Supplementary Materials: Water Treatment and Child Survival: A Meta-Analysis

Michael Kremer, Stephen Luby, Ricardo Maertens, Brandon Tan

Materials and Methods

Figs. S1 to S8

Tables S1 to S11

Table S12 Risk of Bias Table

Materials and Methods

Section 1 presents the meta-analysis models. Section 2 presents a sensitivity analysis. Section 3 presents a cost-effectiveness analysis of chlorine dispensers in Western Kenya and coupons for water treatment solutions. Section 4 presents a comparison between studies included and excluded from our analysis.

1. *Meta-analysis models*

Peto odds ratio model

The main frequentist specification is the Peto one-step odds ratio (OR) method. 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-analyse odds ratios.

However, the sample OR and/or its variance estimates are undefined when there are zero deaths in either the control or treatment group, meaning that the standard meta-analysis approach requires dropping three studies (55, 54, 19).[1] Instead, the Peto one-step method computes an approximation of the log odds ratio which allows for zero deaths in one of either the control or treatment group (59), but is not defined when both control and treatment arms have no events.

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

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

where

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

where a is the number of treatment group participants exposed to the treatment, b is the number of treatment group participants not exposed to the treatment, c is the number of control group participants exposed to the treatment, and d is the number of control group participants not exposed to the treatment; k indexes each study, and m is the total number of studies.

---

[1] Chiller et al., 2006 (18), Semenza et al., 1998 (16) report zero deaths in the treatment group, and Luby et al., 2006 (25) has zero deaths in the control group.

The canonical Peto OR specification (59) uses fixed effects. A random-effects specification may be preferable since treatment effect heterogeneity is expected due to differences across studies in ages of children, baseline child mortality rates, baseline water contamination, treatment compliance, and water treatment technologies.[2] We therefore fit a random-effect model of Peto log odd ratios by using the default Restricted Maximum Likelihood (REML) estimator as implemented in R package meta (91). We use a typical continuity correction of adding 0.5 events to each of the cells (*a,b,c,* and *d*) in the study where *a=0* and *c=0*.

Bayesian logistic meta-analysis model

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

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

where,

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

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

For the main specification, we use mildly informative priors on the hyper-parameters, similar to (16). For τ, we set a normal distribution with mean 0 and standard deviation of 10. This prior encodes the belief that causal effects should not be thought of as large unless data contains evidence to the contrary. For μ, we use a standard distribution with SD of 10, but centered at -4.59, to encode our knowledge that child mortality is a rare event (approximately $\log(0.01) =$ -4.59). For σ_τ and σ_μ, we use a zero-centered standard distribution with SD of 10, which allows for very large heterogeneity. The discussion of the Bayesian OR estimates throughout the paper refers to 95% posterior credible intervals (CrI) from Bayesian inference, which may not be symmetric.

---

[2] However, meta-analysis simulation studies show that heterogeneity may have a minor impact on estimates when rare events are considered (61, 55).

Figure 3B uses no pooling model. that is, one where σ_τ^2 is infinite (assuming that individual studies do not influence each other). In the sensitivity analyses section (see below),following the literature (16), model fit across full pooling (fixed effects) and partial pooling (random effects) specifications are compared using cross validation. Full pooling model is one where $\sigma_\tau^2 = 0$, that is, there are no differences between studies. All other priors are unchanged across no, partial, and full models. For each specification, 13 Bayesian hierarchical models are fitted to the data, leaving out one study at a time and then calculating expected log predictive density (ELPD) for each study (60). This measures the out-of-sample predictive performance of the model for each study, automatically penalizing the model for number of parameters. The ELPD averaged over the thirteen models is used as the cross-validation information criterion. A value closer to zero implies a better fit.

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

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

We report the meta-analysis weights in Table S5.

2. Sensitivity analysis

Over a set of the following 50 sensitivity analysis models (8 Peto OR models looking at lengths of follow-up, additional inverse variance model, Bayes and Peto models repeated for 15 choices of dropping one study and, 6 other choices of studies to include/exclude) the study estimates remain qualitatively similar to our main estimate. In this set of sensitivity analyses, the mean OR estimates range from 0.65 to 0.81 and are all significant, in the sense of the upper bounds on credible/confidence intervals being below 1.

The exclusion of any single particular study. The Peto odds ratio estimates and Bayesian estimates are given in Table S5. For the Bayesian model, the mean OR ranged from 0.66 to 0.76, with the lowest 95% CrI lower bound of 0.43 and the highest 95% CrI upper bound of 0.99. For Peto OR, the means ranged from 0.67 to 0.80, with the lowest lower bound of 0.49 and the highest upper bound of 0.97.

Combining studies that cover continuations of a program. Since one of the studies (18) relies on a continuation[3] of a program from another study (23), it is possible to combine the microdata

---

[3] In (18), 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 (23). 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 (23) were excluded from the sample. (18) uses

from both into one study. The two studies cover different populations, time-periods, and effects, with Haushofer et al., 2020 (18) using data on non-treated individuals collected four to five years after the initial roll-out of the program studied in Null et. al (2018) (23), and measuring the combined effect of the roll-out and continuation of the program. Nevertheless, the meta-analysis estimates remain quantitatively similar and significant with a mean reduction in mortality odds of 29-31%, depending on the model (see Table S6).

Including studies with contaminated control group. Adding the solar disinfection trial (26) to the meta-analysis results in a mean reduction in mortality odds of 28-29%; adding the blinded filtration study (19) results in a mean reduction of 26-27% (see Table S6).

Alternate control group in study with active and passive arms. In one study (23), 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 (23) for the meta-analysis, leads to a mean reduction in mortality odds of 28% in the Peto OR model (see Table S6).

Alternate treatment group in spring protection study. In one study (56), 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 (56) for the meta-analysis, the estimated mean reduction in mortality is 28% in the Peto OR model (see Table S6).

Dropping studies which combine water treatment with other interventions. Dropping studies where the water treatment intervention was combined with the provision of cookstoves (23) or other hygiene and sanitation interventions (22) leads to significant Peto OR and Bayesian OR estimates, with a mean reduction in mortality odds around 35% (Peto OR 0.66, Bayes OR 0.64), see Table S6.

Restricting to studies with longer monitoring periods. The studies included in the meta-analysis have differing lengths of follow-up, ranging from 9.5 to 260 weeks. Meta-analysis models of event data may overweight the contribution ("effort") of shorter studies. However, the weights in both the Peto OR and Bayesian logit model assigned to short studies are low, as seen in Table S5. For the Peto OR model, estimates are expected to be imprecise for studies with shorter monitoring periods, owing to the shorter period over which events can occur. In our Bayesian logit model, when mortality event rates are low, this is reflected in the model by estimation of imprecise baseline risk and, hence, imprecise estimation of treatment effects, which ultimately leads to a lower weight in the meta-analysis estimate. As a sensitivity check, we

---

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 (23), over twice as large as that analyzed in the original study.

repeat Peto OR analysis by excluding studies that are shorter than any given follow-up length in our dataset (104, 78, 65, 52, 37, 20, 13 and 9.5 weeks).

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

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

Inverse variance estimation. The inverse variance method assigns to each study a weight proportional to the inverse of the variance of the effect estimate. As a result, larger studies are given more weight than smaller studies, which have larger standard errors. To perform an inverse variance random effects estimation, we use a normal approximation of the log odds ratios and drop studies with zero deaths in either the treatment or control group: Chiller et al., 2006 (55) and Semenza et al., 1998 (54) report zero deaths in the treatment group; Luby et al., 2006 (19) reports zero deaths in the control group; Quick et al. reports zero deaths overall. The results are presented in Figure S1. Inverse variance random effects estimation implies an average reduction in all-cause odds of child mortality of 26% with random effects (OR 0.74, 95% CI 0.59 to 0.93).

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

Contribution of studies with published mortality outcomes only. At the beginning of the study, five randomized controlled trials (RCTs) were identified which reported mortality outcomes as part of their analysis (23, 29, 30, 53, 19).[4] Two of the five studies did not pre-specify mortality as an outcome, yet reported large effects on mortality in their published manuscript (20, 15).

---

[4] Papers which reported mortality in Clasen et al., 2015 (50) and other studies that we were aware of.

By including studies which did not report mortality outcomes, we are able to increase statistical power to detect significant effects in our main results.[5] Another key reason is to avoid bias if those with positive point estimates are more likely to publish.

