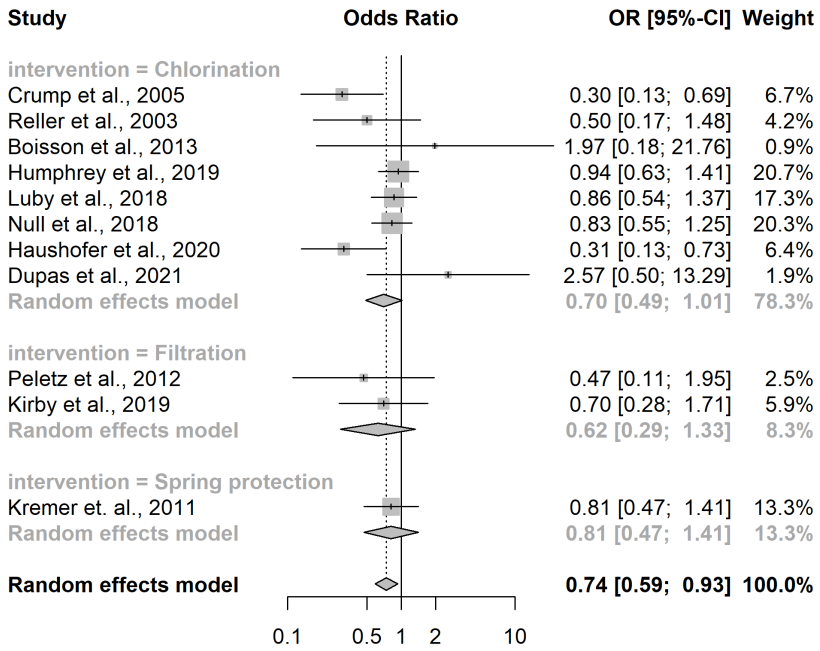
such as the misclassification of causes of death or assignment of deaths to causes that cannot be primary causes of death (63). For example, estimated effects of water treatment on caregiver-reported diarrhea may be subject to reporting bias (11, 12).

There is also uncertainty in our estimates and in the estimates from the Clasen et al. meta-analysis, although the portion of uncertainty due to sampling variation in these is more easily quantified.

As discussed above, the linear model likely underestimates the impact of interventions to improve water quality on child mortality. However, even under the predictions from this model, chlorine dispensers would remain cost-effective according to the 1x GDP "highly cost-effective" threshold. Thus, independent of the tightness of their priors, an analyst who starts with priors based on the linear model and updates based on evidence from this meta-analysis would conclude that water treatment is also cost-effective.

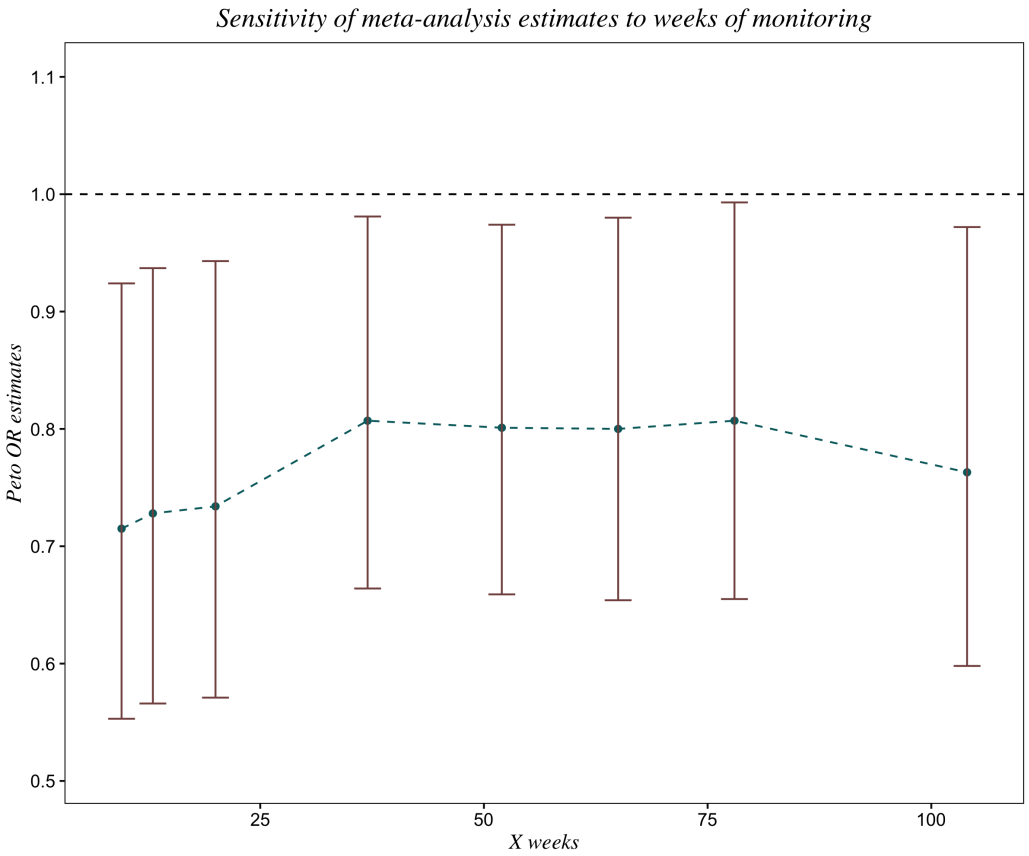
Supplementary Figures

Fig. S1. Random-effects forest plot of child mortality estimates of the impact of water quality interventions (Odds Ratios Inverse Variance)



