WORKING PAPER · NO. 2022-26

Water Treatment and Child Mortality: A Meta-analysis and Cost-effectiveness Analysis

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

JANUARY 2023 (UPDATED APRIL 2024)

Development Innovation Lab at University of Chicago

An Affiliate of Becker Friedman Institute

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Water treatment and child mortality: a meta-analysis and cost-effectiveness analysis

Authors: Michael Kremer[1], Stephen Luby[2], Ricardo Maertens[3], Brandon Tan[3*], Witold Więcek[4]

[1]Department of Economics, University of Chicago; [2]Center for Innovation in Global Health, Stanford University; [3]Department of Economics, Harvard University; [4]Development Innovation Lab, University of Chicago

*Correspondence to: btan2@imf.org

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.

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 18 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 25-28% (frequentist odds ratio, OR, 0.75; 95% CI 0.60 to 0.93; Bayes OR 0.72; 95% CrI 0.51 to 0.94). The results were qualitatively similar under alternative modeling and data inclusion choices. Taking into account heterogeneity across studies, the expected reduction in a new study is 24%.

We used the results to examine the cost-effectiveness of three water treatment approaches, point-of-collection chlorine dispensers, inline chlorination, and a program providing free chlorine solution through maternal and child health (MCH) services. After accounting for delivery costs, we estimate a cost per expected DALY averted due to water treatment between USD 27 and USD 66, depending on approach. This suggests that water treatment is one of the most cost-effective health programs available.

Funding

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

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-reported 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 inline chlorination devices, which automatically chlorinate water.

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 programs on child mortality.

Results

Systematic review

Figure 1 illustrates the search and the selection process. We 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 83 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. We screened additional six studies, three of which we included, based on another recent meta-analysis of safe drinking water interventions that was published after we performed our search (13). The sample of 20 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 Supplementary Information, section 3). Raw input data for meta-analysis are in Table S3.[1]

Publication bias

Neither Egger's nor Andrews and Kasy's tests provided evidence of publication bias on diarrhea or mortality outcomes. Based on the larger sample of 83 studies, we also did not find evidence of the magnitude of effect or of positive significant effects on diarrhea being associated with availability of mortality outcomes. Since the power of these tests for mortality outcome may be

[1] 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.

limited when applied to our sample of 18 studies, we also conducted post hoc simulations. We find that even if our search strategy has missed 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) but these were then found and added to our dataset, the meta-analytic estimate of OR would still be significant. We provide more details in Supplementary Information, 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 of the analyzed 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

The studies included 27,316 participants. Out of 18 studies, 13 were of water chlorination, three of water filtration, and one each of spring protection and solar disinfection. For 11 studies we used individual-level data which were obtained from the authors or publicly available. In aggregate, 173 deaths occurred among 12,700 children in treatment arms (1.4%); in the control arms, 345 deaths occurred among 14,616 children (2.4%). Five studies had no deaths in control and/or treatment arms. The annual risk of mortality in the pooled control group was about 1.6%. We found studies to be representative of diarrhea prevalence in LMICs (see Figure S6) and

found no significant differences to the larger set of RCTs which measured diarrhea. We provide detailed characteristics of 18 included studies and details of the comparison with other RCTs in Supplementary Information, section 1.

Meta-analysis

In the full set of 18 studies, using a random-effects model we estimated a significant average reduction in odds of all-cause child mortality of 25% (frequentist OR 0.75; CI 95% 0.60, 0.92) or 28% (Bayes OR 0.72; CrI 95% 0.51, 0.94), depending on the model (see Figure 2).[2] [3] OR confidence/credibility intervals for individual studies were typically wide, as one would expect in modeling rare event data. In fact, in only two studies the frequentist or Bayesian OR 95% intervals were below 1. Restricting the analysis to studies including chlorination, the reduction was 24% and 27% respectively (frequentist OR 0.76; CI 95% 0.60, 0.97; Bayes OR 0.73; CrI 95% 0.42, 1.11).

There is mixed evidence on cross-study heterogeneity. The frequentist restricted maximum likelihood estimator suggested no heterogeneity (tau = 0; SE = 0.05), which means its estimated effect is the same as in the fixed-effects model (as well as the Bayesian fixed effects model; OR = 0.75, 95% CrI 0.60, 0.93). However, frequentist meta-analyses can underestimate between-study variation and "snap" estimates to zero (14). Given the differences in the studied programs and in the settings where the studies were carried out, a random-effects model is most

[2] 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 S7 for details), our Bayesian OR estimate implies mean risk reduction of 29% (Bayes RR of 0.71; 95% CrI 0.50, 0.92).

[3] An earlier version preprint of this paper, available online, used 15 studies and found very similar reductions in ORs.

appropriate. The Bayesian estimate of heterogeneity, in contrast, was considerable (between-study SD of 0.24 on log scale, compared to the mean log(OR) effect of -0.33) but imprecisely determined (95% CrI 0.01, 0.73). [4]

The expected reduction in mortality odds in a new study, which is used in the cost-effectiveness calculations and is based on a Bayesian posterior predictive distribution, was 24%.[5]

We provide results for all sensitivity analyses of data and model choices that we performed in Supplementary Information, section 3. The estimates from sensitivity analyses we attempted remained qualitatively similar, with mean OR estimates from 64 models ranging from 0.66 to 0.81 for alternative data and modeling choices.

Univariate Bayesian meta-regression models found a significant relationship (in the sense of the 95% interval not including zero) between effects on mortality and effects on diarrhea (beta = 0.97, 95% CrI 0.05 to 1.93), but not on compliance, prevalence of diarrhea, year of study, type of intervention, or randomization unit (cluster vs individual). The relationship between study year and mortality effects is shown in Fig S5. However, we conclude that the power to detect these relationships is generally low; we conducted a simple post-hoc simulation analysis to illustrate this in Supplementary Information, section 6. Therefore more data should be collected to explore

[4] The mean Bayesian estimate of I-squared statistic, defined as the share (in percent) of variation due to underlying variation in true ORs, was 20% (95% CrI from 0%, 69%). 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. We discuss this further in Supplementary Information, section 3.

[5] See details in Supplementary Information, section 2. The effect in a new study is different to the mean across 18 studies because we model logarithms of ORs, which are approximately normal. The heterogeneity in effects 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, the total 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.

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 are based on the expected 24% reduction in the odds of mortality in a new implementation. Results are shown in Table 2 and more details are given in Supplementary Information, section 4 and Table S7.

We calculate two cost-effectiveness metrics, each of which is relevant for a different purpose. First, cost per DALY averted, which is relevant for a decision-maker aiming to maximize health benefits with a small fixed health budget. They would have a given willingness to pay for each unit of health benefit achieved, determined by the cost-effectiveness of their next best option. Second, net benefits per person served. This is relevant for decision makers, such as governments, who can decide how to allocate budgets between different sectors. They would prefer to increase the size of a health program even if the additional spending had a lower cost-benefit ratio, so long as the additional benefit outweighs the additional cost.

We examine three water treatment approaches for cost-effectiveness analysis, chosen because they are appropriate for delivery at scale and data or estimates are available for their costs. First, point-of-collection dispensers of dilute chlorine solution in Kenya, for which we have access to cost data from a large-scale implementation. Second, treatment of piped water through inline chlorination planned in two Indian states. Third, a hypothetical global program delivering free water-treatment through maternal and neonatal healthcare services (MCH), 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.

Point-of-access chlorine dispensers

Cost data was provided by the NGO Evidence Action, which operates dispensers 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 chlorine water treatment. Given an increase in chlorination from baseline of 36% (measured by Evidence Action), approximately 1.1 million people are estimated to treat their water (15). We calculate the per-child cost of provision of chlorine treated drinking water for five years at USD 56.1 (Table 2, Column 1). This leads to cost per DALY averted by the dispenser program of USD 66, far lower than Kenya's GDP per capita (USD 2,099 in 2022).[6] Using a conservative metric of 1x GDP per capita for the value of statistical life, chlorine dispensers have net benefits of about USD 1,720 per child under 5 served.

Inline chlorination

In settings where water is delivered through pipes or from storage tanks, inline chlorination devices can be used to automatically chlorinate the water, meaning that water is treated for everyone using the water source. A range of devices are available for different settings and price points, some of which do not require electricity.

[6] We compare cost per DALY with two commonly used metrics. First, thresholds of 1x and 3x GDP, which were first proposed by the Commission on Macroeconomics and Health and used in earlier editions of the WHO's WHO-CHOICE report (henceforth 1xGDP threshold). Second, cost-effectiveness brackets (e.g. $10 - $100 per DALY averted) used by lists of the most cost-effective child health interventions, including the most recent edition of WHO-CHOICE publication and the World Bank's DCP-3.

We use cost estimates from India, where state and national governments have expanded access to piped water in rural areas, and at least two state governments are planning to incorporate inline chlorination, with technical assistance from non-profit Evidence Action. We assume an increase in chlorination of 69% (observed by a study of inline chlorination in Bangladesh**)**. The estimated per-child cost for five years is about USD 60 (Table 2 column 2). Based on this, we calculate USD 63 per DALY averted. Again this is substantially lower than the 1x GDP per capita threshold for India (USD 2,388 in 2022). This implies net benefits of about USD 2,209 per child under 5 served.

Integrating water treatment into existing maternal and child health services

Point of use water treatment can be delivered through existing maternal and child health services, as with other preventive health products. This approach could scale widely, including in areas where dispensers and inline chlorination are not appropriate, since maternal and child health services now reach the vast majority of mothers. Since it uses existing physical infrastructure, distribution systems, and staff, it would be low-cost. It also targets those most at risk of waterborne disease (young children). To address concerns about wastage, health facilities could distribute water treatment using a coupon system. Coupons for free water treatment solution can be distributed during routine maternal and child health visits, for redemption at local shops, pharmacies, or clinics. This approach has been evaluated at small scale (1,118 participants) in Kenya, and modest scale (14,522 participants) in Malawi. Both studies find that it efficiently targets water treatment to those who will use it.

A 150-milliliter bottle of dilute chlorine solution sufficient for treating one household-month of water costs USD 0.31 and we estimate the total program costs per-child at USD 13 for five years (Table 2, column 3). This implies a cost of USD 27 per DALY averted. This is substantially lower

than the weighted average of GDP per capita for these countries (USD 2,200). MCH delivery appears more cost-effective than the other examples discussed here, in part because our analysis only considers the benefits of reduced child mortality and, unlike inline chlorination or dispensers, the coupon program specifically targets children under-5. We estimate that this program has net benefits of about USD 1,035 per child under 5 served.

Discussion

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 18 studies, we estimate that water treatment reduced the odds of all-cause child mortality by about a quarter on average.

Cost-effectiveness thresholds and lists of priority health interventions

Some multilateral organizations produce lists of the most cost-effective, evidence-backed health approaches and recommend that governments prioritize these approaches for investment. The above cost-effectiveness estimates of $27-$66 per DALY averted would place water treatment near the top of these lists[7]. For example, the WHO-CHOICE's latest publication for maternal, newborn and child health (16) lists 39 interventions which cost less than $100 per DALY averted[8] (including childhood vaccination, nutritional supplementation, and malaria treatment) and eight interventions which cost $100-$1,000 per DALY averted. The World Bank's Disease

[7] Some lists only include goods and services delivered through health facilities, so water treatment delivered through MCH would be in scope, but inline chlorination and dispensers might not.
[8] WHO-CHOICE uses 'healthy life years', which are equivalent to DALYs but represent a year gained rather than a year lost. We use DALYs here for consistency.

Control Priorities 3 (DCP-3) "highest priority package for Essential Universal Health Coverage" requires interventions to be "very good value for money in low-income countries, less than US $200–$300 per DALY averted" (17).[9]

A third widely used cost-effectiveness threshold, first suggested by the WHO's Macroeconomics and Health, is 3x GDP per capita per DALY averted for cost-effective interventions, and 1x GDP per capita for highly-cost effective. Our estimates suggest that water treatment exceeds the 1x GDP threshold 30 to 80 times. Indeed, because the cost of water treatment is low, even small effects on mortality would meet these thresholds. For example, repeating the calculation for chlorine dispensers in Kenya, we find that the threshold of 1x GDP is reached at 0.7% reduction in odds of under-5 mortality.

Selecting delivery methods to maximize net benefits

The appropriate water treatment approach will likely vary by setting. Standard decision theory suggests that policymakers with flexible budgets should allocate budgets so as to maximize expected net benefits rather than cost-benefit ratios. Delivering an intervention to more people increases net benefits, insofar as the benefit is larger than the cost for each additional person served. Decision makers may therefore prefer interventions which are able to reach more people, even if they have lower cost-benefit ratios.

In a location where piped water connections or drinking water storage tanks are widespread, a decision maker would likely prefer inline chlorination, since it has higher take-up. However, in a setting where piped water and storage tanks are rare, a decision maker might prefer MCH

[9] It also categorizes water treatment under "Injury and Environmental Health", with a briefer mention in the volume on reproductive, maternal, newborn and child health.

delivery because it can reach a larger portion of the population. Many countries will have both areas with piped water systems or storage tanks where it is efficient to use inline chlorination, and areas without such infrastructure, where a coupon program would be appropriate.

Magnitude of benefits

To illustrate the potential magnitude of the benefits of water treatment, in Table S6 we present a back-of-envelope calculation, which suggests that a global coupon program delivering water treatment to all households with children under-5 who do not yet have access to piped water in low- and lower-middle income countries, could save over 300,000 under-five lives at a cost of approximately USD 550 million each year. The number of lives saved per dollar invested varies with the mortality rate of the setting. Our illustrative calculation suggests that, assuming constant per capita costs, this program would save roughly four times as many lives in sub-Saharan Africa as it would in East Asia and the Pacific.

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. To illustrate, we calculate 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. The point estimate from this model is a 3.9% reduction in mortality risk. This conservative model is unlikely to accurately describe the

relationship between water treatment and child mortality, for reasons we discuss in Supplementary Information, section 7. We include it here to illustrate how robust our finding on cost-effectiveness is to conservative assumptions. Assuming a 3.9% reduction in mortality, any of the interventions discussed would be cost-effective according to the 1xGDP threshold.