The estimated reduction in all-cause odds of child mortality with the Peto OR model was 33%, however, the result was not significant at the 95% confidence level (Peto OR 0.67: CI 95% 0.41, 1.11). The Bayesian logistic odds ratio estimate is similar in point estimate, and the uncertainty interval includes 1 (Bayes OR 0.73: CI 95% 0.28, 1.44). Dropping the study with zero deaths in its treatment arm (19) and using inverse variance OR results in significant estimates for only the fixed effect specification (random OR 0.65: CI 95% 0.40, 1.05; fixed OR 0.73: CI 95% 0.55, 0.97), see Table S4.

Risk difference (RD) model. As discussed, RD specification is not appropriate when there are very large differences in baseline risks, as is the case in this meta-analysis: one study had no events, and another had a 10% event rate in controls. RD model is also not appropriate when probabilities are low, as this would imply that expected event rates for some studies are negative. This is also the case in our meta-analysis, since several studies had close to 0 events. However, we include results using an RD model for transparency. Fitting a Bayesian RD model we found a non-significant reduction in mortality risk of 0.3 percentage points (d = -0.003, 95% CrI -0.008, 0.001).

Fixed and random effects model. We used both full (fixed effect) and partial (random effect) pooling specifications for the Bayesian logit model. Under a fixed effect Bayesian logit model the reduction in odds was 25% (OR 0.75, 95% CrI 0.61, 0.91), compared to 30% under the random effects model. Using a leave-one-study-out cross-validation (LOO CV) procedure, the expected log predictive density for the partial pooling model was -1210 (with SE of 383) and for the full pooling model -1190 (SE of 379). This suggests no significant differences in the out-of-sample performance of both models, with slight preference for the full pooling model.

3. Cost-effectiveness analysis

The appropriate way to deliver water treatment is likely to vary by context. For example, in urban areas of middle-income countries and many densely settled rural areas it may be through piped water systems, whereas in rural areas of lower income countries where people live on their farms this may be impractical. Here we do not take on the task of identifying the best approaches, but simply use a couple of illustrative examples to argue that there is likely to be tremendous potential to cost effectively reduce child mortality. To the extent that other delivery technologies can do so more effectively, benefits will be even greater. The present estimates imply that free provision of water treatment is a very cost-effective way to reduce child mortality.

---

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

For the first estimate, we consider a chlorine dispenser program operating at scale in Western Kenya using detailed cost and take-up data. Second, we estimate the likely impact of scaling up programs distributing coupons for free dilute chlorine solution to households with young children along the lines discussed in previous studies (35, 36). We use the main meta-analysis estimate that combines water interventions. However, using Peto OR estimate obtained for the subset of chlorine interventions only (OR = 0.688) would not change our calculations substantially, decreasing USD cost per DALY to 31 USD compared to 34 in Table 2, Column 1.

Chlorine dispensers in Western Kenya

The cost effectiveness of chlorine dispensers for point-of-collection water disinfection in Western Kenya (see Table 2 Column 1) is calculated using data from Evidence Action which operates 18,405 dispensers with 1,138, 964 people using the dispensers in Western Kenya (62). Only benefits of reduced child mortality risk are included, while possible health gains through reduced child morbidity and health gains for people over the age of 5 years - such as those with suppressed immune systems (e.g., HIV+) - are ignored. Based on the mean Peto OR estimate of the effect of interventions to improve water quality on child mortality among the included studies (OR 0.72), we estimate that the expected effect of water treatment on <5y mortality, per year is 0.324 p.p. reduction, adjusting for usage rates (see Table 2, row 7). The estimated cost of installing and maintaining chlorine dispensers at scale in western Kenya is about USD 9.13 per child under five served, per year (see Table 2, row 8).[6] Thus, the cost of operations is USD 2,811 per death of a child under 5 averted (see Table 2, row 9). Assuming that a death within the first 5 years of life leads to the loss of 81.25 disability-adjusted life years (DALYs)[7], the cost of chlorine dispensers for point-of-collection water disinfection is USD 36per DALY averted (see Table 2, row 10). This cost is far lower than the Kenyan gross domestic product (GDP) per capita (about USD 1,838 in 2020), which is the threshold suggested by the WHO to determine if interventions are "highly cost-effective" and, of course, even lower than three times the GDP per capita which is the threshold to determine if interventions are "cost-effective" (80). If we assume a discount rate of 3%, death within the first 5 years of life leads to the loss of 31 DALYs and the cost of dispensers is USD 97 per DALY averted.[8]

Coupons for water treatment solution

Programs providing coupons for free water treatment solution to families with young children have so far only been conducted at modest scales (35, 36), but back of the envelope calculations suggest coupon programs would also be highly cost effective (see Table 2 Column 3). These calculations are based on rates of usage and coupon redemption in Kenya (35) and Malawi (36). Approximately 32% of all households who receive coupons treat their water with

---

[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.
[7] As recommended by the World Health Organization.
[8] Assuming a discount rate of 0% instead will increase the cost-effectiveness of chlorine dispensers.

37% of all coupons being redeemed.[9] The under 5 mortality rate among populations without access to safe drinking water is estimated as 5.02% using data from the UN Interagency Group for Child Mortality Estimation – slightly lower than that for rural Kenya used in the calculation above.[10] Based on the present estimate of the effect of interventions to improve water quality on child mortality (OR 0.72) and adjusting for usage rates, it is estimated that the program would reduce <5y mortality per year by 0.15 p.p. (see Table 2, row 7). Each 150-milliliter bottle of WaterGuard costs USD 0.30 and is enough for roughly one month's supply of treated water (for drinking and cooking) for a household. The studies that evaluated this intervention (35, 36) focused on environments where people did not have access to clean water. In scaling up such a coupon program, it may be difficult to exclude areas where much of the population already has access to clean water and it is possible that people in these areas might also treat their water. However, even considering that for every two households targeted the program covers an additional untargeted household which already has clean water, and that the administrative costs of running a coupon program were as large as the retail price of the chlorine solution, the cost of a scaled-up program would still only be USD 2,675 per death of a child under 5 averted – or USD 34 per DALY averted (see Table 2, rows 9 and 10).

Coupon programs could potentially be operated almost everywhere in the world, and rough calculations suggest that a global coupon program which provides coupons for free water treatment solution to all families with under 5 children without access to safe drinking water in low- and middle-income countries could avert up to half a million under five deaths each year (see Table S7). 2.2 billion people around the world do not have access to safely managed drinking water services (1), see row 1.[11] This number is similar in magnitude to the global estimates from other studies (37, 38). Of this population approximately 274 million are children under the age of 5 (2, 63).[12] The under 5 mortality rate among populations without access to safe drinking water implies 2.74 million deaths per year in absence of water treatment. Based on the meta-analysis estimate of the effect of interventions to improve water quality on child mortality (OR 0.72) and adjusting for coupon usage rates, it is estimated that a program that targets this population would save approximately half a million under five lives at a cost of approximately a billion USD each year.

Using alternative estimates that 1.8 billion people globally use a source of drinking water which suffers from fecal contamination (37), the total cost of the program would be USD 809 million and 461,000 under five lives would be saved annually. Alternatively, a coupon program targeting the 1.1 billion people estimated as only having access to drinking water that is of at least 'moderate' risk (>10 E. coli or triphenyltetrazolium chloride (TTC) per 100 ml) (37) would save approximately one-quarter of a million lives at an annual cost of around a half billion

---

[9] Average across Dupas et al., 2016 (35) and Dupas et al., 2020 (36) who find water treatment rates of 34.5% and 30.0% and coupon redemption rates of 41.1% and 33.4%, respectively.

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

[11] 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.

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

dollars. If coverage of the program was restricted to the twenty countries[13] where Population Services International already markets hypochlorite solution at scale, approximately one-eighth of a million deaths would be averted annually at a cost of around a quarter billion USD.

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

4. *Comparison of characteristics between included and excluded studies*

We additionally compare some key characteristics of the water treatment studies included with those excluded from the analysis, but included in (9). There were 73 studies in (9), yielding 80 observations. Some studies had multiple observations on account of multiple study locations, and hence yielded multiple effect estimates. 7 of these observations were included in our meta-analysis, resulting in 73 observations excluded from our meta-analysis and 15 observations included.