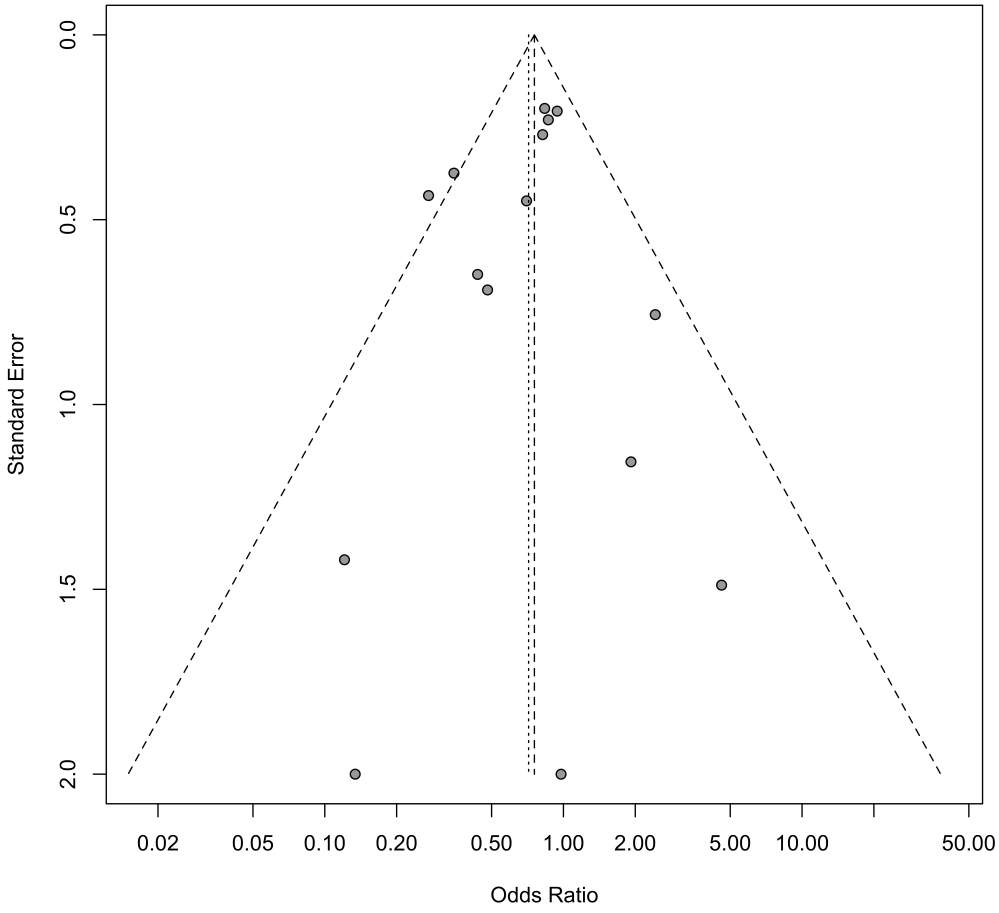
Notes: Dots and horizontal lines represent point estimates and 95% confidence intervals-from individual studies, respectively. The area of the square around each dot represents the weight given to each study in the fixed-effects estimation. Diamonds are centered around the random-effects estimate (by intervention type or overall), their widths indicate the 95% confidence interval. 4 studies are dropped due to zero deaths in either the treatment or control group.

Fig. S2. Restricting set of studies to longer follow-up lengths



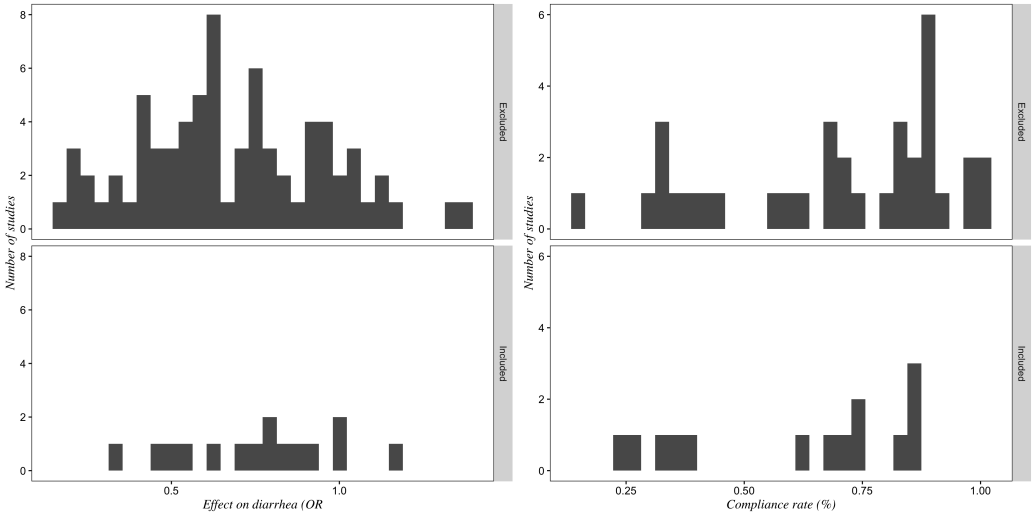
Notes: This figure presents the odds ratio estimated by the frequentist (Peto) meta-analysis model with studies shorter than X weeks removed. Each point is the Peto OR estimate, and the bars represent the 95% Confidence Interval for each estimate. All 15 studies in the main sample are included for X = 9.5 weeks, and 4 studies are included for X = 104 weeks (2 years).

Fig. S3. Funnel plot to examine publication bias



Notes: This figure presents a funnel plot. Symmetry on either side of the vertical line (representing the overall effect) suggests that publication bias is not present. Results for Egger's and Begg's test are reported in Materials and Methods, Section 5.

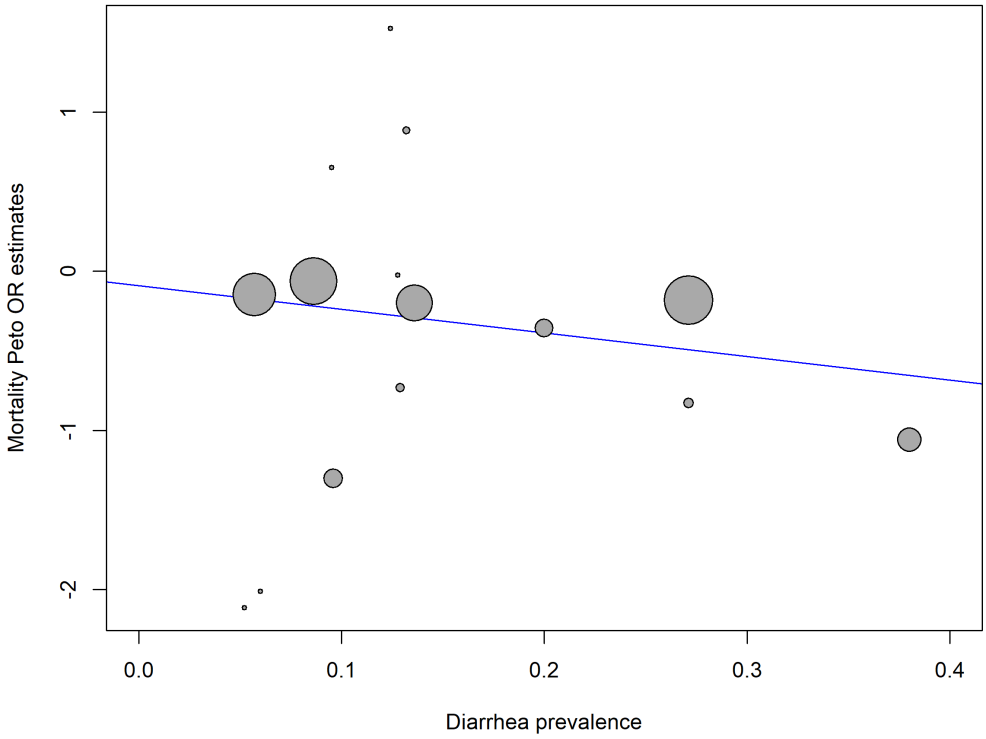
Fig. S4. Diarrhea effect estimates and compliance rates across included and excluded studies