In the same Supplementary Information section 7 we conduct an analysis that formally uses the alternative model (centered at 3.9% reduction) as a Bayesian prior, to estimate whether water treatment meets the most-demanding cost-effectiveness threshold that we identified. We find that the water treatment meets this threshold (USD 200 per DALY averted), unless the analyst is highly confident in the prior. Put in simple terms, the analyst needs to put only 6% weight on the meta-analysis result in the case of chlorine dispensers and 16% for inline chlorination (both reaching the threshold at OR = 0.92). For MCH delivery any precision will suffice, since 3.9% reduction meets the threshold.

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.

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 (Supplementary Information, section 5).

While including short studies does not have a major impact on this analysis (Supplementary Information, 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.[10] 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 (Supplementary Information, 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 18 studies.

Despite high expected value from water treatment, uncertainty remains both about the size of the effect and how it may vary across contexts. For example, differences in cost-effectiveness between the delivery methods discussed here are driven by estimated compliance. However, the relationship between compliance and effect size could have a different form.

[10] 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.

Compliance may also vary across contexts or change over time. Our main estimate for dispensers uses compliance data from NGO Evidence Action. However, data from other implementations suggests lower effects of dispenser provision on water treatment, as take-up might decline over time (32), or counterfactual take-up might be higher than expected (36). However, the effective take-up in water treatment needed to reach 1x GDP per capita threshold ranges from 0.4% (for MCH delivery) to 1.9% (for ILC). We discuss the drivers of cost effectiveness further in Supplementary Information, section 4.

In addition to compliance, several other factors could influence the effect of water treatment on child mortality: counterfactual levels of water treatment, local disease burden from diarrhea 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 Supplementary Information, 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 (Supplementary Information, 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 (Supplementary Information, 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.

We identified 18 studies but for only 10 of them mortality was available in publications. Relying on reported data only would reduce the total sample size from 27,316 to 12,660. Moreover, by requesting data from authors we were able to estimate ORs using the same model, correcting for clustering where necessary. 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.

Materials and Methods

This systematic review was registered within the American Economic Association (AEA), registration number AEARCTR-0005977 (18). We assessed the meta-analysis under PRISMA 2020 guidelines (19), 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 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 (Supplementary Information, section 1).

Two review authors independently assessed the risk of bias using the same Newcastle-Ottawa scale (20) 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.[11] Since death is a rare outcome (typically 1-2% annualized risk, with several studies having zero events in treatment or control arm) and some of the RCTs were cluster-randomized, we first used the same Bayesian method to estimate OR in each study. (We also used frequentist continuity corrections as a

[11] 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.

sensitivity analysis.) We then used two models: a frequentist model and a Bayesian model with diffuse priors (21, 22). All models are described in Supplementary Information, section 2.

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, but we also report results from the fixed-effects models. We decided to estimate average effects for all studies, and the sub-sample of studies that include water chlorination.

We examined potential heterogeneity in treatment effects using Bayesian meta-regression models, fitting one variable at a time for the following: log of mean diarrhea risk ratio estimates, level of compliance (as defined in Supplementary Information), baseline prevalence of diarrhea, type of intervention, unit of randomization (cluster vs household), 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. We motivate these analyses case-by-case, as well as provide detailed results, in Supplementary Information, section 3.

We examined potential publication bias through visual inspection of funnel plots and the use of Egger's and Andrews and Kasy's tests (23). We made two checks: one for mortality outcome, using estimates from 18 studies meta-analyzed here, 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.[12]

Cost effectiveness

Methodological details, references for assumptions, and calculations of cost estimates are given in Supplementary Information, section 4 and in Table S7.

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.

We divide estimated costs for three illustrative delivery mechanisms by the expected number of under-5 deaths and DALYs averted, based on the meta-analysis result. This is conservative, since it does not consider benefits other than through reduced child mortality. We also calculate net benefits by multiplying expected DALYs averted by 1x GDP per capita.

Data sharing

All data and code to replicate the results (including all figures and tables) of this meta-analysis have been made publicly available at https://github.com/DevInnovationLab/i-h2o-meta.

[12] More details on software version and dependencies can be found together with project code at https://github.com/DevInnovationLab/i-h2o-meta.

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, Arthur Baker, 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, Luiza Andrade, Meera Nair, Joy Dada, Alexandre Simoes Gomes, Tessa Han, Cynthia Ji, Edward Jee, Dustin Marshal, Jack Marshall, Shubhalakshmi Nag, Andreas Petrou-Zeniou 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

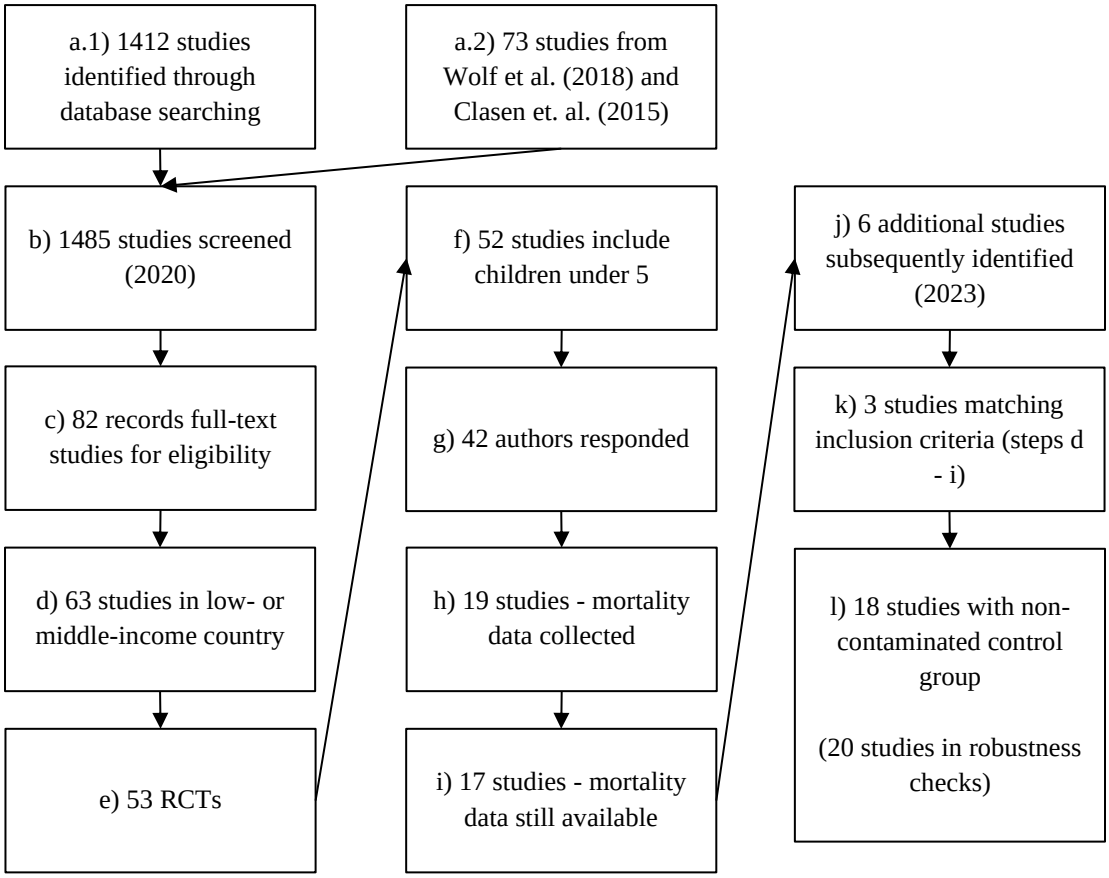
1. United Nations Children's Fund (UNICEF) and World Health Organization, Progress on drinking water, sanitation and hygiene 2000–2017 (2019) (February 21, 2022).
2. , Global Burden of Disease Results Tool. *Glob. Burd. Dis. Results Tool Inst. Health Metr. Eval. IHME* (2015) (February 21, 2022).
3. A. 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. H. S. Waddington, E. Masset, S. Bick, S. Cairncross, Impact on childhood mortality of interventions to improve drinking water, sanitation, and hygiene (WASH) to households: Systematic review and meta-analysis. *PLOS Med.* **20**, e1004215 (2023).
14. Y. Chung, S. Rabe-Hesketh, A. Gelman, J. Liu, V. Dorie, Avoiding Boundary Estimates in Linear Mixed Models Through Weakly Informative Priors. *UC Berkeley Div. Biostat. Work. Pap. Ser.* (2012).
15. , Data provided by Evidence Action.
16. 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.* **10**, 706–723 (2021).
17. D. A. Watkins, *et al.*, "Universal Health Coverage and Essential Packages of Care" in *Disease Control Priorities: Improving Health and Reducing Poverty*, 3rd Ed., D. T. Jamison, *et al.*, Eds. (The International Bank for Reconstruction and Development / The World Bank, 2017) (October 2, 2023).
18. B. Tan, Water Treatment and Child Survival: A Meta-Analysis https:/doi.org/10.1257/rct.5977-1.0 (October 2, 2023).
19. M. J. Page, *et al.*, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).
20. G. A. Wells, *et al.*, The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses (February 21, 2022).
21. 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).
22. W. Wiecek, Meta-analysis of binary data with baggr (2022) (February 21, 2022).
23. I. Andrews, M. Kasy, Identification of and Correction for Publication Bias. *Am. Econ. Rev.* **109**,

2766–2794 (2019).
24. 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).
25. 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).
26. 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).
27. 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).
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. 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).
31. 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).
32. 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).
33. 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).
34. 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).
35. 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).
36. J. Haushofer, M. Kremer, R. Maertens, B. J. Tan, "Water Treatment and Child Mortality: Evidence from Kenya" (National Bureau of Economic Research, 2021).
37. 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* **15**, 272–305 (2023).
38. 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).
39. 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).
40. J. F. Morris, *et al.*, A Randomized Controlled Trial to Assess the Impact of Ceramic Water Filters on Prevention of Diarrhea and Cryptosporidiosis in Infants and Young Children—Western Kenya, 2013. *Am. J. Trop. Med. Hyg.* **98**, 1260–1268 (2018).
41. R. M. Conroy, M. E. Meegan, T. Joyce, K. McGuigan, J. Barnes, Solar disinfection of water reduces diarrhoeal disease: an update. *Arch. Dis. Child.* **81**, 337–338 (1999).
42. 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).

43. 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).