- The distribution of effect estimates of water treatment on diarrhea and compliance rates are similar across included and excluded data (see Fig. S4).
- 47 out of 73 observations (64%) are conducted in a rural setting, with 15.1% and 20.5% being conducted in mixed and urban settings respectively. Similar to this, among the included studies, 73.3% (11 out of 15 studies) of the studies are set in rural areas and the proportion of the studies conducted in mixed and urban settings is 13.3% for both.
- In terms of the water source, the primary source of water at baseline (or in the control group) was an unimproved water source in 49 out of 73 observations (67%.) This is comparable to 86.6% (13 out of 15 studies) among the included studies.

5. *Exploratory simulation of small-study publication bias*

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

---

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

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

6. *Exploratory assessment of power to detect heterogeneous effects*

As shown in Fig S5-S9, univariate meta-regressions did not find statistically significant linear relationships between five predictors and treatment effect estimated using the frequentist model. However, given small sample size and uncertain estimates in individual studies a meta-regression model that would typically be used in such situations may not have sufficient power to detect linear relationships between the predictors and the treatment effects. To assess this, we conducted a simple post-hoc exploratory analysis of whether a meta-regression model would have sufficient power to detect the relationship between treatment effects and threecontinuous predictors: prevalence of diarrhea, compliance, and year of implementation. The linear relationship between the first two is easiest to hypothesize, since at x=0 (no compliance, no prevalence) we would expect the true effect to be 0; we include 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, compliance, prevalence, or year of implementation. We parameterise these such that the expected average effect in the population corresponds to the estimated mean OR. In the case of prevalence we set slope ($y=ax$) to a = -0.335*x/0.183, where -0.335 is log of 0.7149, i.e. the OR estimated by the Peto OR model; 0.183 is population weighted prevalence across 15 studies. In the case of compliance we set it to a = -0.335*x/0.462, where 0.462 is the population weighted compliance across 14 studies (see Table 1), since 1 study did not report compliance. In the case of year of implementation we set y = 0.0335*(x-2010) - 0.335, that is, we assume that in 2010 (weighted average of year of implementation in 15 studies) the mean effect was log of 0.7149 and it decreased linearly to the point where by 2020 half of the effect disappears, which we consider a very strong effect.

We simulate new datasets with some noise ($y = ax + epsilon$), using observed compliance/prevalence values for *x* and for epsilons using posterior SDs for 15 studies from the main Bayesian model, to obtain a crude but realistic estimate of variation in each study. For each simulated dataset we fit a univariate linear regression model and check if the coefficient is significant. We repeat this 10,000 times. Simulated power is the fraction of coefficients that were significant. We find it to be 43% for compliance, 51% for diarrhea prevalence, and 29% for year of implementation. This should be seen as the best case scenario, since we assumed no confounding and a strictly linear relationship. This indicates that there is not sufficient power to detect the relationship between compliance, year of implementation, or prevalence and mortality, even under the assumption of strong effects.

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

Study	Odds Ratio	OR [95%-CI]	Weight
intervention = Chlorination			
Crump et al., 2005		0.30 [0.13; 0.69]	6.7%
Reller et al., 2003		0.50 [0.17; 1.48]	4.2%
Boisson et al., 2013		1.97 [0.18; 21.76]	0.9%
Humphrey et al., 2019		0.94 [0.63; 1.41]	20.7%
Luby et al., 2018		0.86 [0.54; 1.37]	17.3%
Null et al., 2018		0.83 [0.55; 1.25]	20.3%
Haushofer et al., 2020		0.31 [0.13; 0.73]	6.4%
Dupas et al., 2021		2.56 [0.50; 13.24]	1.9%
Random effects model		0.70 [0.49; 1.01]	78.3%
intervention = Filtration			
Peletz et al., 2012		0.47 [0.11; 1.95]	2.5%
Kirby et al., 2019		0.70 [0.28; 1.71]	5.9%
Random effects model		0.62 [0.29; 1.33]	8.3%
intervention = Spring protection			
Kremer et. al., 2011		0.81 [0.47; 1.41]	13.3%
Random effects model		0.81 [0.47; 1.41]	13.3%
Random effects model		**0.74 [0.59; 0.93]**	**100.0%**

0.1   0.5 1 2   10

Note: Dots and horizontal lines represent point estimates and 95% confidence intervals from individual studies, respectively. The area of the square around each dot represents the weight given to each study in the fixed-effects estimation. Diamonds are centered around the random-effects estimate (by intervention type or overall), their width indicate the 95% confidence interval. 3 studies are dropped due to zero deaths in either the treatment or control group.

1

Fig. S2. Restricting set of studies to longer follow-up lengths
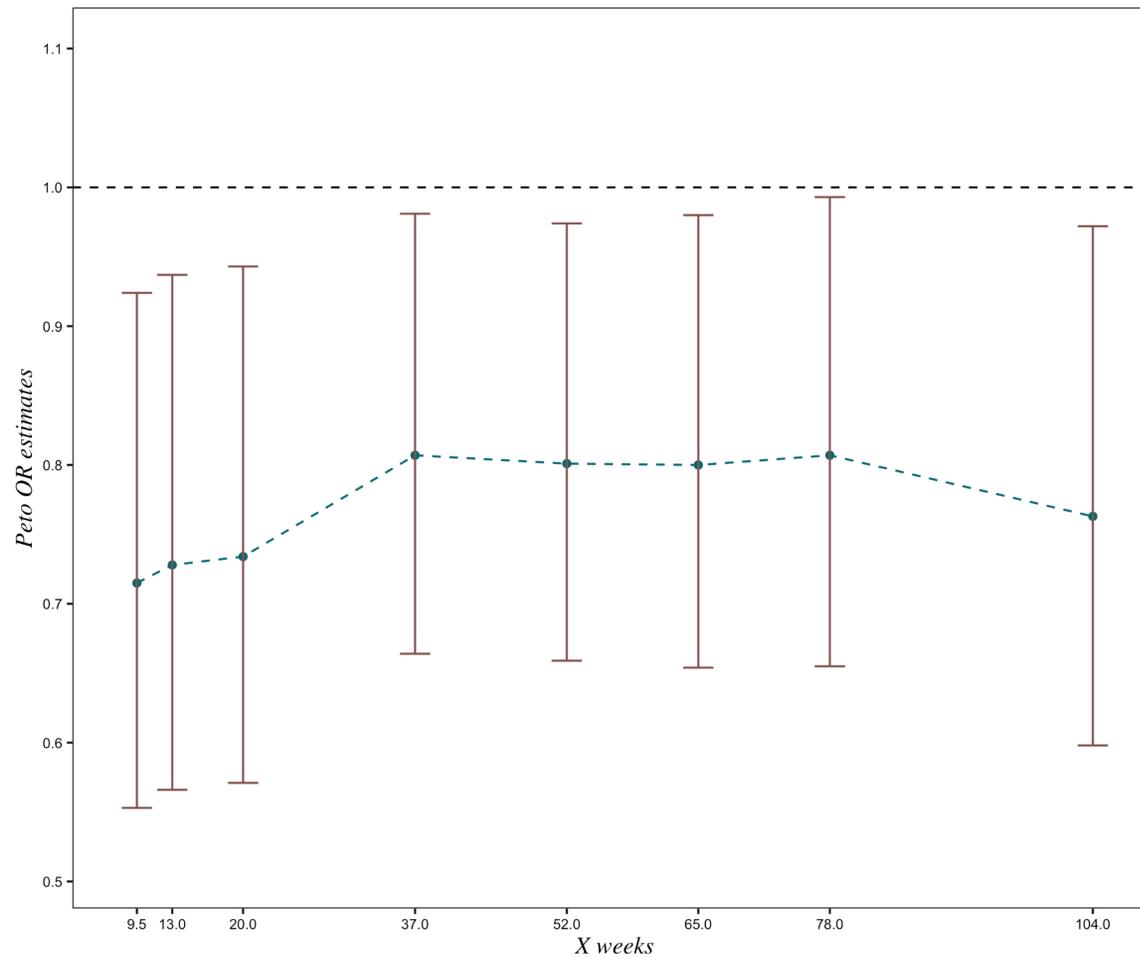
Notes: This figure presents the odds ratio estimated by the frequentist (Peto) meta-analysis model with studies shorter than X weeks removed. Each point is the Peto OR estimate, and the bars represent the 95% Confidence Interval for each estimate. All 13 studies in the main sample are included for X = 9.5 weeks, and 4 studies are included for X = 104 weeks (2 years).