(A) (B)

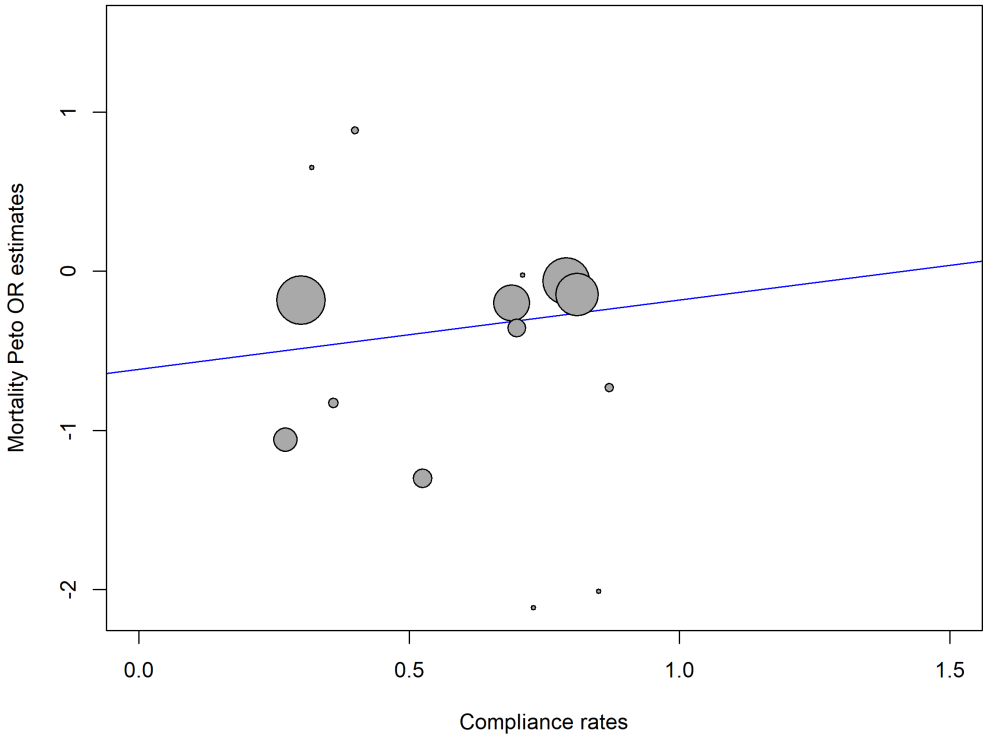
Notes: Figure (A) presents the diarrhea effect size across included (bottom panel) and excluded (top panel) studies. Figure (B) presents the compliance rate across included (bottom panel) and excluded (top panel) studies.

Fig. S5. Heterogeneity in treatment effects, by diarrhea prevalence



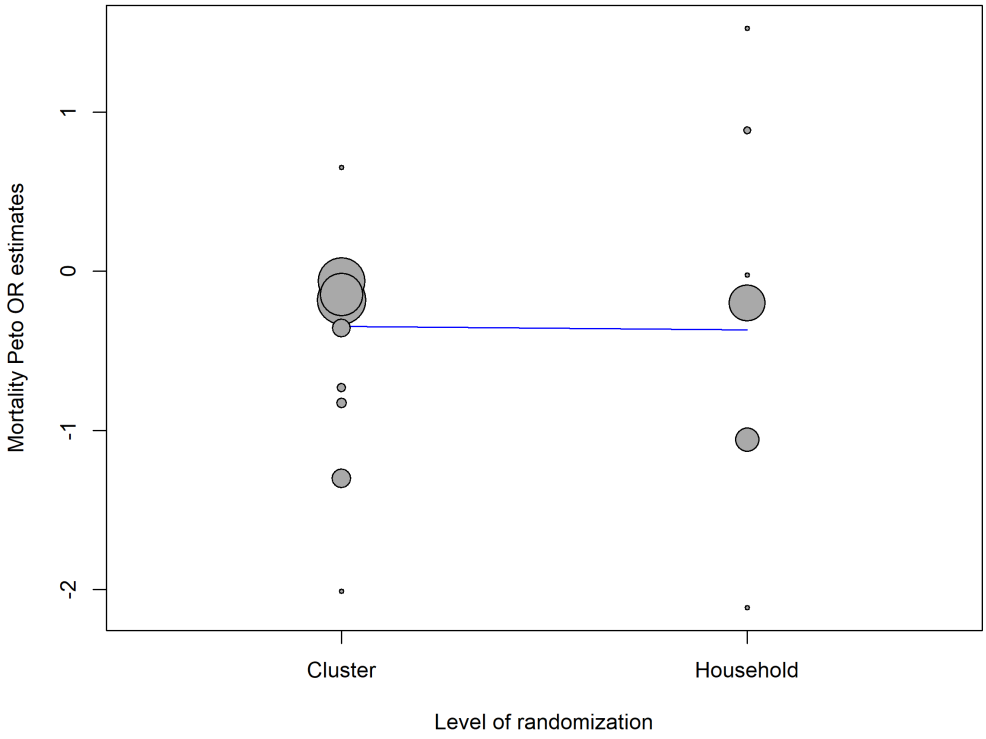
Notes: This figure presents the relationship between child mortality Peto odds ratio estimates and the level of diarrhea prevalence across 15 studies in the sample. We find no significant differences (slope of -1.483 per unit increase in diarrhea prevalence rate, pval 0.252) in effect estimates by prevalence of diarrhea. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study.