Figures and Tables

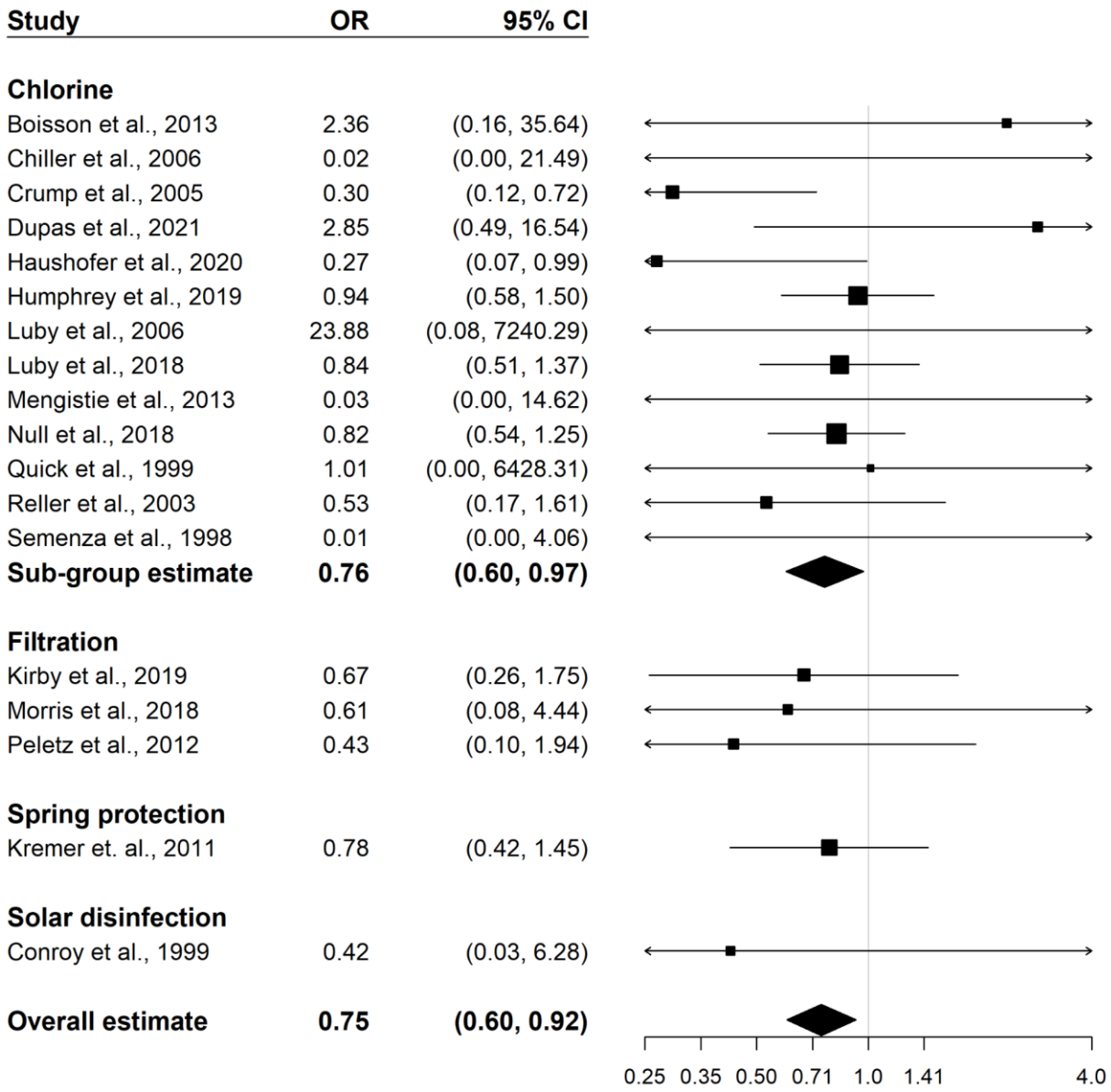
Fig. 1. Study selection



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

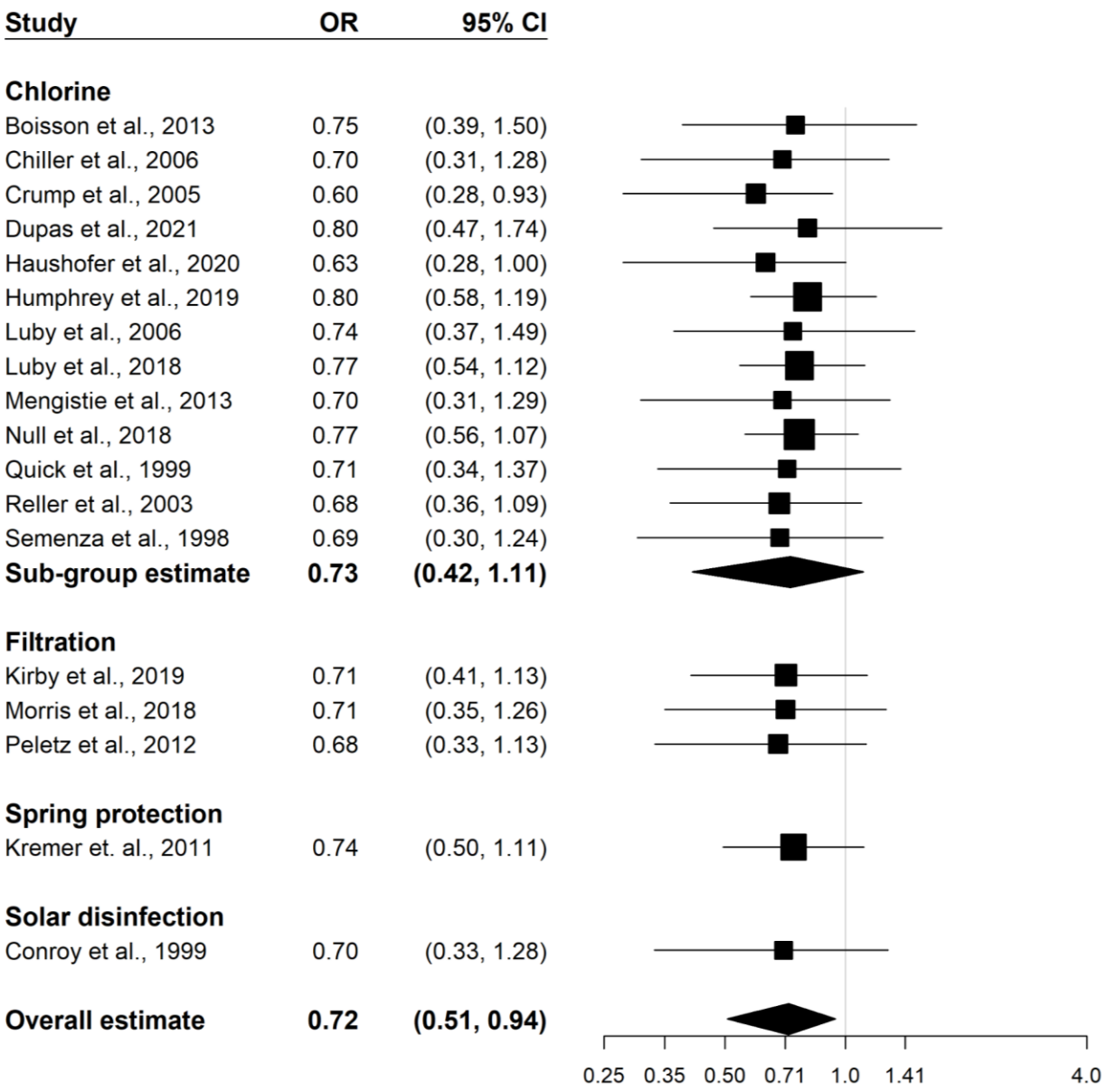
Fig. 2. Forest plots of meta-analysis results

(A) Frequentist model, showing unpooled effects for individual studies



Note: Dots and horizontal lines represent point estimates of the effect of water treatment and their 95% confidence intervals from individual studies. Estimates in rows for individual studies are input 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 model, showing posterior (pooled) effects for individual studies



Note: Dots and horizontal lines represent posterior mean effects (OR) of water treatment and 95% credible intervals for individual studies under a Bayesian model (rather than inputs, unlike the panel A). Diamonds are centered around the posterior mean effect of water treatment for chlorine studies and overall, based on a partially pooled (random effects) model. See Supplementary Information, 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 in each arm	Country	Population of study	Observation period	Infectious environment indicators		Compliance rate**
						Contamination levels	Diarrhea prevalence	
Semenza et al., 1998 (24)	(1) Chlorination	(1) 62 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 (25)	(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 (26)	(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 (27)	(1) Chlorination (2) Flocculant- disinfectant water treatment	(1) 252 hh (2) 261 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 (30)	(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 (31)	(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 (32)	(1) Chlorination	(1) 888 ind. (C) 2811 ind. (active + passive)	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 (33)	(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 (34)	(1) Chlorination + sanitation + hand washing + play space + hygiene counseling +	(1) 918 children. (C) 884 children.	Zimbabwe	Households with a <18m old child	1.5 years	63% of household collected water from	9.5% in control	58%

	construction of improved pit latrines (WASH)					improved water sources at baseline		
Kirby et al., 2019 (35)	(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 (36)	(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*	23%
Dupas et al. 2023(37)	(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 (38)	(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%
Mengistie et al. (2013) (39)	(1) Chlorination	(1) 427 children. (C) 422 children.	Ethiopia	Households with a <5y child	16 weeks	Median E. coli colony count in drinking water (baseline, control): 70 Number of households with E. Coli (baseline, control): 109	24.7% at baseline	79.9%

Morris et al. (2013) (40)	(1) Filtration	(1) 120 children. (C) 120 children.	Kenya	Children 4-10mo	26 weeks	Seroincidence of Cryptosporidum in children (baseline, control): 23	8.9% in control households	71%
Conroy et al. (1999) (41)	(1) Solar disinfection	(1) 175 children. (C) 174 children.	Kenya	Householdswith a <6y child	32 weeks	All source water samples had >200 NTU (nephalometric turbidity units)	58.1% in control households	High
Studies included for robustness checks								
Boisson et al., 2010 (42)	(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 (43)	(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 (32) 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 Supplementary Information, 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	Inline Chlorination in India	Hypothetical Global MCH delivery
Estimated mean OR effect of water treatment on child mortality, mean (95% CrI)		0.72 (0.51, 0.94)	
Posterior predictive estimate (RR) of effect, mean	0.77	0.76	0.77
Effective take-up	0.36	0.69	0.26
Expected DALYs averted per eligible <5 child	0.85	0.95	0.47
Cost of provision per <5 child, 5 years (USD)	56.1	59.6	12.7
Cost per expected death of 5< child averted (USD)	5,256	4,970	2,125
Cost per expected DALY averted (USD)	66	63	27
Net benefits per child <5 served (USD)	1,720	2,209	1035

Notes: this cost-effectiveness calculation is based on the Bayesian model and incorporates uncertainty in predicting effects to a new setting. This calculation only includes the benefits of reduced child mortality. We calculate DALYs averted assuming that a death within the first 5 years of life leads to 79.25 disability-adjusted life years (DALYs) (assuming 81.25 years of life, as recommended by the WHO, and average age of U5 death at 2 years). Further details and references for calculations, assumed costs, and other assumptions are given in Table S7.

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

Supplementary Materials

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

Supplementary Information

Figs. S1 to S7

Tables S1 to S8

1. Details of included studies and comparison with other RCTs

Summary of 18 included studies

The median follow-up length for mortality was 52 weeks, with the longest follow-up being 4-6 years, in two studies conducted in Kenya (1). Twelve studies were conducted in lower-middle income countries, and four were conducted in low-income countries, according to the World Bank classification at the time of the study (2). 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 18 studies, 14 were conducted in rural areas, two were conducted in both rural and urban areas, and two were conducted in a peri-urban setting. The compliance rate (see later in this section for definitions) in the sample ranged from a low of 23% to a high of 87%, with a median of 69%. Average use of water cleaning treatment (see later for definitions) in treatment groups was 56% and in control groups it was about 3%. Based on this, we assumed an effective compliance across meta-analysis (used for cost-effectiveness calculations) of 53%.

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.

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 18 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 0.2% to 56.7% across sub national units, with a median of 13.7% (3). Diarrhea prevalence rates (at baseline or, if baseline not available, in the control group) in our sample of studies range from 5% to 58%, with a weighted mean (using meta-analytic weights from the frequentist model) of 15.6%; this corresponds to the 61st percentile of the distribution of sub national diarrhea estimates, see Figure S6.

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 (1, 3) which report the impact of closely related interventions on different samples, we report sensitivity to combining these studies. For 16 out of 18 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 (4, 5). In two out of the eighteen cases, water treatment was combined with another intervention, cookstoves (6) or other sanitation and hygiene interventions (7).

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. 2021; 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. 2023; Quick et al. 1999; Null et al. 2018, Mengistie et al. 2013). 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. Morris et al. (2018) required that water in the intervention arm have low measured bacteria concentration during three sample rounds. 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: 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: residual free chlorine concentration > 0.1 mg/l during unannounced visits, average across two treatment groups (chlorination + disinfectant)
- 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): 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 WASH households with detectable free chlorine above 0.1 ppm at 12 months
- Kirby et al: "Filter observed and reports last filled since day before yesterday"
- Haushofer et al: free chlorine residual test results above 0.1 mg/L (unannounced visit)
- Dupas et al. (2023): positive chlorine test
- Quick et al.: Proportion of stored water samples with detectable levels of total chlorine
- Mengistie et al. (2013): Presence of free residual chlorine exceeding 0.2 mg/L (unannounced, regular weekly visit)
- Conroy et al. (1999) report that random checks for compliance with solar disinfection (mothers were instructed to place drinking water on the household roof, in sunlight, or indoors) by project workers were conducted, but do not report on compliance
- Morris et al.: proportion of households for which filtrate samples contained <1 MPN E. coli / 100 mL at month 1 and month 6 (in both cases nearly the same value was reported; 71% at month 1 and 70% at month 6).

Similarly, takeup of the intervention in the control group was tracked in several studies.

Definitions for control group takeup for each study are as follows:

- Crump et al: effective free chlorine detected during unannounced visit
- Reller et al: proportion of control households drinking water with detectable free chlorine > 0.1 mg/L
- Chiller et al: households reporting past use of floculant disinfectant, at baseline
- Kremer et al. (2011): households regularly using a treated spring
- Peletz et al. (2012): households with chlorine exceeding over 0.2 mg/mL in a sample taken at baseline
- Boisson et al: presence of residual chlorine in child's drinking water
- Null et al. (2018): detectable free chlorine measured in one-year follow-up
- Luby et al. (2018): households in the control group reporting having treated drinking water yesterday, at baseline
- Kirby et al: households in the control group reporting chlorinating water, at baseline
- Haushofer et al: percentage of households with free chlorine residual test results above 0.1 mg/L (unannounced visit)
- Dupas et al. (2023): positive chlorine test
- Quick et al: detectable chlorine at baseline

- Mengistie et al. (2013): households reporting chlorinating water
- Morris et al: households negative for E. coli at baseline

Several studies were not included in calculation of mean water treatment take-up in controls due to lack of data. In Humphrey et al only available variable was percentage of households in the control group reporting treating water at baseline. For other studies we were not able to find any relevant values.

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 (8). There were 73 studies in (8), yielding 80 observations. Some studies had multiple observations on account of multiple study locations, and hence yielded multiple effect estimates. Nineof these studies were included in our meta-analysis. This means we have a data set of 71 observations not included in our meta-analysis against which we can compare our 18 studies.

The distribution of compliance rates and effect estimates of water treatment on diarrhea are similar across included and excluded data (see Fig. S2).

48 out of 73 studies (65.8%) were conducted in rural settings, with 12.3% and 23.3% being conducted in mixed and urban settings respectively. Similar to this, 14 out of the 18 included studies (78%) are set in rural areas and 4 out of the 18 (22%) included studies are conducted in mixed and urban settings.

In terms of the water source, the primary source of water at baseline (or in the control group) was an unimproved water source[1] in 53 out of 73 observations (73%.) This is comparable to 88.9% (16 out of 18 studies) among the included studies.

We did not find statistically significant differences between the included and excluded studies on any of these variables. A t-test of mean difference between included and excluded studies yield insignificant differences for the diarrhea effect size (p-value=0.42), compliance rate (p-value=0.48), setting (binary variable indicating whether the setting was rural, p-value=0.25), and presence of improved water sources at baseline, as classified in (8) (p-value=0.06). We note that this final p-value is low. However, since we present the results from multiple tests, we consider several multiple testing corrections. Under the Bonferroni correction the revised p-value cutoffs for each test would be 0.025 at 10% significance. (Alternatively, the stepwise method due to Holm provides revised cutoffs of 0.033, 0.05, 0.1 and 0.025, respectively.)

Age characteristics of included children

Three 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
- Morris et al. 2013 studies children between 4-10 months of age.

Additionally, Conroy et al. (1999) included children under 6, not under 5.

[1] per WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (20), unimproved water sources include unprotected wells or springs, or surface water

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., 2021 – most of data collected came from children under the age of 2, while Kremer et al. 2011, Boisson et al. 2013, Dupas et al., 2023 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., 2023	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 individual-level data (and hence no age information other than the inclusion/exclusion criteria into the study) 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., 2021 (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., 2021	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., 2023[2]	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.

[2] For Dupas et al study we had access to individual level data and summary level data. We calculated the numbers of events used for meta-analysis based on summary-level data (2,616 observations), as these counts were confirmed by the study authors, but distribution of ages and person-years calculation is based on individual-level data.

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

	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

Obtaining input odds ratios for the model

As seen in Table 1, eleven of the meta-analyzed studies are cluster RCTs (c-RCTs). We managed to obtain individual-level data (mortality indicators, cluster ID, treatment status) for seven of the studies. For Humphrey et al., Luby et al. (2018), Mengistie et al. and Morris et al. studies we could only extract the summary data from the papers.

For modeling of ORs in each of the 18 studies we used a Bayesian logistic regression model, regressing death outcome on treatment status, fitted using *rstanarm* package in R. For intercept (mortality rate in controls) we used a Gaussian prior distribution with SD of 10, but centered at log(0.01), to encode our knowledge that child mortality is a rare event. For treatment effect we used a Gaussian, zero-centered and with standard deviation of 5.

For c-RCTs with individual-level data (which, with only one exception of Haushofer et al. study, did not report ORs in their papers) we repeated the above, but adding a random effect of cluster ID.