Fig. S3. Funnel plot to examine publication bias
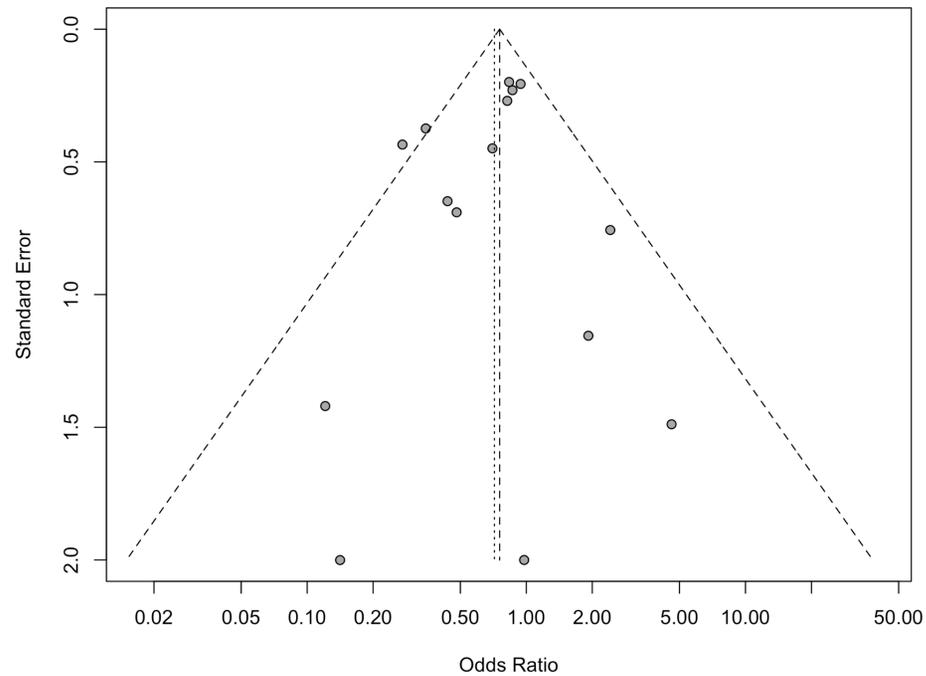
Notes: This figure presents a funnel plot. Symmetry on either side of the vertical line (representing the overall effect) suggests that publication bias is not present. The p-value corresponding to the Begg and Mazumdar test is 0.7839. Two studies - Crump et al., 2005 (30) and Haushofer et al. 2020 (18) fall outside the funnel.

Fig. S4. Diarrhea effects and compliance rates across included and excluded studies

(A)

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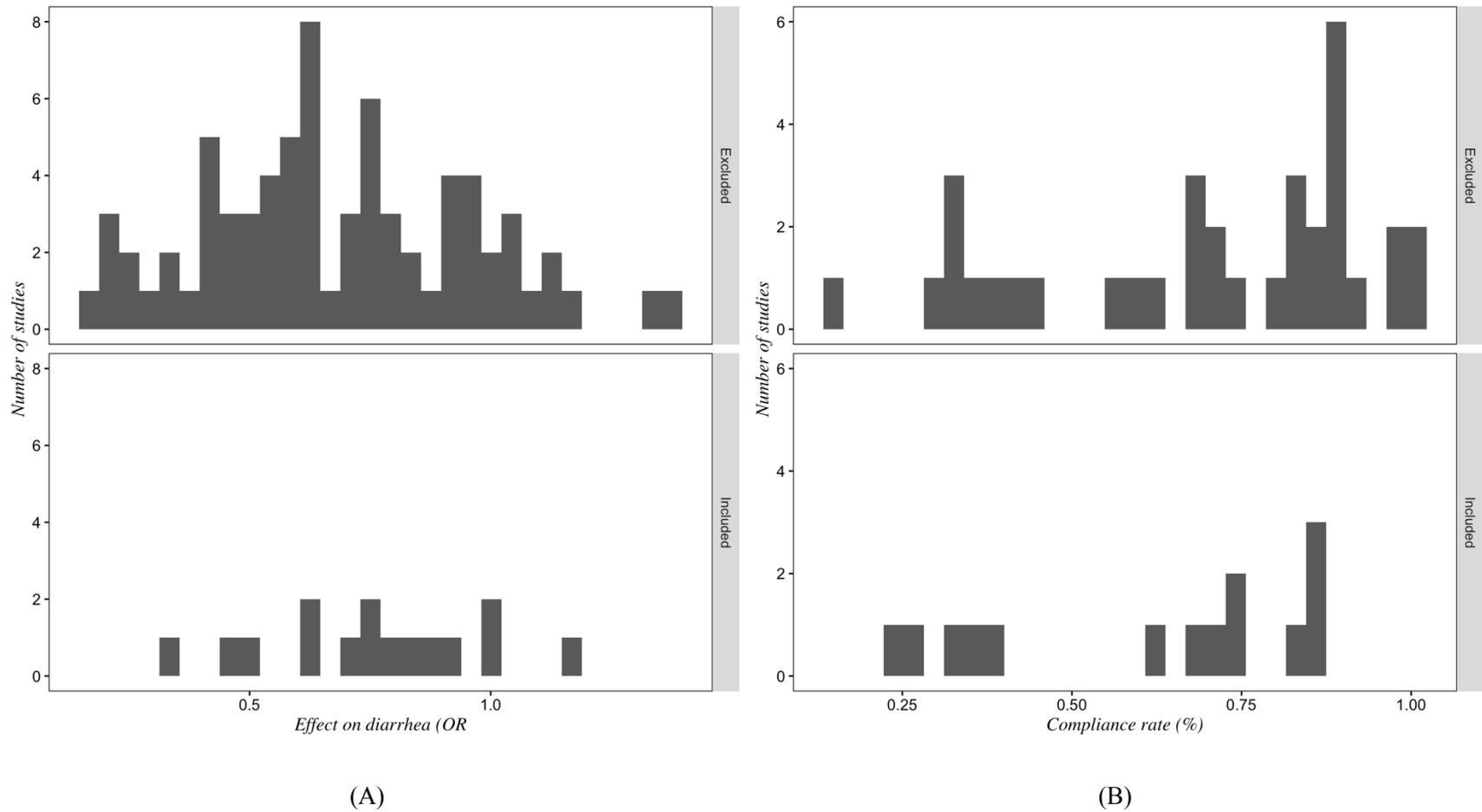
Notes: Figure (A) presents the diarrhea effect size across included (bottom panel) and excluded (top panel) studies, Figure (B) presents the compliance rates across included (bottom panel) and excluded (top panel) studies

Fig. S5. Heterogeneity in treatment effects, by diarrhea prevalence
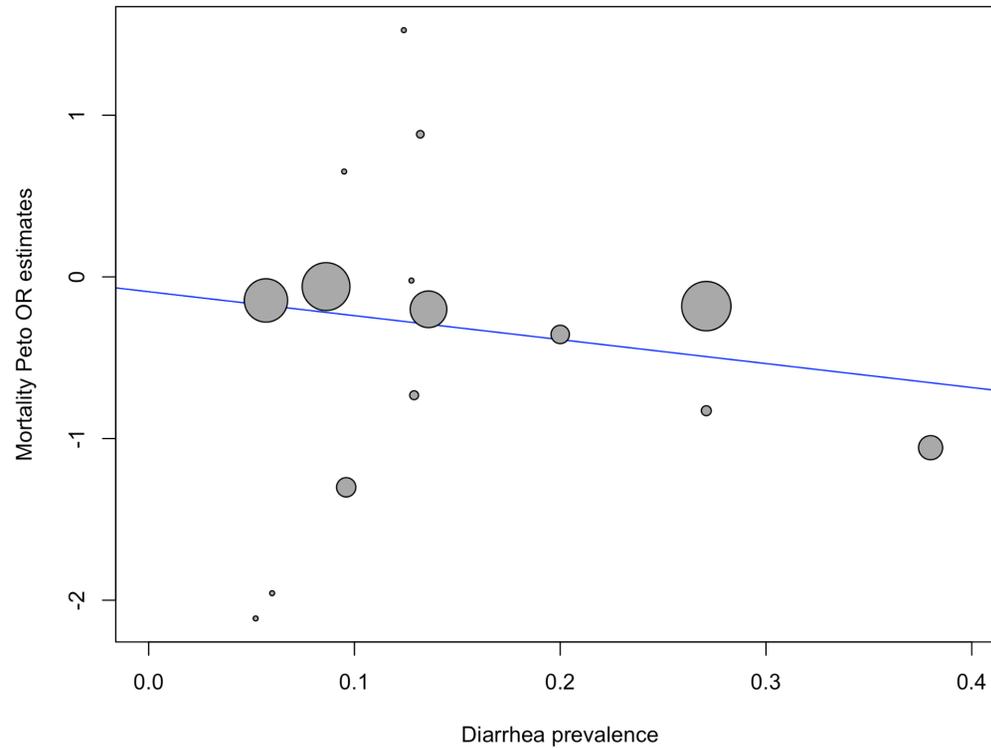
Notes: This figure presents the relationship between mortality Peto Odds Ratio estimates and the level of diarrhea prevalence across 15 studies in the sample. We find no significant differences (slope of -1.483 per unit increase in diarrhea prevalence rate, pval 0.252) in effect estimates by prevalence of diarrhea. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study.