Fig. S6. Heterogeneity in treatment effects, by compliance rate



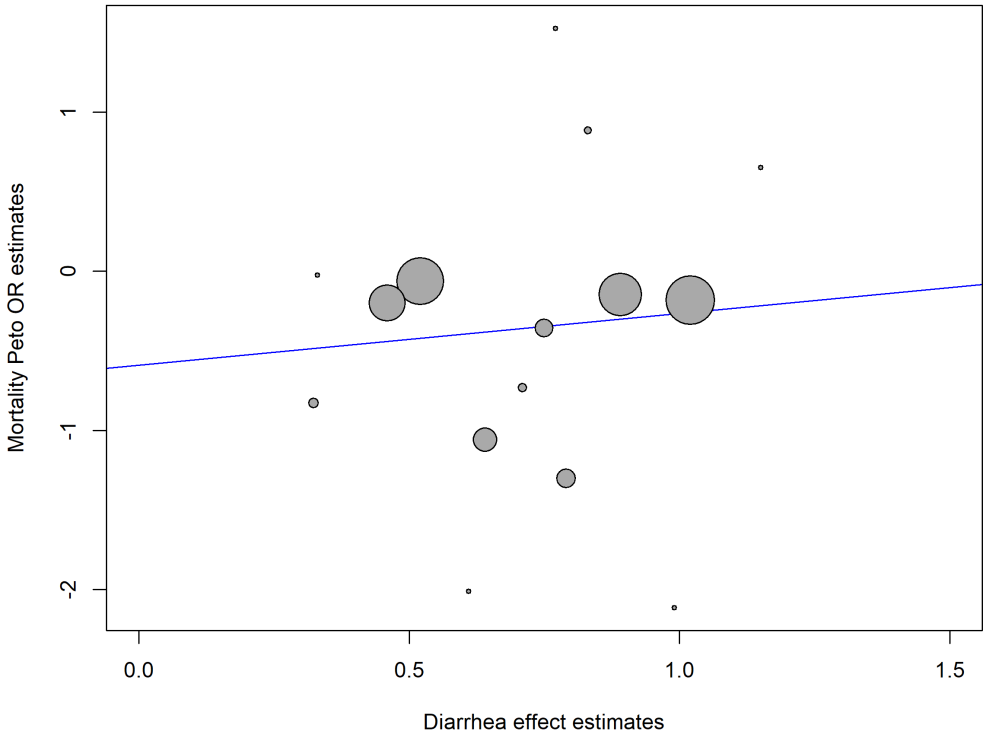
Notes: This figure presents the relationship between child mortality Peto odds ratio estimates and the level of compliance across 14 studies in the sample (one study (19) did not report any measure of compliance). We find no significant differences (slope of 0.439 per unit increase in compliance rate, pval 0.488) in effect estimates by the level of compliance. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study.

Fig. S7. Heterogeneity in treatment effects, by unit of randomization



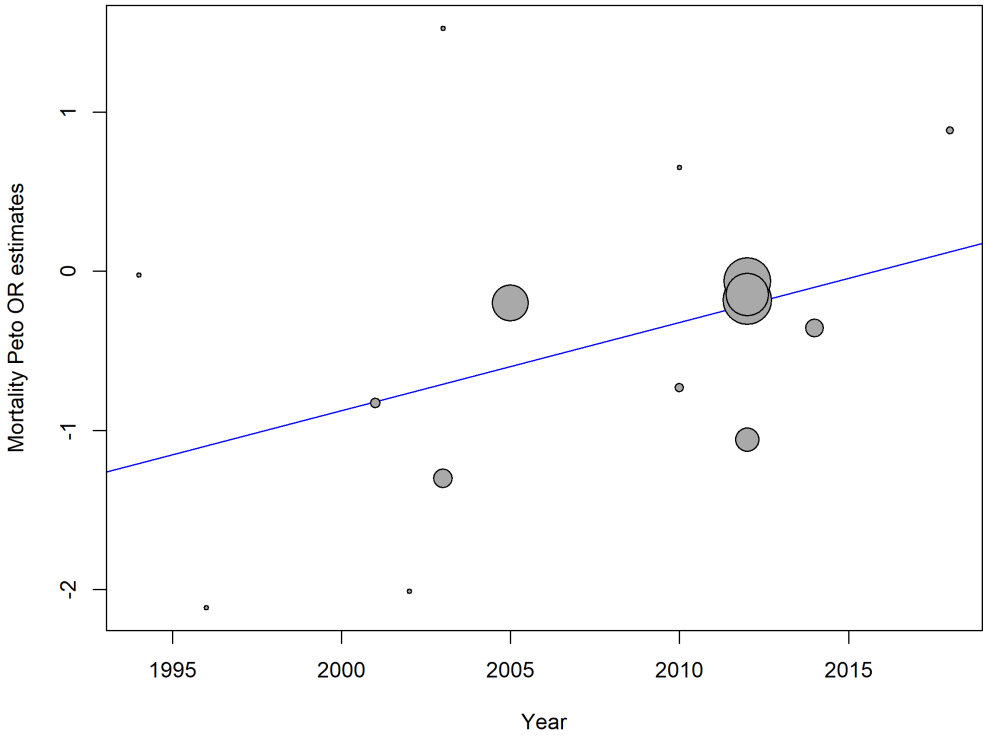
Notes: This figure presents the relationship between child mortality Peto odds ratio estimates and the unit of randomization across 15 studies in the sample. We find no significant differences (Decrease of -0.023 for randomizing at the household level, p-value = 0.944) in effect estimates by the unit of randomization. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study.

Fig. S8. Heterogeneity in treatment effects, by diarrhea effect estimates



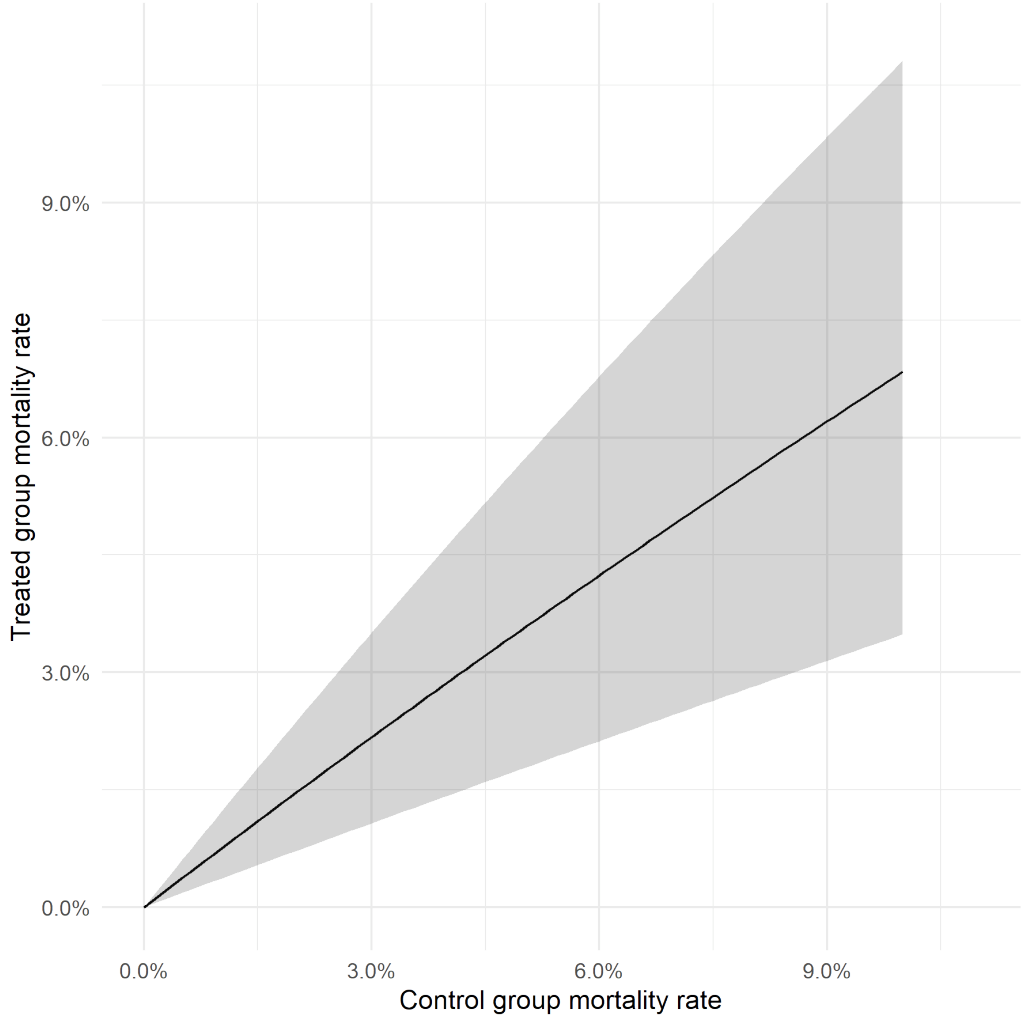
Notes: This figure presents the relationship between child mortality Peto odds ratio estimates and the diarrhea effect estimates across 15 studies in the sample. We find no significant association (slope of 0.324 per unit increase in the diarrhea effect estimate, p-value = 0.644) between mortality and diarrhea effect estimates. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study.