For another three c-RCTs (Mengistie et al., Humphrey et al., Luby et al. 2018) we were able to obtain cluster sizes (ranging from 7 to 24) based on reported numbers of clusters and total sample size.[3] We used them to correct the sample sizes to account for clustering, by dividing them by design effect *d* (*d = 1 + (m-1)*ICC*, where *m* is the average cluster size). We used the ICC value of 0.018, which is the sample-weighted average from the seven cluster-randomised studies which had individual-level data.[4] We then fit the same Bayesian models.

An alternative approach for calculating ORs in studies without clustering

We also recalculated the ORs using continuity corrections, that is, adding a fixed value to each of *a*, *b*, *c*, and *d*, assuming that SE(log(OR)) = sqrt(1/a + 1/b + 1/c + 1/d) (95) . We use this method of calculating inputs into meta-analysis in a sensitivity analysis, assuming a fixed continuity correction of 0.25 for all studies (with exception of c-RCTs with individual-level data, where we rely on the other model in order to appropriately account for clustering).

Fitting frequentist meta-analysis models

Having obtained posterior mean OR's (on logarithmic scale) and their standard deviations from the Bayesian model described above, we used these values as inputs into both frequentist and Bayesian meta-analysis models.

[3] For the Du Preez et al. study the number of villages was not reported and therefore we made no corrections, but only six deaths occurred in that study. For Morris et al. there was no need for adjustments.

[4] The procedure was as follows: for each study with individual-level data available we would calculate the SE of treatment effect with and without cluster random effects. Since ratio of effective sample size to actual sample size is roughly square of the ratio of SE's, we would obtain the implied design effect d for each study. The ICC was then calculated as *(d-1)/(m - 1)*, as per the design effect formula. All of the ICC's obtained this way were then averaged, with weights based on study sample sizes.

For the frequentist model, we used command *rma.uni* from R package *metafor*, with default settings and the Restricted Maximum Likelihood (REML) estimator, which is also a standard choice for all frequentist packages in R (96).

Fitting Bayesian meta-analysis models

We also estimate the effect under a Bayesian hierarchical model. Under this formulation, we sometimes refer to hyperparameters: the mean (referred to as hypermean) on log odds ratios is denoted by τ and the hypervariance (the true variation in mean effects across settings) by σ_τ^2.

To fit the model, we used R package *baggr*. For the main specification, we used mildly informative priors on the hyper-parameters, similar to (10): for τ (the mean effect) we set a normal distribution with mean 0 and standard deviation of 5. This prior encodes the belief that causal effects should not be thought of as large unless data contains evidence to the contrary. We used the same zero-centered normal distribution with SD of 5 for σ_τ^2, the heterogeneity parameter.

In Bayesian meta-regression models, we used the same priors and added a normal prior of coefficients, with mean 0 and SD of 2.5.

The discussion of the Bayesian OR estimates throughout the paper refers to the 95% posterior credible intervals (CrI) from Bayesian inference, which may not be symmetric.

Cross validation approach

Following the literature (10), in the sensitivity analysis section (see below) the model fit for Bayesian models was formally compared between full pooling (fixed effects) and partial pooling

(random effects) specifications by using a cross validation approach. Full pooling model is one where $\sigma_\tau^2 = 0$, i.e., there are no differences between studies (with the same prior for mean effect as in the random-effects model). For each specification, 18 Bayesian hierarchical models were fitted to data by leaving out one study at a time and then calculating expected log predictive density (ELPD) for the left out study (11). 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 all studies is used as the cross-validation information criterion. A value closer to zero implies a better fit.

Study weights in Bayesian model

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 S4.

Posterior predictive distribution

By posterior predictive distribution we mean a distribution of effect that we would expect to see in an additional study, by which we mean either a new implementation of the intervention or an unobserved study that was already conducted. This is the quantity that we use to conduct cost-effectiveness analyses throughout the paper.

To derive a distribution of predicted effects we combine two parameters, τ and σ_τ^2, or hypermean and hypervariance, as per the notation above. This means that the predictive distribution has the

same mean as the historical mean but higher variance (as it combines uncertainty in the mean with between-study variance). However, since we are modeling on log(OR) scale, taking the exponent means that historical mean and predictive distribution mean can differ (since the mean of an exponent of a normal variable depends on variance)

As presented in the main text, the hyper-SD is considerable. Consequently, the dispersion in the predictive distribution is much larger than in the historical mean (Bayes OR of 0.76; 95% CrI 0.33, 1.40). That is, the model assigns considerable probability to a single new study having no positive effects on mortality, even though on average we expect significant positive effects.

3. Sensitivity analysis

List of sensitivity analyses

To understand the impact of model choice on treatment effect estimates, we fitted: (i) fixed effects models instead of random effects, (ii) a model that uses continuity corrections to calculate ORs (see Section 2 above).

For sensitivity to choice of data, we considered the following: (i) exclusion of one study at a time from the analysis, (ii) combining two studies that measure impacts of a similar program on different populations (1, 3), (iii) the inclusion of studies with contaminated control groups (7, 12), (iv) the use of an alternative control group in a study with active and passive control arms (13), (v) use of an alternative treatment group in a spring protection study (14), (vi) restricting to studies with long monitoring durations, and (vii) dropping studies where water treatment was combined with another intervention (6, 7), (viii) contribution of studies with published mortality

outcomes only. Exact results are presented in Tables 5 (i), Table 6 (ii-vii) and Figure S1 (viii), but we summarize the results below.

Summary of result

Over a set of all sensitivity analyses (64 models) we found that the study estimates remain qualitatively similar to our main estimate. In this set of sensitivity analyses, the mean OR estimates range from 0.66 to 0.81. 58 out of 64 models have upper 95% intervals below one.

Case-by-case details of sensitivity analyses

The exclusion of any single particular study. The frequentist and Bayesian odds ratio estimates are given in Table S4. For frequentist OR, the means ranged from 0.71 to 0.79, with the lowest lower bound of 0.55 and the highest upper bound of 0.98. For the Bayesian model, the mean OR ranged from 0.68 to 0.78, with the lowest 95% CrI lower bound of 0.45 and the highest 95% CrI upper bound of 1.01.

Combining studies that cover related programs. Haushofer et al (1) relied on a continuation[5] of a program from another study, Null et al (3). The two studies cover different populations, time-periods, and interventions. In Null et al (3), households have access to both chlorine dispensers, and home delivery of water treatment solution. Haushofer et al use data from some of

[5] In (1), 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 (3). 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 (3) were excluded from the sample. (1) 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 (3), over twice as large as that analyzed in the original study.

the villages included in Null et al study, but using subjects who were not enrolled in the previous study. Mortality data cover the period of four to five years after the initial roll-out of the program studied by Null et al. In Haushofer et al, home delivery of water treatment was discontinued, some chlorine dispensers had closed, and others had opened. As a sensitivity test, we combine these into one study. The meta-analysis estimates remain quantitatively similar and significant with a mean reduction of 32% (see Table S5).

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 (14), the placebo filter removed more than 90% of the source water bacterial contaminant. Participants in the solar disinfection trial (15) 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 S5. Adding the solar disinfection trial (15) to the meta-analysis results in a mean reduction in mortality odds of 25-27%; adding the blinded filtration study (14) results in a mean reduction of 24-26% (see Table S5).

Alternate control group in study with active and passive arms. In the Null et al study, 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 (3) for the meta-analysis, leads to a mean reduction in mortality odds of 24-27% (see Table S5).

Alternate treatment group in spring protection study. In the Kremer et al study (16), 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 (16) for the meta-analysis, the estimated mean reduction in mortality ranges from 25-27% (see Table S5).

Dropping studies which combine water treatment with other interventions. Dropping studies where the water treatment intervention was combined with the provision of cookstoves (6) or other hygiene and sanitation interventions (7) leads to significant OR estimates, with a mean reduction in mortality odds around 30% (Frequentist OR 0.71, Bayes OR 0.68), see Table S5.

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 overweigh the contribution ("effort") of shorter studies. However, the weights assigned to short studies in our models are low, as seen in Table S4. Estimates are expected to be imprecise for studies with shorter monitoring periods, owing to the shorter period over which events can occur. As a sensitivity check, we repeat frequentist 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 S1. We find that the mean reduction in mortality odds ranges from 19% to 25%. Estimates remain significant at 5% until the eleven shortest studies are excluded, but the trend is mild: all estimates are significant at the 10% level.

We conducted an additional check of whether short studies may be unduly impacting the model. We started from 11 studies in the dataset that include one year or more of follow-up data and fit the frequentist 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, 1178. 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 60.6% to 93.0%. In the model with 10 long and 10 simulated short studies the 95% interval was 66.9% to 95.2% (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.

Fixed and random effects model. We fitted both fixed-effects and random-effects models. Under a fixed effect Bayesian logit model the reduction in odds was 24% (OR 0.76, 95% CrI 0.63, 0.91), compared to 28% under the random effects model. Using a leave-one-study-out cross-validation (LOO CV) procedure, the ELPD (for details see Section 2, Cross validation approach) for the partial pooling model was -26.1 (with SE of 4.2) and for the full pooling model -24.9 (SE of 4.8). This suggests no significant differences in the out-of-sample performance of both models, with a slight preference for the full pooling model.

Using continuity corrections to calculate inputs ORs. We also recalculated the ORs using a fixed continuity correction of 0.25 and an assumption of an ICC of 1.8% (in the clustered RCTs; see Section 2, “Obtaining input odds ratios for the model”) for the 11 studies for which we don’t have individual-level data. For the remaining seven studies we used the Bayesian model to

appropriately account for clustering and no continuity corrections. With these inputs, a Bayesian model estimates a 24% reduction in mortality odds (95% CrI 0.61, 0.93), while a frequentist model estimates a 27% reduction (95% CrI 0.53, 0.95).

4. Cost-effectiveness analysis

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 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. In practice, the relevant threshold for a decisionmaker depends on the type of decision maker.

First, decisionmakers with a small fixed health budget can maximize their impact by selecting the most cost-effective approaches. They would have a threshold for their willingness to pay for each unit of health benefit achieved, determined by the cost-effectiveness of their next best option. We therefore calculate the cost per life saved and per DALY, and compare to common cost-effectiveness thresholds.

Second, decision theory suggests that policymakers who can decide how to allocate budgets between different areas should aim to maximize net benefits. They would prefer to increase the size of a health program even if the additional spending had a lower cost-benefit ratio, so long as

the additional benefit outweighs the additional cost. We therefore also calculate net benefits per person served, using a conservative estimate of the value of statistical life.

Since different approaches to water treatment are likely to be suitable for different contexts depending, for example, on whether there are piped water connections already in place, we assess the overall cost-effectiveness of several different approaches to water treatment, comparing them to existing cost-effectiveness thresholds rather than to each other.

In this section we provide more detail for calculations of cost per death averted, cost per DALY averted, and net benefits, and then discuss how decision maker's priors would impact the analysis.

Cost-effectiveness model

Cost effectiveness results are given in the main text and more detailed calculations are given in Table S7 of this supplement. Here, we present the general method and discuss the source of some of the assumptions.

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.

For each of the three approaches (dispensers, ILC, MCH delivery) we conduct the following calculation. First, take the expected percentage reduction of water treatment on child mortality in a new implementation. Second, multiply by the ratio of the expected compliance in the specific delivery method to the average compliance in the meta-analysis (average across 18 studies

weighted by meta-analytic weights). Third, multiply by the under-5 mortality rate in the relevant context. This produces the expected number deaths averted per child under-5 treated.

To estimate the cost per under-5 death averted, we divide the estimated cost per under-5 (over 5 years) by the expected number of deaths averted. To estimate the cost per DALY averted, we divide the cost per death averted by 81.25 (as recommended by the World Health Organization).

To estimate the net benefits per person, we multiply the expected DALYs averted per person by GDP per capita.

Only benefits of reduced child mortality risk are included. Not included are health gains through reduced child morbidity, morbidity or mortality for people over the age of 5 years, or foregone medical expenses.

Data sources for these calculations are discussed next.

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 1.1mln people using the dispensers in western Kenya (17). Some other studies of chlorine dispensers estimate different adherence rates. See below for discussion of how lower adherence would affect cost-effectiveness. The estimated cost of installing and maintaining chlorine

dispensers at scale in western Kenya is USD 9.13 per child under five served, per year (see Table S7).[6]

Inline chlorination

Cost estimates for inline chlorination are provided by Evidence Action, which is providing support and technical assistance to the Government of India on a scale-up of water treatment in Andhra Pradesh and Madya Pradesh. Evidence Action estimates a cost of $134mln over the first seven years of the program. They project that this will be rolled out to reach a target population of 42 million people by year six, leading to 142mln person-years of treatment over seven years. DHS data suggests that 7.9% of this target population is under 5. This leads to an estimated cost of $11.9 per child under 5 for the first seven years of this program.

Estimates of the impact of inline chlorination on water treatment come from a study in Bangladesh.

Integrating water treatment into existing maternal and child health care

We conduct a back-of-the envelope estimate of the cost of a hypothetical global MCH delivery program, covering those in low- and middle-income countries who do not yet have access to piped water. Although such programs have so far only been conducted at modest scales (18, 19), several characteristics of this program suggest that it could be scaled rapidly.

Two randomized evaluations estimate the effect of MCH delivery through a coupon program on take-up of water treatment, a small study (1,118 participants measuring take-up after 3 - 5 months of treatment) in Kenya (19) and a larger study (14,522 participants measuring take-up over 18

[6] 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.

months) in Malawi (18). In Malawi, the program increased the likelihood of household drinking water testing positive for chlorine residual by 26pp (30% in treatment and 4% in control). The Kenya study finds slightly higher take-up in the treatment group (34%), but does not have a pure control group. The Kenya study also found similar take-up rates from household delivery of water treatment solution by community health workers, suggesting that MCH delivery without a coupon system may lead to similar take-up. We therefore use 26pp as our estimate for the impact of MCH delivery on water treatment %.

The under-5 mortality rate is estimated as the mean under-5 mortality rate across low- and lower-middle income countries in the given World Bank region (UN Interagency Group for Child Mortality Estimation), weighted by population without access to piped water (WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene, 2019 (1)). This is a conservative estimate in that it doesn't take into account that within countries, populations without access to piped water are likely to have higher mortality rates. We estimate an under-5 mortality (see Table S7, row 5). To account for potential leakage, we assume that one third of the water treatment is delivered to non-targeted households. Administrative costs of running the program are assumed to be as large as the retail price of the chlorine solution.