Fig. S6. Heterogeneity in treatment effects, by level of compliance
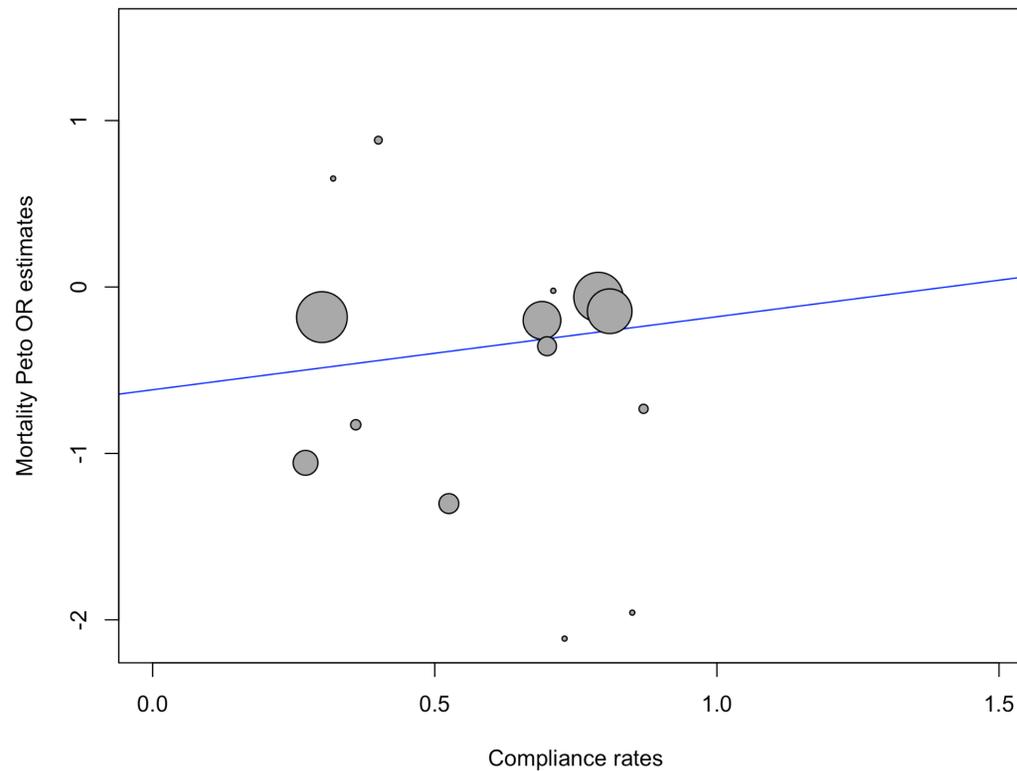
Notes: This figure presents the relationship between mortality Peto Odds Ratio estimates and the level of compliance across 14 studies in the sample (one study did not report any measure of compliance). We find no significant differences (slope of 0.439 per unit increase in compliance rate, pval 0.6942) in effect estimates by the level of compliance. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study..

Fig. S7. Heterogeneity in treatment effects, by unit of randomization
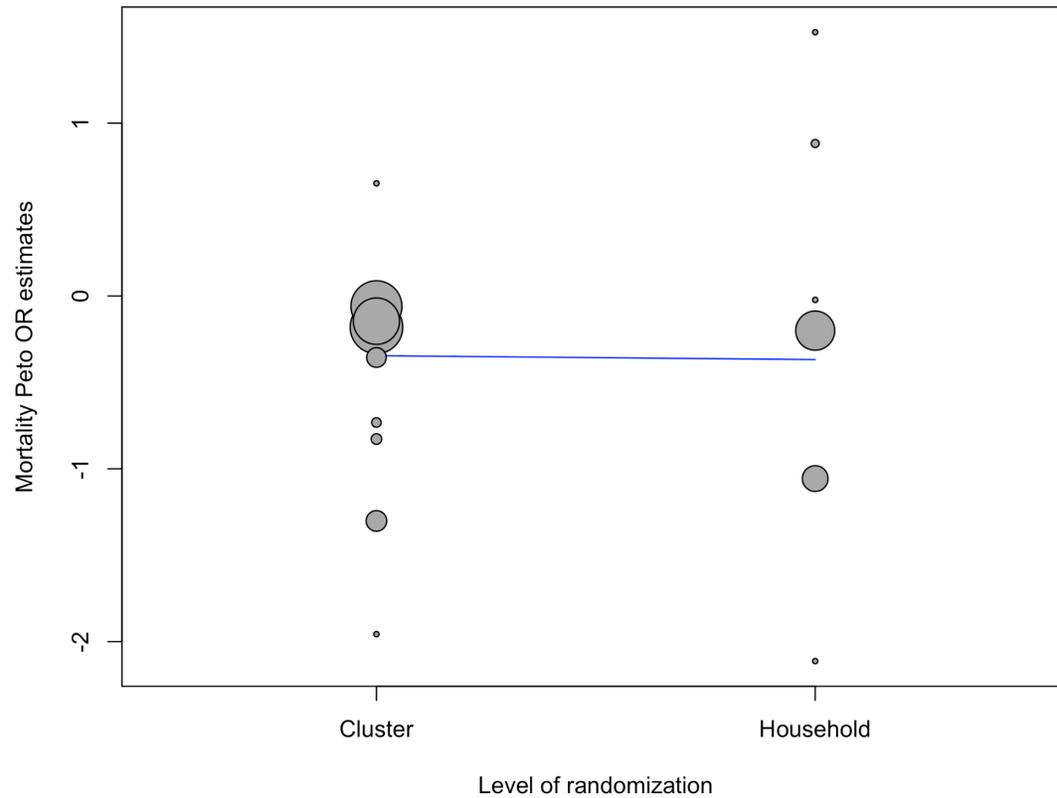
Notes: This figure presents the relationship between mortality Peto Odds Ratio estimates and the unit of randomization across 15 studies in the sample. We find no significant differences (Decrease of -0.023 for randomizing at the household level, p-value = 0.944) in effect estimates by the unit of randomization. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study.**Fig. S8. Heterogeneity in treatment effects, by study year**
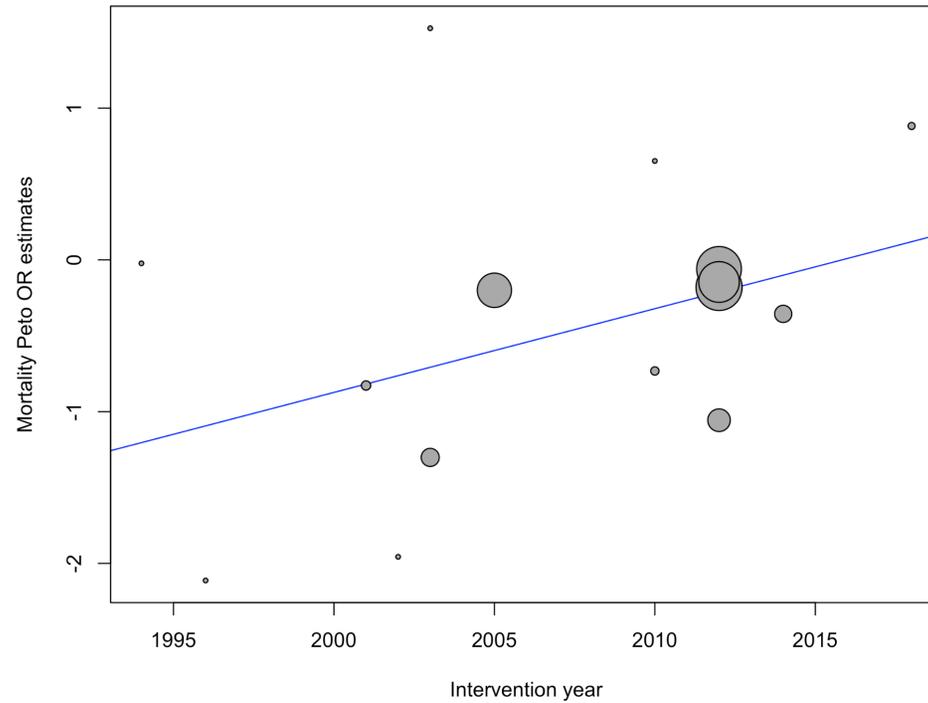
Notes: This figure presents the relationship between mortality Peto OR estimates and the study year across 15 studies in the sample. Year of Intervention is the year the study's intervention was launched. We find no significant association (slope of 0.055 per year, p-value = 0.056) between mortality and study year. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study.

