Proposed Water Treatment Study

Table of contents

1. Study objectives and contributions

While the research project described in this proposal has been designed primarily to provide information on the impact of water treatment on mortality, it will also produce:

1. Shed light on how mortality impact varies with factors such as baseline child mortality and diarrhea rates, and with water treatment rates, thus allowing funders to better target water treatment interventions where they will have the most impact.

2. Evidence on potential pathways of impact, obtained by (a) conducting verbal autopsies to determine the cause of deaths, (b) gathering data on age-specific mortality, and (c) collecting information on morbidity data.

    Such evidence may help resolve the discrepancy between experimental evidence from meta-analyses of water treatment RCTs on mortality and models focused on particular scientific pathways. For example, the proposed study would provide information on whether water treatment reduces mortality only following weaning of children from breastfeeding, consistent with diarrhea as the key pathway, or whether there are also

effects on neonatal mortality. It will also provide information on the benefits and costs of making coupon redemption possible in shops as well as clinics. It may of course also demonstrate that certain approaches are not effective, which could also accelerate scale up by allowing resources to be focused on more promising approaches. For example, if we found that take up rates of water treatment through a coupon program in Nigeria were very low, this would point to focusing on other strategies, such as in-line chlorination (ILC), in Nigeria.

We are and will continue to collaborate closely with partners, including Evidence Action and the governments of India and Kenya, both at operational and senior levels, so that this study will allow us to position programs for rapid scale-up, contingent on positive results.

3. Detailed tracking data to enable long-run follow-ups, enabling future measurement of long-run health and mortality effects, with potential implications for both cost effectiveness and program design.

Below, we outline a phased approach. Phase I includes continued scoping and piloting, and the launch of modest scale initial studies. These would provide information on the feasibility and relative costs and benefits for each country and for both water treatment coupons and ILC prior to larger scale research in Phase II.

It would also guide decisions on whether to focus on estimating the effect of water treatment in one particular setting or whether a multi-site trial would better contribute to understanding how treatment effects vary across settings and across delivery technologies, thus allowing GiveWell s to better target expenditures.

This evidence would not only inform investments by GiveWell, but also those of developing country governments and other donors, with huge potential benefits. For example, it is possible that the collaboration in India could spur the scaling of water treatment at a state or national level as part of the JJM program in India. If water treatment has the impact estimated in Kremer et al. 2022 meta-analysis, allocation of health funds to water treatment could save more than 100,000 lives annually in India alone.

2. Background: strategies for delivering chlorine at scale

As discussed below, distribution of coupons for water treatment through health clinics and inline chlorination (ILC) are each likely to be suited to different contexts and time scales, but each offers the potential to quickly reach many people.

2.1 Inline Chlorination (ILC)

ILC devices can be used to cost effectively achieve high treatment rates in populations that use either piped water systems or communal storage tanks. They automatically dose water with chlorine without requiring users to change the way they collect and manage their drinking water, shifting the burden of water treatment from households to a service provider. ILC devices that do not require electricity for operation are commercially available. We have estimated that the current global target market for ILC is >2 billion people (those using microbially contaminated water sources that are compatible with ILC devices). Because ILC offers the potential to achieve very high adoption rates, it greatly boosts statistical power for measuring effects on mortality. A randomized controlled trial in Dhaka, Bangladesh found a 23% reduction in under-five child diarrhea prevalence and reduced expenditures on health care. Studies in Kenya and Bangladesh have demonstrated effective demand for ILC devices by kiosk owners and landlords. The government of India has expressed interest in incorporating ILC into their ambitious program to deliver piped water to all rural households in the country; we are in talks with the relevant government officials about collaborating on this timely opportunity in India. In Nigeria we are exploring a local partner that is procuring and installing devices in three cities (Lagos, Abuja, Port Harcourt).

2.2 Coupons for water treatment

A coupon program can quickly and cost-effectively deliver chlorine at scale in low-resource contexts in which few people have access to piped water or water from communal storage tanks. Such areas typically have higher baseline mortality.
Although chlorine dispensers have higher take-up rates, coupons can be further scaled where dispensers would not be appropriate.

Coupons could be scaled rapidly through existing government channels. Because most mothers in L&MICs countries already use clinics and governments already provide free goods and services as part of maternal and child health programs, very large-scale delivery is likely to be logistically, bureaucratically, and politically easier than with systems where new systems are needed for last mile delivery. The approach could be scaled by multiple governments with technical assistance from an NGO, analogous to the way that Evidence Action provides technical assistance to governments in scaling up deworming, using public schools for last mile delivery.