Drivers of cost-effectiveness

Our model illustrates that cost-effectiveness is driven by the baseline child mortality rate[7], the cost of the intervention, and the effect of the intervention on water treatment rates. Each of these varies

7

substantially between implementations. For example, the U5 mortality rate is 5.4%. The rate in Nigeria is more than double this, and the rate in Kano State is almost triple.

Similarly, the effect of even similar programs on usage of water treatment can vary over time and between different implementations. The case of chlorine dispensers illustrates this, since various studies have data on take-up for this approach.

The WASH B study (Null 2018) enrolled 8,246 pregnant women between 2012 and 2016. The study had 8 arms, six of which received various combinations of water, sanitation, hygiene, and health information programs, and two of which were control groups. In the arms which included water treatment, chlorine dispensers were provided at the primary water sources of enrolled women. In addition, households were given 1-liter bottles of dilute chlorine solution for point-of-use water disinfection (approximately a 6-month supply) at the start of the study and one more time six months later. The study collected take-up data between January 2014 and July 2016. John and Orkin (2022) collected new data from women in 205 WASH B villages. By this time implementation under the WASH B study had ended, but NGO Evidence Action had taken over the management of the chlorine dispensers. Within each WASH B treatment group, villages were selected randomly. John and Orkin avoided women who had been enrolled in WASH B, but otherwise included all households with a woman aged 18 - 35 who were the most senior woman in their household. Haushofer (2021) analyze the subset of John and Orkin's sample assigned either to pure water (W) treatment or passive control arms in the WASH B study, in order to isolate the impact of water treatment from other interventions.

The WASH B study found that between January 2014 and June 2015, 41% of households in treatment arms receiving water treatment had chlorine residual in their drinking water, falling to 20% for the period February 2015 to July 2016.[8]

John and Orkin found that between April and May 2018, 29% of households in villages eligible to receive dispensers had detectable chlorine in their household drinking water. They found that 19% of households in passive control areas had detectable chlorine. Using the same data for a subset of villages, Haushofer 2021 find that 23% of households in villages assigned to W arms under WASH B had detectable Free Chlorine Residual (FCR), compared to 16% in villages assigned to passive control under WASH B.

This data illustrates that even for a similar approach (chlorine dispensers), usage can vary between implementations and over time. Estimates of take-up of water treatment where chlorine dispensers are provided varies between 23% (for a small subset of villages in April - May 2018), to 55% (the first year of WASH B) . It also illustrates that counterfactual water treatment can vary, from 3% (WASH B) and 16% (Haushofer 2021, April - May 2016). Some of this difference may be driven by seasonality. Households may consider rainwater to be safer than groundwater, and therefore be less likely to treat water in the rainy season.

This leads to a wide variation in the estimated effect of chlorine dispensers on chlorination of drinking water. Between 44pp (the first year of WASH B)and 7pp (Haushofer 2021). We conduct an illustrative cost-effectiveness calculation assuming that dispensers increase water treatment by 7%.

[8] Point-of-use water treatment was delivered to hosueholds during this period in addition to chlorine dispensers. Water treatment rates may have been lower had only water treatment only been delivered through dispensers.

Estimating magnitude of benefits for hypothetical global delivery of water treatment through MCH

Table S6 presents estimates of the magnitude of benefits of a hypothetical global program to deliver water treatment through MCH. MCH programs are most relevant in settings where piped water (which can be treated using inline chlorination) is not available. As such table S6 considers the cost of a program targeting households in low- and lower-middle income countries who do not yet have access to piped water.

2.2 billion people around the world do not have access to safely managed drinking water services (9), see row 1.[9] This number is similar in magnitude to the global estimates from other studies (20, 21). Of this population approximately 220 million are children under the age of 5 (22, 23).[10]

The under-5 mortality rate among populations without access to safe drinking water implies 2.3 million deaths per year in absence of water treatment. Based on the expected effect of water treatment on child mortality in a new implementation and adjusting for usage rates, it is estimated that a program that targets this population would save approximately 305,000 under-5 lives at a cost of approximately 555 million USD each year.

---

[9] 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.

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

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.

5. Publication bias

Publication bias on diarrhea outcome

We used the same dataset of diarrhea outcomes as in section 1 of this document, but with diarrhea outcomes from several studies included in this meta-analysis added to the original dataset, for a total of 88 observations. 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 Wolf et al meta-analysis (8).

We created funnel plots (Figure S4) and estimated Egger's tests for funnel plot asymmetry for all studies and studies that include chlorination interventions only. While asymmetry can arise due to several factors, it is often caused by some form of publication bias. We failed to rejected the hypothesis of symmetry in both cases (p-value=0.854, p-value=0.279 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 (13) on the joint dataset. We assumed 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 -0.38 (SE = 0.06) and (hyper)SD of 0.31 (SE = 0.04). The relative probability of publication (between studies with |z| less/more than 1.96), beta_p, is 1.014, with a standard error of 0.365. That is, insignificant results have virtually the same estimated probability of getting published as significant results.

Repeating the same method for the subset of studies that considered chlorination, the results are noisier, with a (hyper)mean of -0.39 (SE = 0.1), (hyper)SD of 0.32 (SE = 0.06), and beta_p of 1.861 (SE=1.1). Here, too, we cannot reject the hypothesis that relative publication probability is equal to 1.

In summary, funnel plots, Egger's test, and Andrews and Kasy's tests do not suggest evidence of publication bias in papers investigating interventions targeting diarrhea.

Relationship between mortality and diarrhea

We also tested if studies with large estimated diarrhea effects were more likely to report mortality data and found no evidence to support this hypothesis. As mentioned, point estimates and z-values for diarrhea outcome are available for 89 observations, including 17 out of 18 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 failed to reject the null hypothesis, with p-values of 0.46 and 0.99 respectively.

Publication bias on mortality outcome

Next, we repeated the same publication bias checks as for the diarrhea outcome, but now for mortality, this time on the smaller sample of studies that collected mortality data. We report a

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

We also found no evidence of publication bias using the meta-study adjustment methods from Andrews and Kasy. 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 = 2.7). The mean of the true treatment effect across 18 studies (adjusted for publication bias) is not meaningfully different from our estimate, OR = 0.71 (logOR = -0.34, SE = 0.1). Mean hyper-SD (heterogeneity) parameter across studies is 0 (SE = 0), exhibiting similar behavior to the frequentist model in the main text. (13)

Exploratory simulation of small-study publication bias

As described in the main text, it is likely that the power of statistical tests for publication bias is low due to the small number of mortality studies. 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 1178 (the average across 18 studies included in our dataset). We added the simulated unpublished studies to the original dataset of 18 studies and fit all data using the default frequentist OR model. We calculated averages over 250 replications.

With 5 additional studies the estimated reduction in odds was 22%. With 15 additional unpublished simulated studies with a true OR of 1 (i.e. 15 real studies with OR of 0.74 + 15 simulated studies with OR of 1), the meta-analysis estimate had a mean of 0.81, with 95% interval of 0.69 to 0.96. 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.

6. Exploratory assessment of power to detect heterogeneous effects

As discussed, univariate meta-regressions do not find statistically significant linear relationships between predictors and treatment effect (with exception of another outcome in studies, log of reduction in risk of diarrhea). 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 (zero take-up/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 = -0.319*x/0.184, where -0.319 = log(0.727), i.e. the logarithm of the OR estimated by the Bayesian OR model; 0.184 is population-weighted prevalence across 18 studies. In the case of compliance we set it to a = -0.319*x/0.47, where 0.47 is the population-weighted compliance across 16 studies (see Table 1); two studies did not report compliance. In the case of year of implementation we set y = (0.0319/2)*(x-2010) - 0.319, that is, we assume that in 2010 (weighted average of year of implementation in 15 studies) the mean effect was log(0.727) 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 log(OR) for new datasets with some noise ($y = ax + e$), using observed compliance/prevalence/year values for each *x* (for compliance we impute the missing values as mean) and for noise *e* using SDs for 18 studies from the inputs. For each simulated dataset we fit a frequentist RE meta-regression model and check if the meta-regression coefficient is significant. We repeat this 10,000 times. Simulated power is the fraction of coefficients that were significant. We find it to be 23% for compliance, 25% for diarrhea prevalence, and 7% 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 (22), the effect of water treatment on diarrhea is taken from the central estimate in the Clasen et al meta-analysis (24), and mortality is assumed to be linear in diarrhea cases, so reductions in diarrhea deaths are proportional to reductions in diarrheal cases. Future research could investigate to what degree this discrepancy can be accounted for by factors outside this model.

A recent Cochrane review (24) found a reduction in under-five diarrhea due to water quality interventions of 39% (CI 95% 25%, 51%).[11] Under a simple model in which deaths are 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 almost 30% reduction in the odds of all-cause child mortality.

[11] 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, 24).

One potential 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) (25, 26). 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 (27–30). Continued exposure to diarrheal pathogens alters the gut microbiome, increasing susceptibility to infection (31). Such subclinical or clinical episodes of infection can induce impairments in gut function and undernutrition phenotypes leading to increased mortality (32, 33) . Relatedly, diarrhea can lead to malnutrition (34, 35), 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 kill enteroviruses, *Salmonella Typhi* and *Salmonella Paratyphi*, prevent hepatitis A and hepatitis E, and reduce worm loads. Water treatment could also prevent deaths from sepsis among infants by facilitating cleaner births and postnatal care practices (36). Poor water quality and exposure to a more pathogenic environment is associated with preterm birth and low birth weight (37).

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 (38, 39). 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 (40). For example, estimated effects of water treatment on caregiver-reported diarrhea may be subject to reporting bias (12, 39).

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.

Incorporating informative priors in the analysis

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, water treatment would remain cost-effective according to the 1x GDP "highly cost-effective" threshold in all three cases that we studied in this paper (dispensers, inline chlorination, MCH delivery), since to reach 1x GDP threshold the reductions in odds of mortality need to be tiny, less than 1%.

In Figure S7 we illustrate the relationship between the cost per DALY with the effect size (odds ratio on all-cause mortality). We also show the most stringent threshold of 200 USD per DALY averted, which is reached at OR = 0.92 for dispensers and inline chlorination, and OR = 0.97 for MCH delivery.

As we discussed in the main text, a decision maker considering investing in water treatment may wish to combine evidence from the meta-analysis with other sources of evidence. A formal way of combining these estimates in a way that takes model uncertainty into account is to use other sources of evidence as a Bayesian prior. To illustrate this, we use the same Bayesian model as in the main text (see Section 2 of this document), but now include an informative prior on the hypermean that is centered at 3.9% reduction in log of odds, with varying precision (defined as inverse variance).

In the posterior estimate from our main Bayesian model (with diffuse priors) the mean effect is -0.33, with SD of 0.15, on log OR scale. Precision in this model is 1/(0.15^2) = 43.7. For the alternative model, the prior distribution for the mean (of logarithm of OR) is normal, with the mean equal to log(1-0.039) = -0.0398 and varying SD.

We find that in order for cost-effectiveness to exceed the most stringent cut-off of 200 USD per DALY averted, the prior would have to be very precise: SD of 0.06 for dispensers, 0.07 for inline chlorination. For MCH delivery meeting the threshold is guaranteed, since at 3.9% reduction the cost-effectiveness is below 200 USD per DALY averted.

One way to interpret this is by translating it into probabilities of a ten percent or greater reduction in mortality odds (OR = 0.9), asking how confident (a priori) would the analyst have to be that the reduction will not exceed 10%. The answer is 83% for dispensers and 86% for inline chlorination.

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	(water[tw] AND (supply[tw] OR availability[tw] OR access[tw] OR connect*[tw] OR distance[tw] OR improved[tw] OR	TITLE-ABS-KEY (water W/3 (supply OR availability OR access OR connect* OR distance OR improved	(Water near/3 (supply or availability or access or connect* or distance or improved or distribut* or quantity or volume)):ti,ab,kw or

	quantity or volume)).mp or exp water supply/	distribut*[tw] OR quantity[tw] OR volume[tw])) OR "Water Supply"[MeSH]	OR distribut* OR quantity OR volume))	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	("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	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*	((((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

	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 diarrhea/) and (disease$1 or infection$1 or episode$1 or illness$2).mp	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 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])	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*) AND (TITLE-ABS-KEY (disease* OR infection* OR episode* OR illness*))	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 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 , 2013) OR LIMIT-TO (PUBYEAR , 2012)	Publication Year from 2012 to 2016
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) (41), Colford (2005) (42), Rodrigo (2011) (43)
Not a randomized control trial	Kirchhoff (1985) (44), Alam (1989) (45), Mahfouz (1995) (46), Conroy (1996) (47), Xiao (1997) (48), Quick (2002) (49), Jensen (2003) (50), Majuru (2011) (51), Johri et al. (2019) (52), Reese et al. (2019) (53)
Does not include children under 5 years in age	Abebe (2014) (54)
Authors responded but no mortality data collected	Gruber (2013) (55), Günther (2013) (56), Jain (2010) (57), Opryszko (2010a, b, c) (58), Patel (2012) (59), Roberts (2001) (60), Tiwari (2009) (61)), - (1995a, b) (62), Boisson (2009) (63), Doocy (2006) (64), Stauber (2009, 2012a, b) (65–67), Lindquist (2014a, b) (68) , Fabiszewski (2012) (69), Clasen (2004b, c) (70, 71), Pickering et al. (2019) (72), Handzel (1998) (73),
Authors responded and mortality data was collected but no longer available	Gasana (2002) (74), Brown (2008) (75)
Authors did not respond	Torun (1982) (76)* , Austin (1993a,b) (77), Mengistie (2013) (78), McGuigan (2011) (79), Mäusezhal (2009) (80), Lule (2005) (81), du Preez (2008, 2010) (82)