Fig. S9. Heterogeneity in treatment effects, by diarrhea effect estimates
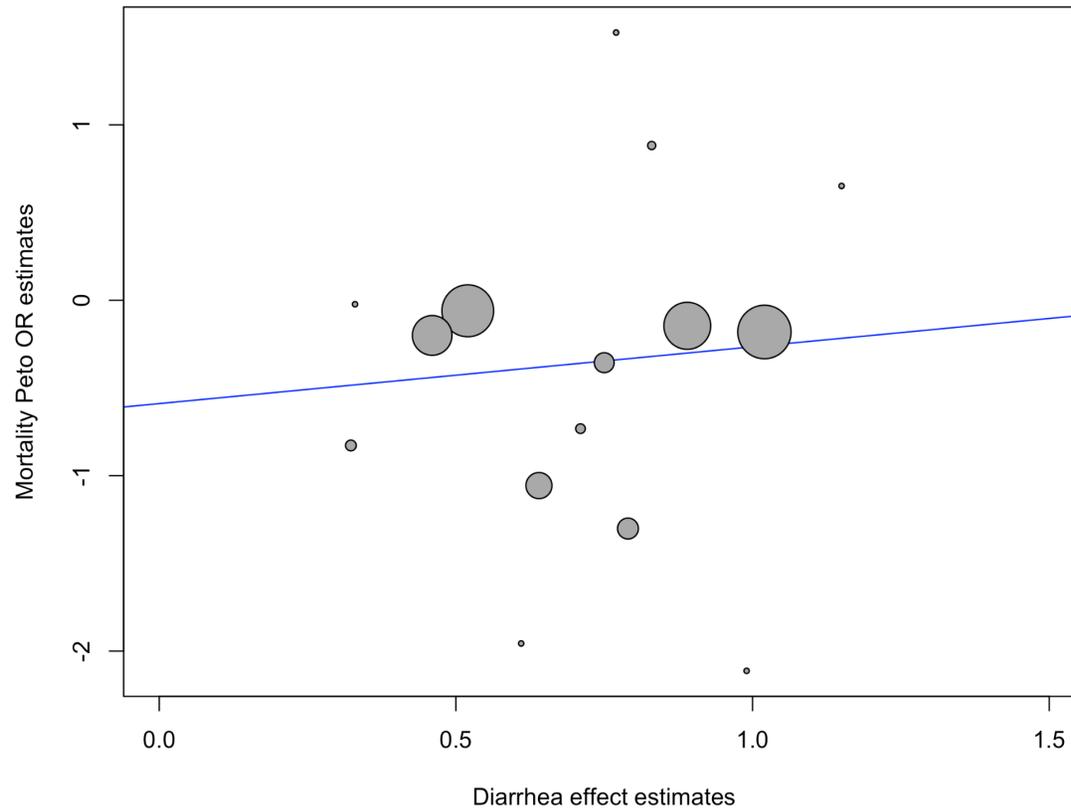
Notes: This figure presents the relationship between mortality Peto Odds Ratio estimates and the diarrhea effect estimates across 15 studies in the sample. We find no significant association (slope of 0.324 per unit increase in the diarrhea effect estimate, p-value = 0.644) between mortality and diarrhea effect estimates. Each point represents a study. The size of the bubble is inversely proportional to the variance of the estimated Peto Odds ratio for each study**Table S1. Search strategy and search terms**

Search set	Embase (Ovid)	Pubmed	Scopus	Cochrane Library
Water Quality				
1	((Water adj3 (treatment or quality or cleaning or purif* or chlorin* or decontamination or filt* or disinfect* or floccul* or storage or recontamination or re-contamination)).mp. or exp water quality/ or exp water management/) and ((water.mp. or exp water/) adj3 (drinking or consumption).mp)	((treatment[tw] OR quality[tw] OR cleaning[tw] OR purif*[tw] OR chlorin*[tw] OR decontamination[tw] OR filt*[tw] OR disinfect*[tw] OR floccul*[tw] OR storage[tw] OR recontamination[tw] OR "re-contamination"[tw]) OR "Water Quality"[MeSH] OR "Water Purification"[MeSH]) AND ((water[tw] OR water[MeSH]) AND (drinking[tw] OR consumption[tw]))	TITLE-ABS-KEY (water W/3 (treatment OR quality OR cleaning OR purif* OR chlorin* OR decontamination OR filt* OR disinfect* OR floccul* OR storage OR recontamination OR "re-contamination")) AND (TITLE-ABS-KEY (water W/3 (drinking OR consumption)))	((water near/3 (treatment or quality or cleaning or purif* or chlorin* or decontamination or filt* or disinfect* or floccul* or storage or recontamination or "re-contamination")):ti,ab,kw or MeSH descriptor: [Water] explode all trees or MeSH descriptor: [Water Quality] explode all trees or MeSH descriptor: [Water Purification] explode all trees) and ((Drinking or consumption) near/3 water):ti,ab,kw
Water Access				
2	(Water adj3 (supply or availability or access or connect* or distance or improved or distribut* or quantity or volume)).mp or exp water supply/	(water[tw] AND (supply[tw] OR availability[tw] OR access[tw] OR connect*[tw] OR distance[tw] OR improved[tw] OR distribut*[tw] OR quantity[tw] OR volume[tw])) OR "Water Supply"[MeSH]	TITLE-ABS-KEY (water W/3 (supply OR availability OR access OR connect* OR distance OR improved OR distribut* OR quantity OR volume))	(Water near/3 (supply or availability or access or connect* or distance or improved or distribut* or quantity or volume)):ti,ab,kw or MeSH descriptor: [Water Supply] explode all trees
Sanitation				
3	toilet*.mp. or latrine*.mp. or pit.mp. or pits.mp. or sanita*.mp. or ecosan.mp. or sewage.mp. or sewer$1.mp. or sewerage.mp. or exp sewage/ or open defecation.mp or (((feces or faeces or fecal or faecal or excre* or waste).mp.	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	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	(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]

10

or exp feces/) adj3 (disposal or manag* or service*).mp.) or exp sanitation/ or exp environmental sanitation/	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]	faecal OR excre* OR waste) W/3 (disposal OR manag* OR service*)))	explode all trees or MeSH descriptor: [Toilet Training] explode all trees or MeSH descriptor: [Sanitation] explode all trees or MeSH descriptor: [Feces] explode all trees or MeSH descriptor: [Sewage] explode all trees
Diarrhoeal disease			
4 (((f?ecal adj1 coliform$1) or bacterial or microbiological or viral or diarrh?ea? or intestinal or enteric or gastro-enteric or protozoa$1 or waterborne or water-borne or enterovirus or "enteric virus" or poliovirus or rotavirus or norovirus or "norwalk-like virus" or hepatitis or campylobacter or helicobacter or legionellos$ or vibrio or cholera or escherichia or salmonell$ or shigell$ or cryptosporidi$).mp. or exp diarrhea/) and (disease$1 or infection$1 or episode$1 or illness$2).mp	("fecal coliform"[tw] OR "fecal coliforms"[tw] OR "faecal coliform"[tw] OR "faecal coliforms"[tw] OR bacterial[tw] OR microbiological[tw] OR viral[tw] OR diarrhoea*[tw] OR diarrhea*[tw] OR intestinal[tw] OR enteric[tw] OR "gastro-enteric"[tw] OR protozoa*[tw] OR waterborne[tw] OR "water-borne"[tw] OR Diarrhea[MeSH] OR enterovirus[tw] OR "enteric virus"[tw] OR poliovirus[tw] OR 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	TITLE-ABS-KEY (((fecal OR faecal) PRE/1 coliform*) OR bacterial OR microbiological OR viral OR diarrhoea* OR diarrhea* OR intestinal OR enteric OR "gastro-enteric" OR protozoa* OR waterborne OR "water-borne" OR enterovirus OR "enteric virus" OR poliovirus OR rotavirus OR norovirus OR "norwalk-like virus" OR hepatitis OR campylobacter OR helicobacter OR legionellos* OR vibrio OR cholera OR escherichia OR salmonell* OR shigell* OR cryptosporidi*) AND (TITLE-ABS-KEY (disease* OR infection*	((((fecal or faecal) next coliform*) or bacterial or microbiological or viral or diarrhoea* or diarrhea* or intestinal or enteric or gastro-enteric or protozoa* or waterborne or water-borne or enterovirus or enteric virus or poliovirus or rotavirus or norovirus or norwalk-like virus or hepatitis or campylobacter or helicobacter or legionellos* or vibrio or cholera or escherichia or salmonell* or shigell* or cryptosporidi*):ti,ab,kw or MeSH descriptor: [Diarrhea] explode all trees) and (disease* or infection* or episode* or illness*):ti,ab,kw