Fig. S9. Heterogeneity in treatment effects, by study year



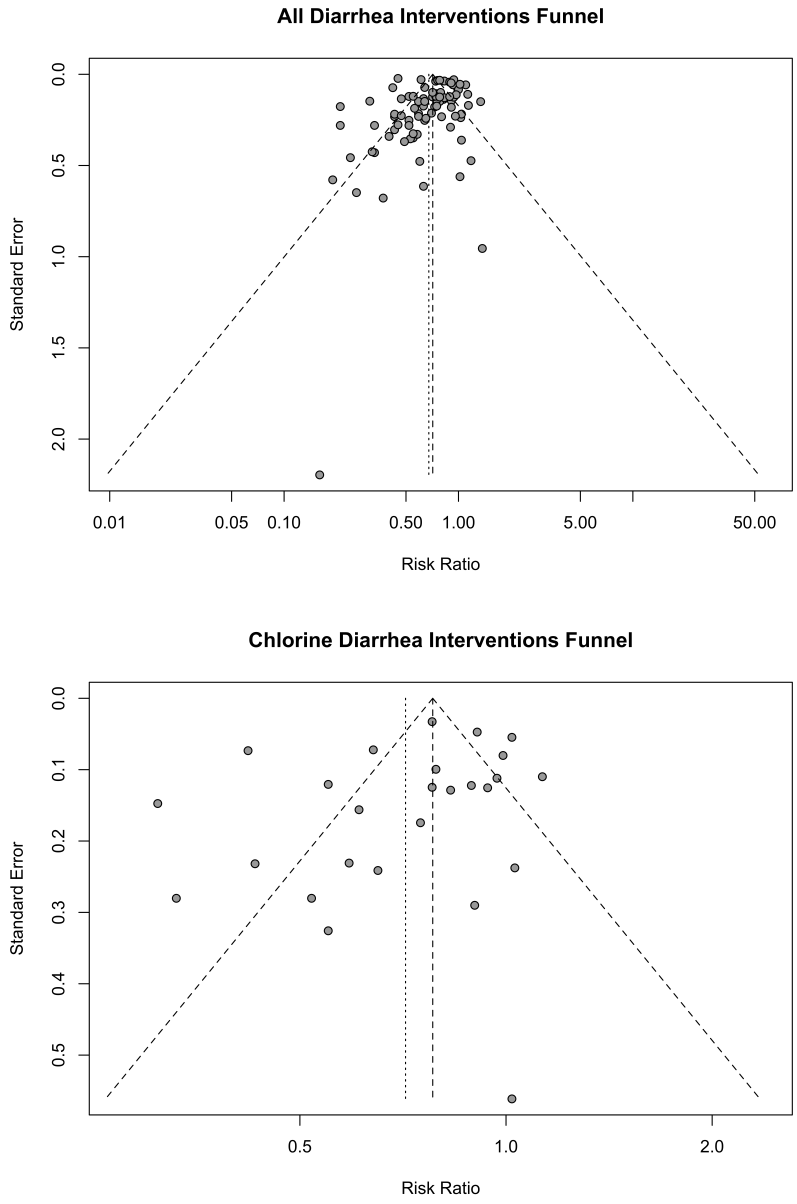
Notes: This figure presents the relationship between child mortality Peto odds ratio estimates and the study year across 15 studies in the sample. Year of Intervention is the year the study's intervention was launched. We find no significant association (slope of 0.055 per year, p-value = 0.056) between mortality and study year. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study.

Fig. S10. Predicted mortality among treated as a function of control group mortality



Notes: The relationship between a given mortality rate in control group, p (x axis) and mortality among treated (y axis) is, for a given OR, [OR*p/(1-p)]/[1 + OR*p/(1-p)]. For small values of p the relationship is nearly linear, as seen in the figure. We use the posterior predictive distribution (of ORs in a new setting) from the Bayesian model to construct mean (solid line) and 90% interval (shaded area) for mortality among treated.

Fig. S11. Funnel plot for all diarrhea interventions and chlorine diarrhea intervntions



Notes: Funnel plot to assess publication bias in risk ratio estimates of diarrhea morbidity in all augmented WaSH studies and chlorination WaSH studies.

Supplementary Tables

Table S1. Search strategy and search terms

Search set	Embase (Ovid)	Pubmed	Scopus	Cochrane Library
Water Quality				
1	((Water adj3 (treatment or quality or cleaning or purif* or chlorin* or decontamination or filt* or disinfect* or floccul* or storage or recontamination or re-contamination)).mp. or exp water quality/ or exp water management/) and ((water.mp. or exp water/) adj3 (drinking or consumption).mp)	((treatment[tw] OR quality[tw] OR cleaning[tw] OR purif*[tw] OR chlorin*[tw] OR decontamination[tw] OR filt*[tw] OR disinfect*[tw] OR floccul*[tw] OR storage[tw] OR recontamination[tw] OR "re-contamination"[tw]) OR "Water Quality"[MeSH] OR "Water Purification"[MeSH]) AND ((water[tw] OR water[MeSH]) AND (drinking[tw] OR consumption[tw]))	TITLE-ABS-KEY (water W/3 (treatment OR quality OR cleaning OR purif* OR chlorin* OR decontamination OR filt* OR disinfect* OR floccul* OR storage OR recontamination OR "re-contamination")) AND (TITLE-ABS-KEY (water W/3 (drinking OR consumption)))	((water near/3 (treatment or quality or cleaning or purif* or chlorin* or decontamination or filt* or disinfect* or floccul* or storage or recontamination or "re-contamination")):ti,ab,kw or MeSH descriptor: [Water] explode all trees or MeSH descriptor: [Water Quality] explode all trees or MeSH descriptor: [Water Purification] explode all trees) and ((Drinking or consumption) near/3 water):ti,ab,kw
Water Access				
2	(Water adj3 (supply or availability or access or connect* or distance or improved or distribut* or quantity or volume)).mp or exp water supply/	(water[tw] AND (supply[tw] OR availability[tw] OR access[tw] OR connect*[tw] OR distance[tw] OR improved[tw] OR distribut*[tw] OR quantity[tw] OR volume[tw])) OR "Water Supply"[MeSH]	TITLE-ABS-KEY (water W/3 (supply OR availability OR access OR connect* OR distance OR improved OR distribut* OR quantity OR volume))	(Water near/3 (supply or availability or access or connect* or distance or improved or distribut* or quantity or volume)):ti,ab,kw or MeSH descriptor: [Water Supply] explode all trees