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 (83)	0	88	2	78
Reller et al., 2003 (5)	10	729	5	182
Crump et al., 2005 (84)	9	872	15	447
Luby et al., 2006 (4)	2	1013	0	533
Chiller et al., 2006 (85)	0	549	1	543
Kremer et. al., 2011 (16)	18	691	47	1465
Peletz et al., 2012 (86)	3	58	6	54
Boisson et al., 2013 (87)	2	1505	1	1483
Null et al., 2018 (3)	30	858	114	2697
Luby et al., 2018 (88)	27	629	62	1244
Humphrey et al., 2019 (7)	49	946	50	909
Kirby et al., 2019 (6)	8	1198	12	1252
Haushofer et al. 2021 (1)	7	987	22	965
Dupas et al. 2023 (18)	5	1288	2	1321
Quick et al. 1999 (89)	0	400	0	391
Mengistie et al. 2013 (78)	0	425	1	420
Morris et al. 2018 (90)	2	117	3	115
Conroy et al. 1999 (91)	1	174	2	172
Total	173	12250	345	13970
B. Studies included for robustness checks				
Boisson et al., 2010 (14)	4	81	1	104
du Preez et al., 2011(15)	3	355	3	334

Table S4. Sensitivity of main results to dropping each study

	Panel A: Bayes Odds Ratio			Panel B: Frequentist Odds Ratio		
	(1)	(2)	(3)	(4)	(5)	(6)
Excluded Study	Mean Effect	CrI 95%	% Weight in Meta Analysis	Mean Effect	CrI 95%	% Weight in Meta Analysis
Null et al., 2018	0.69	(0.45, 0.96)	20.13	0.71	(0.55, 0.93)	25.29
Humphrey et al., 2019	0.68	(0.46, 0.93)	17.53	0.71	(0.56, 0.90)	20.37
Luby et al., 2018	0.7	(0.46, 0.95)	16.59	0.73	(0.57, 0.92)	18.71
Kremer et. al., 2011	0.71	(0.47, 0.96)	12.36	0.74	(0.59, 0.93)	12.08
Crump et al., 2005	0.78	(0.57, 1.01)	7.21	0.79	(0.63, 0.98)	5.73
Kirby et al., 2019	0.72	(0.49, 0.97)	6.47	0.75	(0.60, 0.94)	4.98
Reller et al., 2003	0.73	(0.51, 0.98)	5.13	0.76	(0.61, 0.94)	3.7
Haushofer et al., 2021	0.75	(0.54, 0.99)	3.96	0.77	(0.62, 0.95)	2.69
Peletz et al., 2012	0.73	(0.51, 0.97)	3.12	0.76	(0.61, 0.94)	2.02
Dupas et al., 2023	0.70	(0.48, 0.92)	2.38	0.73	(0.59, 0.91)	1.48
Morris et al., 2018	0.72	(0.50, 0.95)	1.91	0.75	(0.60, 0.93)	1.15
Conroy et al., 1999	0.73	(0.51, 0.96)	1.1	0.75	(0.61, 0.93)	0.63
Boisson et al., 2013	0.71	(0.50, 0.93)	1.08	0.74	(0.60, 0.92)	0.62
Luby et al., 2006	0.72	(0.50, 0.94)	0.26	0.74	(0.60, 0.92)	0.14
Semenza et al., 1998	0.73	(0.52, 0.96)	0.25	0.75	(0.61, 0.93)	0.14
Mengistie et al., 2013	0.73	(0.52, 0.95)	0.21	0.75	(0.61, 0.93)	0.11
Chiller et al., 2006	0.73	(0.51, 0.95)	0.18	0.75	(0.61, 0.93)	0.1
Quick et al., 1999	0.72	(0.51, 0.95)	0.11	0.75	(0.60, 0.93)	0.06

Notes: Rows 1 through 18 report meta-analysis estimates of OR obtained by excluding the study in the excluded study column from the full sample. Panel A (columns (1) - (3)) reports Bayesian odds ratio estimates, and Panel B (columns (4) - (6)) reports Frequentist odds ratio estimates. Columns (3) and (6) report the weight of each study in the meta-analysis from Table S3.

Table S5. Additional sensitivity checks

	Combining studies that cover related programs (Haushofer et al and Null et al)		Including study with contaminated control group I		Including study with contaminated control group II		Alternate control in study with active and passive arms (Null et al.)		Alternate treatment in spring protection (Kremer et al.)		Studies that include hand washing interventions/ cookstoves (Humphrey et al, Kirby et al)		Using continuity corrections to calculate inputs ORs	
Study references	(1, 3)		(15)		(14)		(3)		(16)		(6, 7)			
	Freq OR	Bayesian OR	Freq OR	Bayesian OR	Freq OR	Bayesian OR	Freq OR	Bayesian OR	Freq OR	Bayesian OR	Freq OR	Bayesian OR	Freq OR	Bayesian OR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
ITT effect on child mortality	0.68	0.68	0.75	0.73	0.76	0.74	0.76	0.73	0.75	0.73	0.71	0.68	0.76	0.73
CI/CrI 95%	(0.53, 0.87)	(0.49, 0.92)	(0.61, 0.93)	(0.52, 0.95)	(0.61, 0.94)	(0.53, 0.98)	(0.61, 0.94)	(0.51, 0.96)	(0.61, 0.92)	(0.52, 0.94)	(0.55, 0.91)	(0.43, 0.95)	(0.61, 0.93)	(0.53, 0.95)
p-value	0.002		0.008		0.011		0.011		0.007		0.006		0.009	

Table S6. Total lives saved and costs: preliminary calculations for a hypothetical global coupon program

Target Population	Population under five years of age (millions)	Proportion of population without access to piped water (p.p.)	# of <5y children without access to piped water (millions)	<5y mortality rate (p.p.)	Number of deaths among <5y without access to piped water per year (thousands)	Cost to serve full <5 population without access to piped water per year ($ millions)	Total <5y lives saved per year (thousands)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sub-Saharan Africa	140.6	62.8	88.3	8.1	1,473.9	224.3	183.9
Middle East & North Africa	12.1	20.1	2.4	2.2	10.6	6.2	1.3
South Asia	169.2	60.1	101.7	4.1	845.4	258.2	105.5
East Asia & Pacific	43	56.4	24.3	2.9	144.4	61.6	18
Total	**362.1**	**58.8**	**213**	**5.2**	**2,283.8**	**540.9**	**284.9**

Notes:
Column 2 is calculated by taking the proportion of the population without access to piped drinking water.
Column 3 is calculated by multiplying (1) by (2).
Column 4 is the mean mortality rate in the under 5 population, calculated as the mean <5 mortality rate across countries in the given World Bank region (UN Interagency Group for Child Mortality Estimation) weighted by population without access to piped water (WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene, 2019 (9)). For Europe, Central Asia, Latin America, and the Caribbean the mortality rates are relatively low and we did not include them in the calculation.
Column 5 is calculated by multiplying (4), annualized, by (3).
Column 6 is calculated by multiplying the cost from Table S7 row 7 by (3) for one year. This does not account for the fact that within a country, children without access to piped water are likely to have a higher mortality rate, and is thus a conservative calculation.
Column 7 is calculated by multiplying (5) by the estimated reduction in child mortality adjusted by usage rates: (1 – posterior predictive estimate of effect (RR)) * usage rate for MCH delivery / usage rate in meta-analysis from mean across Dupas et al., 2016 (19) and Dupas et al., 2023 (18).
The total row is calculated by recreating the regional calculations over the union of Sub-Saharan Africa, Middle East & North Africa, South Asia, and East Asia & Pacific. The total row is not equal to the sum of the regional rows, since (2) is not a linear function of national populations.

Sources: WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene, 2019 (9)

Table S7. Cost–effectiveness analysis

	Chlorine Dispensers in Western Kenya	In-line Chlorination in India	Hypothetical global MCH delivery
(1) <5y mortality rate (in p.p.) [A]	6.9	3.9	5.2
(2) Posterior predictive mean (RR) of effect [B]	0.77	0.76	0.77
(3) Average effective compliance in meta-analysis	0.53	0.53	0.53
(4) Effective take-up rate for intervention [C]	0.36	0.69	0.26
(5) Expected deaths averted, per person [D]	0.0107	0.012	0.006
(6) Expected DALYs averted, <5 child [E]	0.85	0.95	0.47
(7) Cost of provision per <5 child, 5 years (USD) [F]	56.1	59.6	12.7
(8) Cost per death of a <5 child averted (USD) [G]	5,256	4,970	2,125
(9) Cost per DALY averted (USD) [H]	66	63	27
(10) Net benefits per child under 5 served (USD) [I]	1,720	2,209	1,035

[A] We use overall U5 mortality, to be consistent with the meta-analysis result. Cost-effectiveness will vary by the average age of children enrolled, since mortality risks vary by age. Calculations as follows: **ILC**: authors estimates of weighted average from Andhra Pradesh and Madya Pradesh, based on data from IHME Client portal, 2019 (92) (<5 mortality rate weighted by the % of dispensers present in that region); MCH delivery: UN Interagency Group for Child Mortality Estimation (93) (<5 mortality rate across countries weighted by population without access to piped drinking water - unimproved, surface water, improved, but unpiped, per WHO/UNICEF JMP 2021); **DISPENSERS**: Kenya DHS, 2014 (92) (<5 mortality rate weighted by the % of dispensers present in that region).

[B] Converted from posterior predictive distribution of ORs, assuming mortality risk as in row (1)

[C] Calculated as a difference between treatment and control groups. ILC: Pickering et al., 2019 (72); MCH delivery : Dupas et al., 2023 (18); Dispensers: Evidence Action, weighted average of take-up 2015-2021 compared to baseline

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

[E] Calculated as (5) * DALYs lost from child death <5y. The number of DALYs lost 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" (94).

[F] **ILC**: Cost-data shared by Evidence Action, private communication; MCH delivery : 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 * 5 years * 0.37 (Average share of coupons redeemed across Dupas et al., 2016 (19) and Dupas et al., 2023 (18)) * 1.5 (Assumption that for every two households with a child <5y without access to piped drinking water, one untargeted household receives coupons) / 1.58 (number of <5 children per household with at least one <5 child, calculated as a weighted average of the total number of <5 children over time divided by the total number of households with at least one <5 child over time. We take a weighted average across all countries for which household level data is available. All data is from IPUMS [add citation]); **DISPENSERS**: Evidence Action (17);

[G] ((7)/5));

[H] ((7)/(6))

[I] Calculated as (6) * GDP per capita - (7); assumed GDP per capita of USD 3,142 for LICs and LMICs (coupons), USD 2,099 for Kenya (dispensers) and USD 2,388 for India, per World Bank

Table S13. Risk of Bias Table

Study	Selection bias	Response bias	Allocation bias	Follow-up bias	Exposure assessment	Compliance	Outcome assessment	Outcome measurement	Sum of stars
	Is there evidence of selection bias? Please specify as either yes, possible or no (no=1 star). If yes or possible, please provide details. Consider specifically if intervention and control group are representative for a well defined study population.	*Is there evidence of response bias? Please specify as either yes, possible or no (no=1 star). If yes or possible, please provide details.*	*Is there evidence of bias in allocation of intervention? Please specify as either yes, possible, no (but cluster-random) (=1 star) or no (randomized) (=2 stars). If yes or possible, please provide details. Consider also whether in terms of random allocation was concealed to to those enrolling people/children/households.*	*Is there evidence of bias to follow-up? Please specify as either yes, possible or no (no=1 star). If yes or possible, please provide details. Specify the amount of loss to follow-up.*	*How accurate is the exposure classified? Please specify as either poor (uncertain discrimination), adequate or good (clearly described, good discrimination) (good=1 star).*	*How high was the compliance of the intervention group to the intervention? Please specify as absolute number and rate as either low (<20%), medium (20-50%) or high (>50%) (high=1 star).*	*How was outcome assessed? Parent/person recall? Fieldworker assessed (=1 star)? Physician/microbiologically assessed (=2 stars)?*	*Is there evidence of ascertainment bias? Please specify as either yes, possible or no (no=1 star). Has the assessor and/or person under study been blinded to intervention status?*	*Sum of the resulting quality rating out of 11 possible stars.*
Chiller et al. 2006	* (Households with a <1 year old child)	possible, not blinding, repeated visits, same staff for intervention and health outcome measurement	**	* (6.9% in intervention group; 4.8% in control group) Study period: 13 weeks	*	* (as demonstrated by free chlorine)	* (field worker assessed)	possible, no blinding	7
Crump et al. 2005	* (Family compounds with at least one child <2years)	possible, no blinding	* cluster randomized	* (18.1% in flocculant-disenfectant group; 19% in chlorination group; 18.2% in control group) Study period: 20 weeks	*	*	* (field worker assessed)	possible, no blinding	6
Haushofer et al. 2020	* (Children <5y)	possible, no blinding	* cluster randomized	* (6% in total, and not differential across arms; 7% in treatment versus 5% in control) Study period: 4 years	*	medium (31%)	* (field worker assessed)	possible, no blinding	5

Humphrey et al. 2019	* (Households with an <18m old child)	possible, no blinding	* cluster randomized	* (3.5% in the standard care group; 3.3% in infant and young child feeding (IYCF); 2.5% in the WASH group; and 1.4% in the IYCF+WASH group) Note: Unlike the actual study, we consider for attrition only mothers that left the trial or were lost to the follow-up. Study period: 2 years and 4 months	*	* (79%)	* (field worker assessed)	possible, no blinding	6
Kremer et al. 2011	*	possible, no blinding	* cluster randomized	* (5% of respondents lost to follow-up in the first 2 rounds; 20% across all 3 rounds) Study period: 4 years	poor (community intervention)	*	* (field worker assessed)	possible, no blinding	5
Luby et al. 2006	*	possible, no blinding, same staff for intervention and health outcome recording	* cluster randomized	(13% in households that received flocculent-disinfectant; 4% in control households) Study period: 37 weeks	*	not reported	* (field worker assessed)	possible, no blinding	4
Luby et al. 2018	* (Newborns and their siblings under 36m old)	possible, no blinding	* cluster randomized	* 6% across all arms Study period: 2 years	*	* (81%)	* (field worker assessed)	possible, no blinding	6