	salmonell*[tw] OR shigell*[tw] OR cryptosporidi*[tw]) AND (disease*[tw] OR infection*[tw] OR episode*[tw] OR illness*[tw])	OR episode* OR illness*))	

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) (64), Colford (2005) (65), Rodrigo (2011) (66)
Not a randomized control trial	Kirchhoff (1985) (67), Alam (1989) (68), Mahfouz (1995) (69), Conroy (1996) (70), Xiao (1997) (71), Quick (2002) (58), Jensen (2003) (72), Majuru (2011) (73), Johri et al. (2019) (74), Reese et al. (2019) (75)
Does not include children under 5 years in age	Abebe (2014) (76)
Authors responded but no mortality data collected	Gruber (2013) (77), Günther (2013) (78), Jain (2010) (79), Opryszko (2010a, b, c) (80), Patel (2012) (81), Roberts (2001) (82), Tiwari (2009) (83), URL (1995a, b) (84), Boisson (2009) (85), Doocy (2006) (86), Stauber (2009, 2012a, b) (87) (88) (89), Lindquist (2014a, b) (90) (91), Fabiszewski (2012) (92), Clasen (2004b, c) (93) (94), Pickering et al. (2019) (95), Handzel (1998) (96),
Authors responded and mortality data was collected but no longer available	Gasana (2002) (97), Brown (2008) (98)
Authors did not respond	Torun (1982)*, Austin (1993a,b) (99) Mengistie (2013) (100), McGaugan (2011), Mäusezhal (2009) (101), Lule (2005) (102), du Preez (2008, 2010) (103) (104)

Note: *Only author died.

Table S3. Meta-analysis estimates of the child mortality impact of water quality interventions

	Meta-analysis Estimates			
	Mean Peto OR (1)	Mean bayesian OR (2)	Mean Peto OR (3)	Mean Bayesian OR (4)
ITT effect on child mortality	0.72	0.70	0.69	0.68
CI 95%	(0.55,0.92)	(0.49,0.92)	(0.47,1.01)	(0.37,1.03)
Interventions	All	All	Chlorine	Chlorine
Studies	15	15	12	12

Notes: Columns 1 and 2 report the Peto odds ratio and Bayesian logistic odds ratio respectively, for all 15 studies. Columns 3 and 4 report the Peto odds ratio and Bayesian logistic odds ratio respectively, for a subset of 12 chlorine intervention studies.

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

	Crump et al., 2005 (30) (1)	Luby et al., 2006 (19) (2)	Peletz et al., 2012 (53) (3)	Luby et al., 2018 (29) (4)	Null et al., 2018 (23) (5)	Random-Effects OR (6)	Fixed-Effects OR (7)	Mean Bayesian OR (8)	Mean Peto OR (9)
Panel A: Bayesian Odds ratio									
ITT effect on child mortality	0.55	NA	0.67	0.78	0.77	0.65	0.73	0.73	
CI 95%	(0.55,0.89)	NA	(0.67,1.25)	(0.78,1.16)	(0.77,1.08)	(0.40,1.05)	(0.55,0.97)	(0.28,1.44)	
Panel B: Peto Odds ratio									
ITT effect on child mortality	0.27	4.60	0.48	0.86	0.83				0.67
CI 95%	(0.12,0.64)	(0.25,85.10)	(0.12,1.86)	(0.55,1.36)	(0.56,1.23)				(0.41,1.11)
Obs.	1538	1548	121	1962	3699				

Notes: Panel A Columns 1 to 5 report odds ratio estimates for individual studies. There is no estimate for Luby et al., 2006 (29) due to zero deaths in the control group. Panel B Columns 1 to 5 report Peto odds ratio estimates for individual studies. Column 6 reports the odds ratio random effects inverse variance meta-analysis estimate including the studies from columns 1 through 5 except Luby et al., 2006 (29). Column 7 reports the odds ratio fixed effects inverse variance meta-analysis estimate including the studies from columns 1 through 5 except Luby et al., 2006 (29). Column 8 reports the Bayesian odds ratio meta-analysis estimate including the studies from columns 1 through 5. Column 9 reports the Peto odds ratio meta-analysis estimate including the studies from columns 1 through 6.

	Crump et al., 2005 (30) (1)	Haushofer et al., 2020 (18) (2)	Luby et al., 2006 (19) (3)	Luby et al., 2018 (29) (4)	Null et al., 2018 (23) (5)	Peletz et al., 2012 (53) (6)	IV Random-Effects OR (7)	IV Fixed-Effects OR (8)	Mean Bayesian/Peto OR (9)
Panel A: Bayesian Odds ratio									
ITT effect on child mortality	0.55	0.55		0.78	0.77	0.67	0.67	0.67	0.61
CI 95%	(0.55,0.89)	(0.55, 0.89)		(0.78,1.16)	(0.77, 1.08)	(0.67, 1.25)	(0.34,0.92)	(0.52,0.88)	(0.27,1.10)
Panel B: Peto Odds ratio									
ITT effect on child mortality	0.27	0.35	4.6	0.86	0.83	0.48			0.59
CI 95%	(0.12,0.64)	(0.17,0.72)	(0.25,85.10)	(0.55,1.36)	(0.56, 1.23)	(0.12,1.86)			(0.37,0.96)
Obs.	1538	1981	1548	1962	3699	121			

Notes: Panel A Columns 1 to 6 report odds ratio estimates for individual studies. There is no estimate for Luby et al., 2006 (29) due to zero deaths in the control group. Panel B Columns 1 to 6 report Peto odds ratio estimates for individual studies. Column 7 reports the odds ratio random effects inverse variance meta-analysis estimate including the studies from columns 1 through 5 except Luby et al., 2006 (29). Column 8 reports the odds ratio fixed effects inverse variance meta-analysis estimate including the studies from columns 1 through 5 except Luby et al., 2006 (29). Column 9reports the Bayesian and Peto meta-analysis estimates obtained from using the studies from columns 1 through 6. Column 10 reports the Peto odds ratio meta-analysis estimate including the studies from columns 1 through 6.