Sanitation				
3	toilet*.mp. or latrine*.mp. or pit.mp. or pits.mp. or sanita*.mp. or ecosan.mp. or sewage.mp. or sewer$1.mp. or sewerage.mp. or exp sewage/ or open defecation.mp or (((feces or faeces or fecal or faecal or excre* or waste).mp. or exp feces/) adj3 (disposal or manag* or service*).mp.) or exp sanitation/ or exp environmental sanitation/	toilet*[tw] OR latrine*[tw] OR pit[tw] OR pits[tw] OR sanita*[tw] OR ecosan[tw] OR feces[tw] OR faeces[tw] OR fecal[tw] OR faecal[tw] OR excre*[tw] OR "waste disposal"[tw] OR "disposal of waste"[tw] OR "waste management"[tw] OR "management of waste"[tw] OR sewage[tw] OR sewer*[tw] OR sewerage[tw] OR "open defecation"[tw] OR "Toilet Facilities"[MeSH] OR "Toilet Training"[MeSH] OR Sanitation[MeSH] OR Feces[MeSH] OR Sewage[MeSH]	TITLE-ABS-KEY (toilet* OR latrine* OR pit OR pits OR sanita* OR ecosan OR sewage OR sewer* OR sewerage OR "open defecation") OR (TITLE-ABS-KEY ((feces OR faeces OR fecal OR faecal OR excre* OR waste) W/3 (disposal OR manag* OR service*)))	(toilet* or latrine* or pit or pits or Sanita* or ecosan or sewage or sewer* or sewerage or open defecation or ((feces or faeces or fecal or faecal or excre* or waste) near/3 (disposal or manag* or service*))):ti,ab,kw or MeSH descriptor: [Toilet Facilities] explode all trees or MeSH descriptor: [Toilet Training] explode all trees or MeSH descriptor: [Sanitation] explode all trees or MeSH descriptor: [Feces] explode all trees or MeSH descriptor: [Sewage] explode all trees
Diarrhoeal disease				
4	(((f?ecal adj1 coliform$1) or bacterial or microbiological or viral or diarrh?ea? or intestinal or enteric or gastro-enteric or protozoa$1 or waterborne or water-borne or enterovirus or "enteric virus" or poliovirus or rotavirus or norovirus or "norwalk-like virus" or hepatitis or campylobacter or helicobacter or legionellos$ or vibrio or cholera or escherichia or salmonell$ or shigell$ or cryptosporidi$).mp. or exp	("fecal coliform"[tw] OR "fecal coliforms"[tw] OR "faecal coliform"[tw] OR "faecal coliforms"[tw] OR bacterial[tw] OR microbiological[tw] OR viral[tw] OR diarrhoea*[tw] OR diarrhea*[tw] OR intestinal[tw] OR enteric[tw] OR "gastro-enteric"[tw] OR protozoa*[tw] OR waterborne[tw] OR "water-borne"[tw] OR Diarrhea[MeSH] OR enterovirus[tw] OR "enteric virus"[tw] OR poliovirus[tw] OR	TITLE-ABS-KEY (((fecal OR faecal) PRE/1 coliform*) OR bacterial OR microbiological OR viral OR diarrhoea* OR diarrhea* OR intestinal OR enteric OR "gastro-enteric" OR protozoa* OR waterborne OR "water-borne" OR enterovirus OR "enteric virus" OR poliovirus OR rotavirus OR norovirus OR "norwalk-like virus" OR hepatitis OR campylobacter OR	((((fecal or faecal) next coliform*) or bacterial or microbiological or viral or diarrhoea* or diarrhea* or intestinal or enteric or gastro-enteric or protozoa* or waterborne or water-borne or enterovirus or enteric virus or poliovirus or rotavirus or norovirus or norwalk-like virus or hepatitis or campylobacter or helicobacter or legionellos* or vibrio or cholera or escherichia or salmonell* or shigell* or cryptosporidi*):ti,ab,kw or MeSH descriptor: [Diarrhea] explode all

	diarrhea/) and (disease$1 or infection$1 or episode$1 or illness$2).mp	rotavirus[tw] OR norovirus[tw] OR "norwalk-like virus"[tw] OR hepatitis[tw] OR campylobacter[tw] OR helicobacter[tw] OR legionellos*[tw] OR vibrio[tw] OR cholera[tw] OR escherichia[tw] OR salmonell*[tw] OR shigell*[tw] OR cryptosporidi*[tw]) AND (disease*[tw] OR infection*[tw] OR episode*[tw] OR illness*[tw])	helicobacter OR legionellos* OR vibrio OR cholera OR escherichia OR salmonell* OR shigell* OR cryptosporidi*) AND (TITLE-ABS-KEY (disease* OR infection* OR episode* OR illness*))	trees) and (disease* or infection* or episode* or illness*):ti,ab,kw
Epidemiological study				
5	(prevalence or incidence or risk or exposure or exposed or outcome or epidemiology or epidemiological or impact or effect or evaluation or odds).mp	prevalence[tw] OR incidence[tw] OR risk[tw] OR exposure[tw] OR exposed[tw] OR outcome[tw] OR epidemiology[tw] OR epidemiological[tw] OR impact[tw] OR effect[tw] OR evaluation[tw] OR odds[tw]	TITLE-ABS-KEY (prevalence OR incidence OR risk OR exposure OR exposed OR outcome OR epidemiology OR epidemiological OR impact OR effect OR evaluation OR odds)	(prevalence or incidence or risk or exposure or exposed or outcome or epidemiology or epidemiological or impact or effect or evaluation or odds):ti,ab,kw
Limits				
6	Limit to (humans and (english or french) and yr="2012 -Current")	("2012/01/01"[PDat] : "2016/02/05"[PDat]) AND Humans[Mesh] AND (English[lang] OR French[lang])	LIMIT-TO (LANGUAGE , "English") OR LIMIT-TO (LANGUAGE , "French")) AND (LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR ,	Publication Year from 2012 to 2016

		2013) OR LIMIT-TO (PUBYEAR , 2012)	
Search for Water Quality, Water Access, Sanitation and Diarrhoeal Diseases			
(1 or 2 or 3) and 4 and 5 and 6	(1 OR 2 OR 3) AND 4 AND 5 AND 6	(1 OR 2 OR 3) AND 4 AND 5 AND 6	(1 or 2 or 3) and 4 and 5 and 6