Null et al. 2018	* (Newborns)	possible, no blinding	* cluster randomized	* (17% in the active and passive control group, 17% and 14% in the intervention group) Study period: 2 years	*	low (30%)	* (field worker assessed)	possible, no blinding	5
Peletz et al. 2012	Household with a 6month-1year old at enrollment and with HIV + mothers (100 HIV + and 20 HIV -)	possible, no blinding	* randomized	* (13% in intervention group; 9% in control group) Study period: 1 year	*	* (87%)	* (field worker assessed)	possible, no blinding	5
Reller et al. 2003	* (Households with an ≤11m old or pregnant woman in third trimester)	possible, no blinding, same staff recording and encouraging use	**	(13% lost in flocculant-disenfectant alone group; 24% lost in flocculant-disinfectant plus vessel group; 14% in bleach alone group; 13% in bleach plus vessel group; 5% in standard water-handling group) Study period: 50 weeks	*	medium - only 27% of household with effective level of free chlorine	* (field worker assessed)	possible, no blinding	5
Semenza et al. 1998	possible, intervention in those with non-piped drinking water	possible, no blinding	** for the POU chlorine treatment intervention	* no evidence of attrition Study period: 9.5 weeks surveillance time	*	* 73% of water samples contained chlorine residuals	* (field worker assessed)	possible, no blinding	6
Boisson et al. 2013	* (Household with at-least one child <5 year old)	* (double blinded)	** randomized	* (12% across treatment and control groups; attritopm was not associated with treatment arm (p-val=0.94))	*	low (32.0%)	* (field worker assessed)	* double-blinded	8
Kirby et al. 2019	* (Household with children <4 years)	possible, no blinding	* cluster randomized	* (5.1% in round 1, 8.5% in round 2 and 11.1% in round 3 across all arms; reasons for attrition were similar across both arms)	*	* 69.9%	* (field worker assessed)	possible, no blinding	6

Quick et al. 1999	* (All households in study area)	possible, no blinding	** randomized	* no evidence of attrition Study period: 34 weeks surveillance time	*	* ranging from 70-95% across sampling rounds	* (field worker assessed)	possible, no blinding	7
Dupas et al. 2021	* (Household with a child <6 years)	possible, no blinding	* cluster randomized	* (Probability of attrition is similar across study arms) -	*	low - chlorine usage measured through water tests remain at 30%	* (field worker assessed)	possible, no blinding	5
Conroy et al. 1991	Yes - only Maasai children studied	possible, mitigated by Maasai elder fieldworkers	possible, treatment assigned by alternating households	* no attrition	*	* (claims of no evidence of non-compliance)	* (field worker assessed)	possible, no blinding	4
Mengistie et al. 2013	* (Household with a child <6 years)	possible, no blinding	* cluster randomized	* (around 0.4% attrition in both arms)	*	* (80% compliance)	* (field worker assessed)	possible, no blinding	6
Morris et al. 2018	* (Household with a child 4-10 months)	possible, no blinding, medical facility visits tracked	** randomized	Not reported, account for possible 20% attrition rate	*	* (71% compliance)	Weekly questionnaires	possible, no blinding	5

a Please specify as either yes, possible or no (no=1 star). If yes or possible, please provide details. Consider specifically if intervention and control groups are representative for a well defined study population.

b Please specify as either yes, possible or no (no=1 star). If yes or possible, please provide details.

c Please specify as either yes, possible, no (but cluster-random) (=1 star) or no (randomized) (=2 stars). If yes or possible, please provide details. Consider also whether in terms of random allocation was concealed to those enrolling people/children/households.

d Please specify as either yes, possible or no (no=1 star). If yes or possible, please provide details. Specify the amount of loss to follow-up.

e Please specify as either poor (uncertain discrimination), adequate or good (clearly described, good discrimination) (good=1 star).

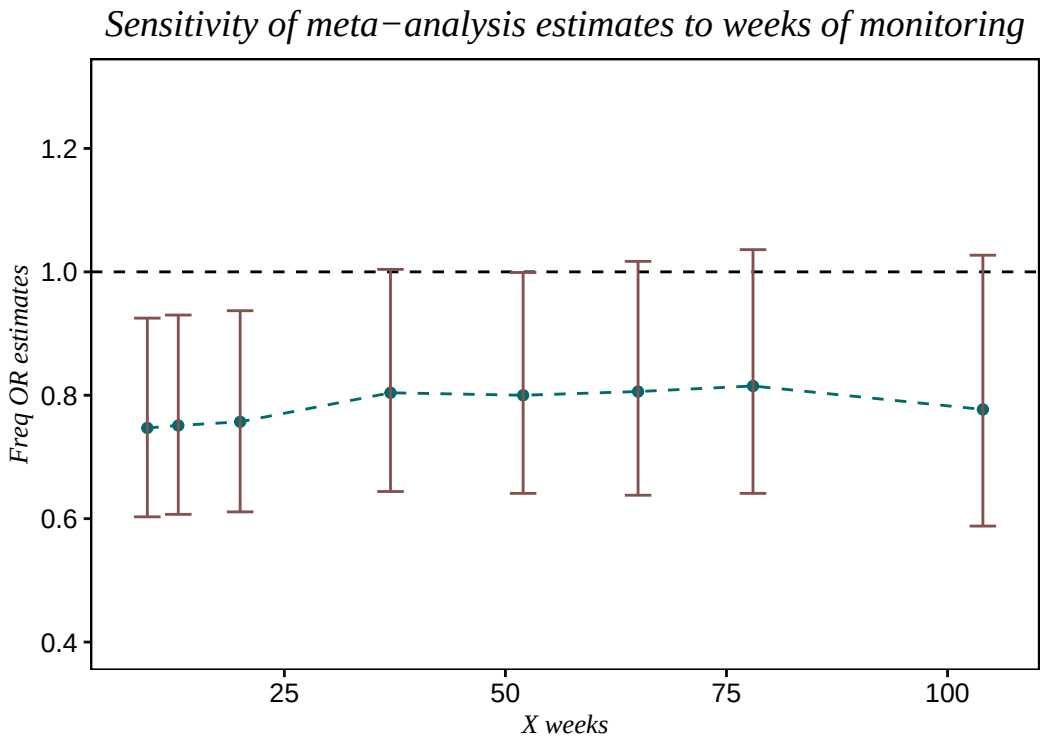
f Please specify as absolute number and rate as either low (<20%), medium (20-50%) or high (>50%) (high=1 star).

g Parent/person recall? Fieldworker assessed (=1 star)? Physician/microbiologically assessed (=2 stars)?

h Please specify as either yes, possible or no (no=1 star). Has the assessor and/or person under study been blinded to intervention status

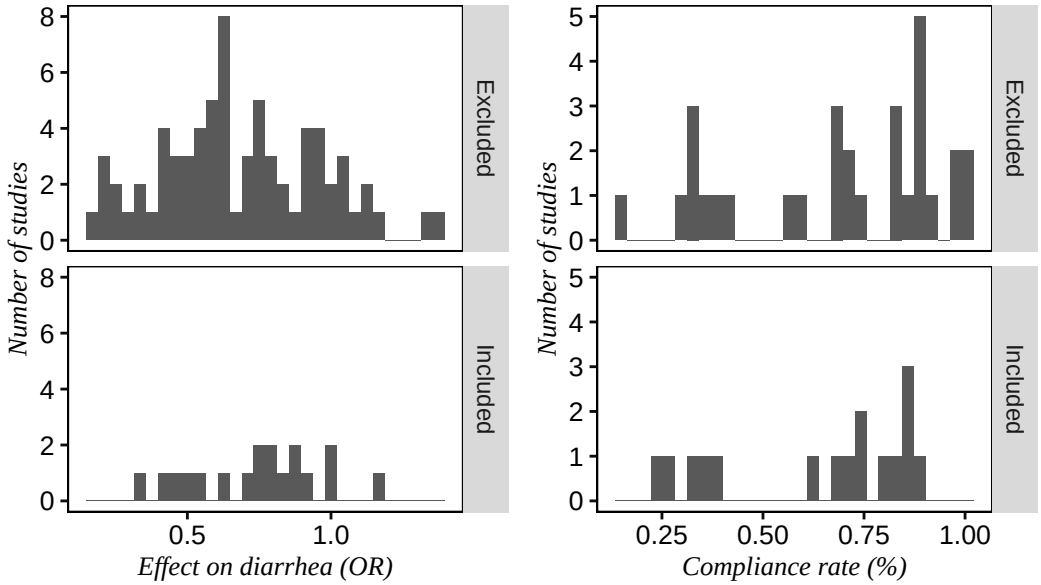
Supplementary Figures

Fig. S1. Restricting set of studies to longer follow-up lengths



Notes: This figure presents the odds ratio estimated by the frequentist meta-analysis model with studies shorter than X weeks removed. Each point is the frequentist OR estimate, and the bars represent the 95% Confidence Interval for each estimate. All 18 studies in the main sample are included for X = 9.5 weeks, and 4 studies are included for X = 104 weeks (2 years).

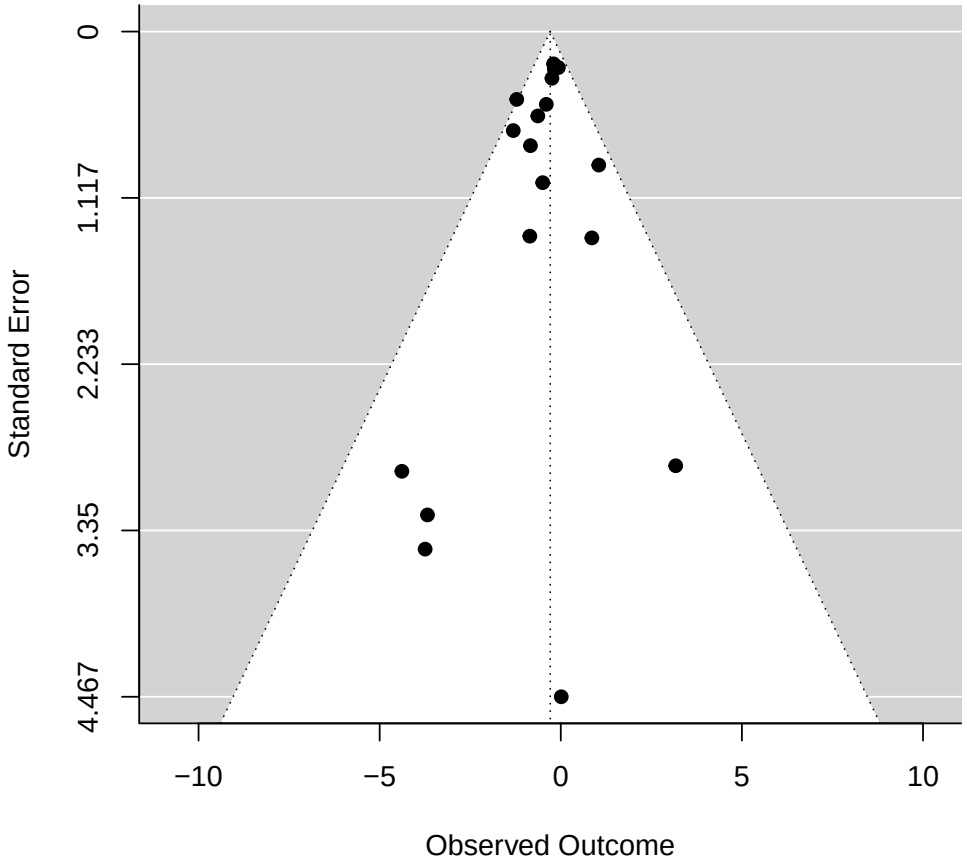
Fig. S2. Diarrhea effect estimates and compliance rates across included and excluded studies



(A) (B)

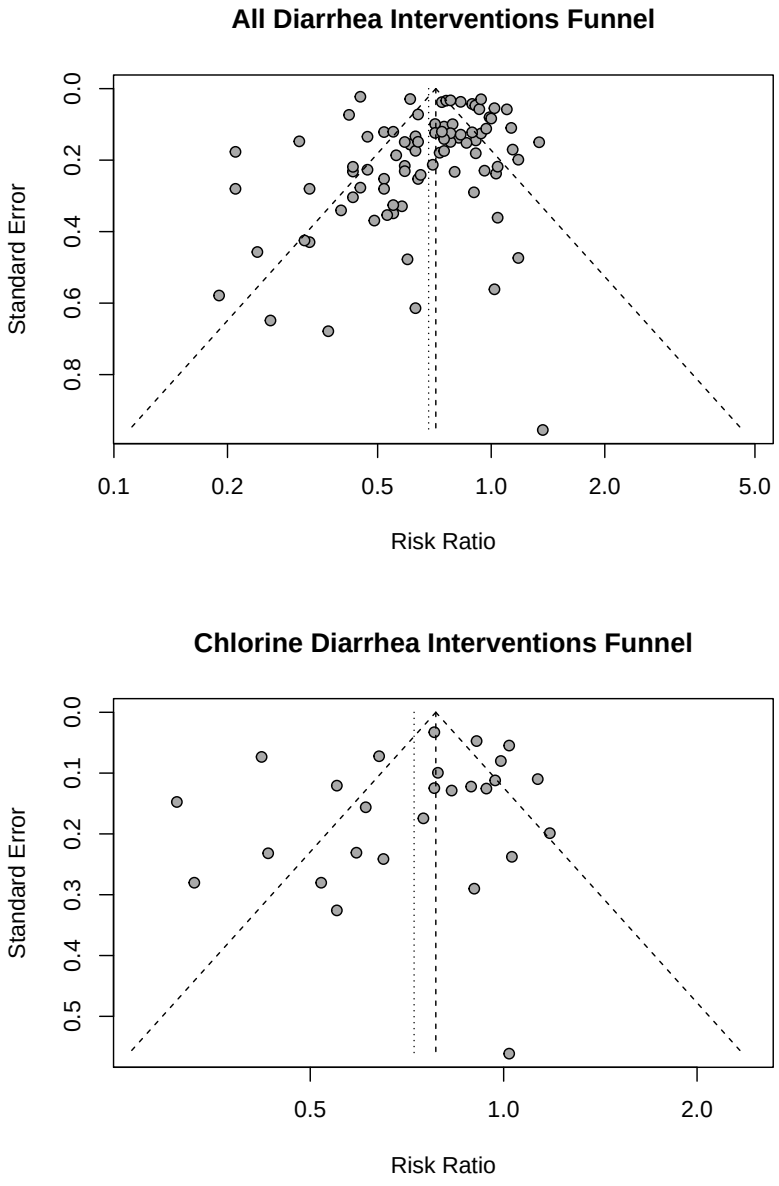
Notes: Figure (A) presents the diarrhea effect size across included (bottom panel) and excluded (top panel) studies. Figure (B) presents the compliance rate across included (bottom panel) and excluded (top panel) studies.