Table S5. Sensitivity of main results to dropping each study

	Meta-analysis effect after dropping one study														
	Haushofer et al., 2020	Luby et al., 2018	Null et al., 2018	Kremer et. al., 2011	Humphrey et al., 2019	Kirby et al., 2019	Dupas et al., 2021	Reller et al., 2003	Boisson et al., 2013	Peletz et al., 2012	Luby et al., 2006	Quick et al., 1999	Crump et al., 2005	Chiller et al., 2006	Semenza et al., 1998
Panel A: Bayes Odds Ratio															
ITT effect on child mortality	0.80	0.68	0.68	0.69	0.67	0.71	0.69	0.73	0.70	0.72	0.70	0.71	0.80	0.72	0.73
CI 95%	(0.66,0.97)	(0.49,0.93)	(0.49,0.94)	(0.50,0.94)	(0.49,0.90)	(0.53,0.94)	(0.53,0.90)	(0.56,0.95)	(0.54,0.92)	(0.554,0.943)	(0.54,0.91)	(0.55,0.92)	(0.66,0.97)	(0.56,0.93)	(0.57,0.94)
Weights in meta-analysis (%)	8.8	15.7	17.9	13.3	17.4	6.7	2.7	3.6	1.2	3.2	0.8	0.4	7.0	0.4	0.8
Panel B: Peto Odds Ratio															
ITT effect on child mortality	0.77	0.69	0.69	0.70	0.67	0.70	0.69	0.73	0.70	0.73	0.70	0.72	0.78	0.72	0.73
CI 95%	(0.54,1.01)	(0.45,0.99)	(0.44,0.99)	(0.45,1.01)	(0.42,0.96)	(0.45,0.99)	(0.46,0.92)	(0.48,1.01)	(0.46,0.95)	(0.50,0.99)	(0.46,0.93)	(0.50,0.96)	(0.56,1.028)	(0.504,0.96)	(0.51,0.99)
Weights in meta-analysis (%)	7.3	15.5	17.6	12.7	17.8	6.9	2.8	5.2	1.4	3.4	0.5	0.3	7.6	0.6	0.5

Notes: Columns 1 through 13 report meta-analysis estimates of OR obtained by excluding the study in the column heading from the full sample. Panel A reports Bayesian odds ratio estimates, and Panel B reports Peto odds ratio estimates. Row 3 of each panel reports the weight of each study in the meta-analysis from Table S3.

Table S6. Additional sensitivity checks

	Combining studies that cover related programs (23, 18)		Including study with contaminated control group I (26)		Including study with contaminated control group I (24)		Alternate control in study with active and passive arms (23)		Alternate treatment in spring protection (56)		Studies where water treatment was combined with another intervention (21, 22)	
	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR	Mean Peto OR	Mean Bayesian OR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ITT effect on child mortality	0.71	0.69	0.72	0.71	0.74	0.73	0.72	0.70	0.72	0.70	0.66	0.64
CI 95%	(0.56, 0.91)	(0.48, 0.92)	(0.57, 0.92)	(0.49, 0.92)	(0.57, 0.96)	(0.50, 0.97)	(0.55, 0.93)	(0.47, 0.94)	(0.56, 0.92)	(0.48, 0.91)	(0.47, 0.93)	(0.41, 0.89)
p-value	0.006		0.01		0.022		0.014		0.009		0.02	

Notes: Columns 1 and 2 present meta-analysis estimates combining Null et al., 2018 (23) and Haushofer et al., 2020 (18) into a single study. Columns 3 and 4 present meta-analysis estimates including du Preez et al., 2011 (26). Column 5 and 6 present meta-analysis estimates including Boisson et al., 2010 (19). Columns 7 and 8 present meta-analysis estimates using only the active control group in Null et al., 2018 (23). Columns 9 and 10 present meta-analysis estimates including both those who received spring protection in year 1 and 2 into the treatment group in Kremer et al., 2011 (56). Columns 11 and 12 present meta-analysis estimates by dropping studies (23,22) where the water treatment intervention was combined with other interventions (cookstoves, sanitation and hygiene).

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

Target Population	Source	(1) Population size (millions)	(2) # of <5y children without access to safe drinking water (millions)	(3) Number of deaths among <5y without access to safe drinking water per year (millions)	(4) Cost of providing coupons to <5 population without access to safe drinking water per year ($ millions)	(5) Total <5y lives saved per year (thousands)
Population without access to safely managed drinking water services	WHO/UNICEF, 2019 (1)	2200	272.8	2.739	10921.2	563
Population using a source of drinking water which suffers from fecal contamination	Bain et al., 2014 (37)	1800	223.2	2.241	892.8	461
Population using a source of drinking water with >10 E coli or TTC per 100 ml	Bain et al., 2014 (37)	1100	136.4	1.369	545.6	282
Population without access to safely managed drinking water services in countries where PSI sells chlorine[1]	WHO/UNICEF, 2019 (1)	561	69.6	0.698	278.4	144

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

Notes: Column 2 is calculated by multiplying (1) by the mean under 5 population share across countries weighted by population without access to safe drinking water. Column 3 is calculated by multiplying (2) by the mean under 5 mortality rate across countries weighted by population without access to safe drinking water. Column 4 is calculated by multiplying the cost of providing coupons from Table 2 row 8 by (2). Column 5 is calculated by multiplying (4) by the estimated reduction in child mortality adjusted by usage rates: (1 − OR) * usage rate in meta-analysis / usage rate in coupons from Dupas et al., 2020 (45).

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

Table S8. Andrews and Kasy (2019) publication bias test

	Risk Difference	Log Odds Ratio
	(1)	(2)
τ	-0.010	-0.192
	(0.003)	(0.053)
σ^2_τ	0.008	0.000
	(0.002)	(0.000)
β_p	7.763	0.449
	(9.383)	(0.411)

Notes: Meta-study estimates using methods from Andrews and Kasy, 2019 (29), with standard errors clustered by study in parentheses. Column (1) uses risk differences and includes all 15 studies. Column (2) uses log odds ratio and includes 11 studies with no zero events in either the control or treatment. τ represents the publication bias adjusted meta-analysis estimate, σ^2_τ is the scale parameter following Andrews and Kasy 2019, and β_p is the relative publication probability for a result that is insignificant at 5% level, compared to a result that is significant at 5% level.

Table S9. Balance of characteristics across included and excluded studies

	Diarrhea effect size	Compliance rate	Setting	Water source type
Test for difference in means - p-val	0.32	0.23	0.51	0.13

Notes: Each column reports the p-value associated with a t-test of mean difference between included and excluded studies for each of the characteristics. Setting is a binary variable reflecting whether the setting was rural. Water source type is a binary variable reflecting whether the primary source of water (in both treatment and control groups)was unimproved.

Table S10. PRISMA Checklist[14]

TITLE			
Title	1	Identify the report as a systematic review.	Y
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Supplementary material: Prisma Abstract checklist
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Y (p.1, p.3)
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Y (p.1, p.3)
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Y (p.4)
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Y (p.3-p.4)
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Y (p.4, p.10-p.13)
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Y (p.4)
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Y (p.4)
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Y (p4-p.5, 8, 12)
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Y (p. 6 - 8, p.36-p.40)
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Y (p.4, p.7)
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Y (p.5)

[14] We closely follow the PRISMA checklist in this meta-analysis, however the flowchart of study selection provided in this paper differs from the PRISMA 2020 style. In contrast to the PRISMA flowchart we do not report the number of duplicate records, reports marked as ineligible by automation tools, reports sought for retrieval, and detailed information on the reports excluded at every stage of the screening process.

Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Y (p.4)
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Y (Added in Github repository)
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Y (p.6)
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Y (Supplementary material: ?p.5, p.6)
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Y (p.8, p.9)
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Y (p.8-p.9)
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Y (p.4, p.7)
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Y (p.3)
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Y (p.6, p.32)
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Y (p.6, p.7)
Study characteristics	17	Cite each included study and present its characteristics.	Y (p.7, p.8, p.36-p.40)
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Y (Supplementary material: Risk of bias table)
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Y (p.36-p.40, p.34, p.35)
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Y (p.7, p.8)
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Y (p.8, p.9)
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Y (p.8)
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Y (p.8, p.9)
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Y (p.7)
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Y (p.8, p.9)

DISCUSSION				
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Y (p.11)
	23b	Discuss any limitations of the evidence included in the review.	Y (p.13-p.14)
	23c	Discuss any limitations of the review processes used.	Y (p.13-p.14)
	23d	Discuss implications of the results for practice, policy, and future research.	Y (p.11-p.12)
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Y (p.1)
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Y (p.3)
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Y (p.2, p.6)
Competing interests	26	Declare any competing interests of review authors.	Y (p.7, p.15)
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Y (p.15)

Table S11. PRISMA Abstract Checklist

TITLE				
Title		1	Identify the report as a systematic review.	Yes
BACKGROUND				
Objectives		2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
METHODS				
Eligibility criteria		3	Specify the inclusion and exclusion criteria for the review.	Yes
Information sources		4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Yes
Risk of bias		5	Specify the methods used to assess risk of bias in the included studies.	Yes
Synthesis of results		6	Specify the methods used to present and synthesise results.	Yes
RESULTS				
Included studies		7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes
Synthesis of results		8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes
DISCUSSION				
Limitations of evidence		9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Yes
Interpretation		10	Provide a general interpretation of the results and important implications.	Yes
OTHER				
Funding		11	Specify the primary source of funding for the review.	Yes
Registration		12	Provide the register name and registration number.	Yes

Table S12. 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