Table S2. Excluded studies

Reason for Exclusion	Studies
Not a developing country	Colford (2002) (64), Colford (2005) (65), Rodrigo (2011) (66)
Not a randomized control trial	Kirchhoff (1985) (67), Alam (1989) (68), Mahfouz (1995) (69), Conroy (1996) (70), Xiao (1997) (71), Quick (2002) (72), Jensen (2003) (73), Majuru (2011) (74), Johri et al. (2019) (75), Reese et al. (2019) (76)
Does not include children under 5 years in age	Abebe (2014) (77)
Authors responded but no mortality data collected	Gruber (2013) (78), Günther (2013) (79), Jain (2010) (80), Opryszko (2010a, b, c) (44), Patel (2012) (81), Roberts (2001) (82), Tiwari (2009) (83)), URL (1995a, b) (84), Boisson (2009) (85), Doocy (2006) (86), Stauber (2009, 2012a, b) (87–89), Lindquist (2014a, b) (90) , Fabiszewski (2012) (91), Clasen (2004b, c) (92, 93), Pickering et al. (2019) (94), Handzel (1998) (95),
Authors responded and mortality data was collected but no longer available	Gasana (2002) (96), Brown (2008) (97)
Authors did not respond	Torun (1982)* , Austin (1993a,b) (98), Mengistie (2013) (99), McGaugan (2011) , Mäusezhal (2009) (100), Lule (2005) (101), du Preez (2008, 2010) (102)

Note: *The only author died.

Table S3. Numbers of events (deaths) and non-events in treatment and control groups

Study	Treatment group		Control group	
	Events	Non-events	Events	Non-events
A. Main sample				
Semenza et al., 1998 (26)	0	88	2	78
Reller et al., 2003 (27)	10	729	5	182
Crump et al., 2005 (21)	9	1009	15	505
Luby et al., 2006 (19)	2	1013	0	553
Chiller et al., 2006 (28)	0	132	1	137
Kremer et. al., 2011 (29)	18	691	47	1465
Peletz et al., 2012 (22)	3	58	6	54
Boisson et al., 2013 (30)	2	1505	1	1507
Null et al., 2018 (18)	30	858	114	2697
Luby et al., 2018 (20)	27	629	62	1244
Humphrey et al., 2019 (31)	49	946	50	909
Kirby et al., 2019 (32)	8	1198	12	1252
Haushofer et al. 2021 (33)	7	987	22	965
Dupas et al. 2021 (34)	5	1288	2	1321
Quick et al. 1999 (35)	0	400	0	391
Total	170	11531	339	13260
B. Studies included for robustness checks				
Boisson et al., 2010 (36)	4	81	1	104
du Preez et al., 2011(37)	3	355	3	334

Table S4. The child mortality impact of water quality interventions – initial studies which report child mortality effects

	Crump et al., 2005 (21)	Luby et al., 2006 (19)	Peletz et al., 2012 (22)	Luby et al., 2018 (20)	Null et al., 2018 (18)	IV Random-Effects OR	IV Fixed-Effects OR	Mean Bayesian/Peto OR (5 studies)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Bayesian OR								
ITT effect on child mortality	0.54		0.67	0.78	0.77	0.65	0.73	0.74
CrI 95%	(0.54,0.90)		(0.67,1.24)	(0.78,1.15)	(0.77,1.09)	(0.40,1.05)	(0.55,0.97)	(0.28,1.50)
Panel B: Peto OR								
ITT effect on child mortality	0.27	4.60	0.86	0.83	0.48			0.67
CI 95%	(0.12,0.64)	(0.25,85.10)	(0.55,1.36)	(0.56,1.34)	(0.12,1.86)			(0.41,1.11)
Obs.	1538	1548	121	1962	3699			

Notes: The studies (Columns 1-5) included are studies which were identified reporting mortality outcomes at the beginning of the meta-analysis. Panel A Columns 1 to 5 report Bayesian odds ratio estimates for individual studies. There is no estimate for Luby et al., 2006 (19) due to zero deaths in the control group. Panel B Columns 1 to 5 report Peto odds ratio estimates for individual studies. Column 6 reports the random effects inverse variance odds ratio meta-analysis estimate including the studies from columns 1 through 5 except Luby et al., 2006 (19). Column 7 reports the fixed effects inverse variance odds ratio meta-analysis estimate including the studies from columns 1 through 5 except Luby et al., 2006 (19). Column 8 reports the Bayesian/Peto odds ratio meta-analysis estimate including the studies from columns 1 through 5.

Table S5. Sensitivity of main results to dropping each study

	Haushofer et al., 2021 (33)	Luby et al., 2018 (20)	Null et al., 2018 (18)	Kremer et. al., 2011 (29)	Humphrey et al., 2019 (31)	Kirby et al., 2019 (32)	Dupas et al., 2021 (34)	Reller et al., 2003 (27)	Boisson et al., 2013 (30)	Peletz et al., 2012 (22)	Luby et al., 2006 (19)	Quick et al., 1999 (35)	Crump et al., 2005 (21)	Chiller et al., 2006 (28)	Semenza et al., 1998 (26)
Panel A: Bayes Odds Ratio															
Mean effect	0.80	0.68	0.68	0.69	0.67	0.71	0.69	0.73	0.70	0.72	0.70	0.71	0.80	0.72	0.73
CrI 95%	(0.66, 0.97)	(0.49, 0.93)	(0.49,0.94)	(0.50, 0.94)	(0.49, 0.90)	(0.53, 0.94)	(0.53, 0.90)	(0.56, 0.95)	(0.54, 0.92)	(0.55, 0.94)	(0.54, 0.91)	(0.55, 0.92)	(0.66, 0.97)	(0.56, 0.93)	(0.57, 0.94)
% weight in meta-analysis	8.8	15.7	17.9	13.3	17.4	6.7	2.7	3.6	1.2	3.2	0.8	0.4	7.0	0.4	0.8
Panel B: Peto Odds Ratio															
Mean effect	0.77	0.69	0.69	0.70	0.67	0.70	0.69	0.73	0.70	0.73	0.70	0.72	0.78	0.72	0.73
CI 95%	(0.54, 1.01)	(0.45, 0.99)	(0.44, 0.99)	(0.45, 1.01)	(0.42, 0.96)	(0.45, 0.99)	(0.46 ,0.92)	(0.48, 1.01)	(0.46, 0.95)	(0.50, 0.99)	(0.46, 0.93)	(0.50, 0.96)	(0.56, 1.028)	(0.50, 0.96)	(0.51, 0.99)
% weight in meta-analysis	7.3	15.5	17.6	12.7	17.8	6.9	2.8	5.2	1.4	3.4	0.5	0.3	7.6	0.6	0.5

Notes: Columns 1 through 15 report meta-analysis estimates of OR obtained by excluding the study in the column heading from the full sample. Panel A reports Bayesian odds ratio estimates, and Panel B reports Peto odds ratio estimates. Row 3 of each panel reports the weight of each study in the meta-analysis from Table S3.