Fig. S3. Funnel plot to examine publication bias



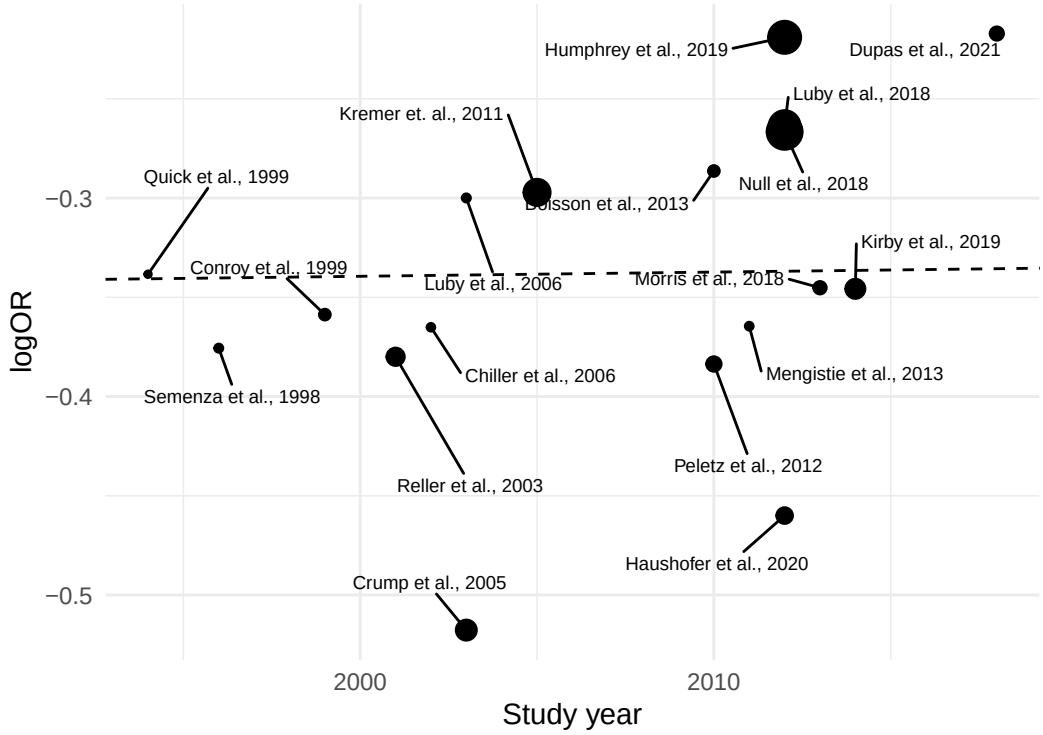
Notes: This figure presents a funnel plot. Symmetry on either side of the vertical line (representing the overall effect) suggests that publication bias is not present. Results for funnel asymmetry test are reported in Materials and Methods, Section 5.

Fig. S4. Funnel plot for all diarrhea interventions and chlorine diarrhea intervntions



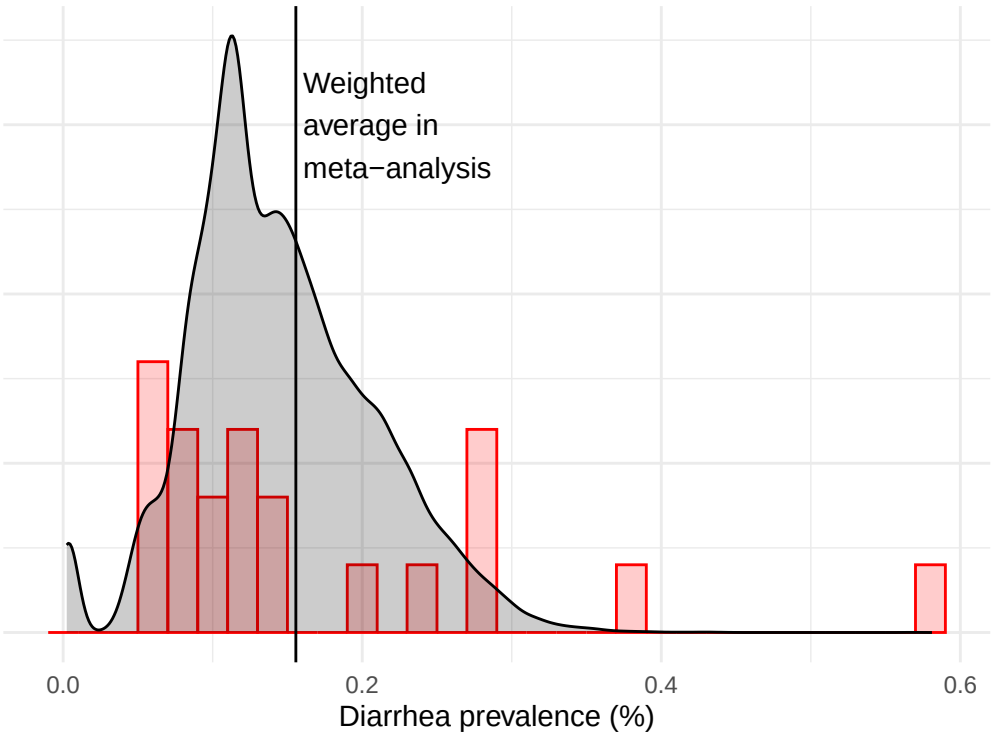
Notes: Funnel plot to assess publication bias in risk ratio estimates of diarrhea morbidity in all augmented available studies (top panel) and the subset of chlorination studies (bottom panel).

Fig. S5. Heterogeneity in treatment effects, by study year



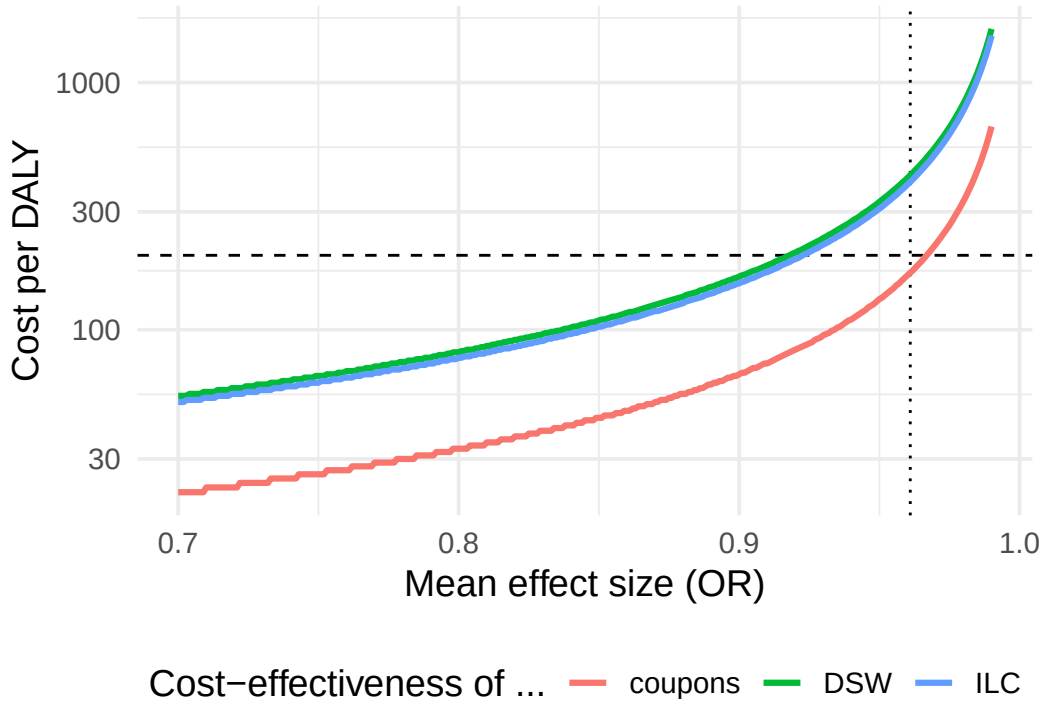
Notes: The relationship between treatment effect estimates (y axis) and the study year (x axis) across 18 studies included in the sample. Year of intervention is the year that each studied intervention was launched. We find no association (slope is less than 0.0002 per year, with SD = 0.0025) between mortality and study year. Each point represents a study. The size of each bubble is inversely proportional to the standard errors of treatment effect estimate.

Fig. S6. Diarrhea prevalence in included studies compared to distribution low- and middle- income countries



Notes: The gray area represents the density over 43,323 observations of diarrhea prevalence at the sub-national level from 94 countries, based on data from Institute for Health Metrics and Evaluation, 2020. The red histogram is of diarrhea prevalence among the studies included in this meta-analysis. The black bar shows the weighted average of diarrhea prevalence among the studies in the meta-analysis, which is 15.6%. This corresponds to the 61st percentile of the distribution of sub national diarrhea estimates.

Fig. S7. Cost-effectiveness as a function of treatment effect



Notes: The calculation from Table S7 is repeated in this plot, varying the odds ratios on all-cause mortality (X axis) and recalculating USD cost per DALY averted for each of the three methods we investigated. The vertical line corresponds to 0.039 reduction in OR, which we discuss in Section 7 of the supplement. The horizontal line is the most stringent cut-off of 200 USD.

References

1. J. Haushofer, M. Kremer, R. Maertens, B. J. Tan, “Water Treatment and Child Mortality: Evidence from Kenya” (National Bureau of Economic Research, 2021).
2. The World Bank, The World by Income and Region. *World Dev. Indic.* (February 21, 2022).
3. 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).
4. 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).
5. 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).
6. 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).
7. 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).
8. 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).
9. United Nations Children’s Fund (UNICEF) and World Health Organization, Progress on drinking water, sanitation and hygiene 2000–2017 (2019) (February 21, 2022).
10. W. Wiecek, Meta-analysis of binary data with baggr (2022) (February 21, 2022).
11. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
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. I. Andrews, M. Kasy, Identification of and Correction for Publication Bias. *Am. Econ. Rev.* **109**, 2766–2794 (2019).
14. 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).
15. 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).
16. 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).
17. Evidence Action, Dispensers for Safe Water: Cost-effectiveness Model Guide (2019).
18. 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* **15**, 272–305 (2023).

19. 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).
20. 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).
21. 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).
22. , Global Burden of Disease Results Tool. *Glob. Burd. Dis. Results Tool Inst. Health Metr. Eval. IHME* (2015) (February 21, 2022).
23. T. Spoorenberg, "Data and methods for the production of national population estimates: An overview and analysis of available metadata" (UN Population Division, 2020).
24. T. F. Clasen, *et al.*, Interventions to improve water quality for preventing diarrhoea. *Cochrane Database Syst. Rev.*, CD004794 (2015).
25. 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).
26. R. J. Evans, *Death in Hamburg: Society and Politics in the Cholera Years* (Penguin Books, Penguin Books) (February 21, 2022).
27. 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).
28. C. L. Coles, *et al.*, Nutritional Status and Diarrheal Illness as Independent Risk Factors for Alveolar Pneumonia. *Am. J. Epidemiol.* **162**, 999–1007 (2005).
29. 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).
30. 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).
31. 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).
32. 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).
33. T. J. Schuijt, *et al.*, The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. *Gut* **65**, 575–583 (2016).
34. S. Subramanian, *et al.*, Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* **510**, 417–421 (2014).
35. 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).
36. 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).
37. 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).
38. H. Ritchie, M. Roser, Clean Water and Sanitation. *Our World Data* (2021) (February 21, 2022).
39. 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).
40. 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).
41. 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).
42. 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).
43. 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).
44. 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).
45. I. A. Alam, M. Sadiq, Metal contamination of drinking water from corrosion of distribution pipes. *Environ. Pollut.* **57**, 167–178 (1989).
46. 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).
47. 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).
48. 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).
49. 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).
50. 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).
51. 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).
52. 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).
53. 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).
54. 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).
55. 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).
56. 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).
57. 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).
58. 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).
59. 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).
60. 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).
61. 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).
62. , Contra la morbilidad infantil: Filtros artesanales y educación (1995).
63. 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).
64. 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).
65. 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).
66. 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).
67. 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).
68. 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).
69. 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).
70. T. F. Clasen, S. Cairncross, Editorial: Household water management: refining the dominant paradigm. *Trop. Med. Int. Health* **9**, 187–191 (2004).

71. 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).
72. 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).
73. 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).
74. 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).
75. 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).
76. B. Torún, "Environmental and Educational Interventions against Diarrhea in Guatemala" in *Diarrhea and Malnutrition: Interactions, Mechanisms, and Interventions*, L. C. Chen, N. S. Scrimshaw, Eds. (Springer US, 1983), pp. 235–266.
77. C. J. Austin, Chlorinating household water in The Gambia in *Proceedings of the 19th WEDC International Conference, Accra, Ghana*, (1993) (February 21, 2022).
78. 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).
79. K. G. McGuigan, P. Samaiyar, M. du Preez, R. M. Conroy, High Compliance Randomized Controlled Field Trial of Solar Disinfection of Drinking Water and Its Impact on Childhood Diarrhea in Rural Cambodia. *Environ. Sci. Technol.* **45**, 7862–7867 (2011).
80. 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).
81. 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).
82. 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).
83. 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).
84. 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).
85. T. Chiller, Reducing diarrhoea in Guatemalan children: randomized controlled trial of flocculant-disinfectant for drinking water. *Bull. World Health Organ.* **84**, 28–35 (2006).
86. 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).

87. 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).
88. 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).
89. 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).
90. J. F. Morris, *et al.*, A Randomized Controlled Trial to Assess the Impact of Ceramic Water Filters on Prevention of Diarrhea and Cryptosporidiosis in Infants and Young Children—Western Kenya, 2013. *Am. J. Trop. Med. Hyg.* **98**, 1260–1268 (2018).
91. R. M. Conroy, M. E. Meegan, T. Joyce, K. McGuigan, J. Barnes, Solar disinfection of water reduces diarrhoeal disease: an update. *Arch. Dis. Child.* **81**, 337–338 (1999).
92. , "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).
93. , Child Mortality Estimates. *UN Inter-Agency Group Child Mortal. Estim.* (2020) (February 21, 2022).
94. Department of Data and Analytics, WHO, WHO methods and data sources for global burden of disease estimates 2000-2019 (December 2).
95. Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. Stat Med. 26, 53–77 (2007).
96. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. J. Stat. Soft. 36, 1–48 (2010).