Table S6. Additional sensitivity checks

	Combining studies that cover related programs		Including study with contaminated control group I		Including study with contaminated control group II		Alternate control in study with active and passive arms		Alternate treatment in spring protection		Studies where water treatment was combined with another intervention	
	(18, 33)		(37)		(36)		(18)		(29)		(31, 32)	
	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ITT effect on child mortality	0.71	0.69	0.72	0.71	0.74	0.73	0.72	0.70	0.72	0.70	0.66	0.64
CI/CrI 95%	(0.56, 0.91)	(0.48, 0.92)	(0.57, 0.92)	(0.49, 0.92)	(0.57, 0.96)	(0.50, 0.97)	(0.55, 0.93)	(0.47, 0.94)	(0.56, 0.92)	(0.48, 0.91)	(0.47, 0.93)	(0.41, 0.89)
p-value	0.006		0.01		0.022		0.014		0.009		0.02	

Notes: Columns 1 and 2 present meta-analysis estimates combining Null et al., 2018 (18) and Haushofer et al., 2021 (33) into a single study. Columns 3 and 4 present meta-analysis estimates including Du Preez et al., 2011 (37). Column 5 and 6 present meta-analysis estimates including Boisson et al., 2010 (36). Columns 7 and 8 present meta-analysis estimates using only the active control group in Null et al., 2018 (18). Columns 9 and 10 present meta-analysis estimates including both those who received spring protection in year 1 and 2 into the treatment group in Kremer et al., 2011 (29). Columns 11 and 12 present meta-analysis estimates by dropping studies (31, 32) where the water treatment intervention was combined with other interventions (cookstoves, sanitation and hygiene).

Table S7. Total lives saved and costs: preliminary calculations for the global Coupon Program

Target Population	Source	Population size (millions)	# of <5y children without access to safe drinking water (millions)	Number of deaths among <5y without access to safe drinking water per year (millions)	Cost of providing coupons to <5 population without access to safe drinking water per year ($ millions)	Total <5y lives saved per year (thousands)
		(1)	(2)	(3)	(4)	(5)
Population without access to safely managed drinking water services	WHO/UNICEF, 2019 (1)	2200	272.8	2.739	1091.2	372
Population using a source of drinking water which suffers from fecal contamination	Bain et al., 2014 (46)	1800	223.2	2.241	892.8	304
Population using a source of drinking water with >10 E coli or TTC per 100 ml	Bain et al., 2014 (46)	1100	136.4	1.369	545.6	186
Population without access to safely managed drinking water services in countries where PSI sells chlorine[1]	WHO/UNICEF, 2019 (1)	561	69.6	0.698	278.4	95

[1] Countries are Zambia, Madagascar, Tanzania, Rwanda, Malawi, Kenya, Afghanistan, Burkina Faso, India, Uzbekistan, Myanmar, Mozambique, Nigeria, Uganda, Nepal, Vietnam, Ethiopia, Burundi, Guinea, and Cameroon.

Notes: Column 2 is calculated by multiplying (1) by the mean under 5 population share across countries weighted by population without access to safe drinking water. Column 3 is calculated by multiplying (2) by the mean annual mortality rate in the under 5 population , calculated as the mean <5 mortality rate across countries (UN Interagency Group for Child Mortality Estimation) weighted by population without access to safe drinking water (WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene, 2019 (1)/5). Column 4 is calculated by multiplying the cost of providing coupons from Table S8 row 7 by (2) for one year. Column 5 is calculated by multiplying (3) by the estimated reduction in child mortality adjusted by usage rates: (1 – posterior predictive estimate of effect (RR)) * usage rate in meta-analysis / usage rate in coupons from mean across Dupas et al., 2016 (45) and Dupas et al., 2021 (34).

Sources: WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene, 2019 (1), Bain et al., 2014 (46)

Table S8. Cost–effectiveness analysis

	Chlorine Dispensers in Western Kenya			Global Coupon Program		
	Meta-analysis, diffuse prior	***Meta-analysis, informative Gaussian prior***	***Meta-analysis, informative Student's T prior***	***Meta-analysis, diffuse prior***	***Meta-analysis, informative Gaussian prior***	***Meta-analysis, informative Student's T prior***
(1) Posterior predictive estimate (RR) of effect, mean [A]	0.75	0.84	0.82	0.75	0.83	0.81
(2) <5y mortality rate (in pp) [B]	6.9	6.9	6.9	5	5	5
(3) Average take-up in meta-analysis	0.59	0.59	0.59	0.59	0.59	0.59
(4) Average take-up rate of intervention [C]	0.51	0.51	0.51	0.32	0.32	0.32
(5) Expected deaths averted, per person [D]	0.0147	0.0098	0.011	0.0067	0.0045	0.0051
(6) Expected DALYs averted, <5 child [E]	1.16	0.77	0.87	0.53	0.36	0.4
(7) Cost of provision per <5 child, 5 years (USD) [F]	45.5	45.5	45.5	20	20	20
(8) Cost per death of a <5 child averted (USD) [G]	3104	4656	4124	2974	4451	3936
(9) Cost per DALY averted (USD) [H]	39	59	52	38	56	50

[A] Converted from posterior predictive distribution of ORs, assuming mortality risk as in row (2)

[B] Dispensers: Kenya DHS, 2014 (103) (<5 mortality rate weighted by the % of dispensers present in that region); Coupons: UN Interagency Group for Child Mortality Estimation (104) (<5 mortality rate across countries weighted by population without access to safe drinking water.)

[C] Dispensers: Evidence Action (43); Coupons: mean across Dupas et al., 2016 (45) and Dupas et al., 2021 (34)

[D] ((2)/100)* ((5)/(4))*(1-(3))

[E] (2)/100)* ((5)/(4))*(1-(3))*DALYs lost from child death <5y; The number of DALYs lost from death under 5 assumes a life expectancy of 81.25 years and average age at death of 2, following the standard approach of calculating DALY outlined in "WHO methods and data sources for global burden of disease estimates 2000-2019" (2).

[F] Dispensers: Evidence Action (43), 0.30 USD (Retail cost per bottle of chlorine) * 2 (Assumption that administrative costs are as large as the price of chlorine bottles) * 12 months * 0.37 (Share of coupons redeemed) * 1.5 (Assumption that for every two households with a child <5y without access to safe drinking water, one untargeted household receives coupons. 0.37 is the average share of coupons redeemed across Dupas et al., 2016 (45) and Dupas et al., 2020 (34).

Coupons: Dupas et al., 2016 (45) and Dupas et al., 2021 (34), average cost over the 2016-2017 period

[G] ((8)/(6))

[H] ((8)/(7))