Coupons delivered through health facilities can also target water treatment to pregnant women and parents of young children, potentially increasing the cost-effectiveness of the program. Coupons could also allow water treatment to reach people without access to tanks or standpipes, and these are often those who need water treatment the most. Water treatment is only provided to those who redeem coupons, further increasing cost effectiveness. Evidence from Kenya and Malawi suggests that the coupon program effectively targets water treatment to those who will

use it, without screening out people who might treat their water if it were delivered to them directly. Evidence from Malawi suggests that people are more likely to take up water treatment to those at higher risk, for example those with lower quality water-sources.

3. Activities: Phased approach

In phase zero (during 2022), we have built up a team, and begun scoping work in Kenya, India, and Nigeria.

In phase one (beginning early 2023), we will continue to scope and pilot in Kenya, India, and Nigeria. Scoping and piloting in several locations is clearly necessary for a multi-site study, but also desirable for a single site study, because it may reveal that a particular country and delivery technology is a much better candidate than the others.

Phase one will include launching an initial coupon study in Kenya and an initial study in either India or Nigeria[1]. Data on mortality, chlorination rates, cause of death, morbidity, and medical expenses from these initial studies will contribute to a meta-analysis. These will also enable refinement of implementation and data collection approaches, to inform a larger scale work.

By mid 2023, we will have information about key parameters in all three sites, enabling a decision to be made about how to best distribute remaining funds between locations and interventions to best deliver on program goals, including better understanding impacts on mortality, how mortality effects vary across contexts, and cost effectiveness at scale, as well as laying the groundwork for program scale up. Phase two (beginning fall 2023) will include launching larger scale studies in India, Nigeria, and/or Kenya.

Phase 0: Update on 2022 activities so far (DIL-funded).

Kenya: Coupon Study Preparations
We have built up a team in Kenya to support this project. This includes a Principal Investigator based at the Kenya Medical Research Institute, Dr. Sammy Khagayi and three full time staff-members based in Kenya, one of whom is working on the project full time, the others splitting their time between this and other projects. We also have US based team-members supporting.
This team has completed many key preliminary steps for a pilot study at two Health and Demographic Surveillance Systems (HDSS) sites (comprising 40 clinics) in Kisumu and Siaya

---

[1] We're budgeting for an initial study in India or Nigeria, because we expect that only one will be promising enough to launch a full RCT. In the event that it is possible to launch initial studies in all three countries, we could do so with smaller sample sizes or shorter follow-up periods.

counties in Kenya. We met with Dr. Amoth, Director General for the Ministry of Health, who is enthusiastic about a study and a potential scale-up of a coupon program. We have obtained IRB approvals from the University of Chicago and KEMRI for a coupon study, and are working on amendments to these for updates to our research design. Team members in Kenya have presented the coupon program to the Siaya County Director of Public Health and received approval to conduct the program (Kisumu approval is expected by December). A system to procure WaterGuard and AquaTabs has been worked out with PSI.

Health facilities visits and consultations have been conducted to refine the enrollment and redemption system to minimize burden on health workers, and local partners consulted to refine program details including the design of coupon booklets and instructions for users of water treatment.

Members of our research team have also been actively working on ILC in Western Kenya and have conducted ILC pilot studies at community water kiosks and health care facilities in collaboration with the NGOs SWAP and CARE, measured effective demand among kiosk operators, and are currently gathering evidence to better understand cost and optimize operational models. Funding from Open Philanthropy is currently supporting establishing local manufacturing capacity in Nairobi for a novel liquid ILC doser that can be refilled with WaterGuard or paired with electrochlorinators (decentralized chlorine production).

India: October 2022 - December 2022. Scoping activities to inform pilot and intervention design.

The team is currently collaborating with the central Department of Drinking Water and Sanitation (DDWS) which runs the Jal Jeevan Mission (JJM), a flagship program to provide a functional household tap connection to all rural households in India by 2024. We met with the Secretary of DDWS, the Mission Director of JJM, and the new CEO of NITI Aayog (the apex public policy think tank in India) in July. The CEO and the Secretary have appointed a nodal officer within the Department for us to coordinate with. They also requested a note from our team estimating the potential number of lives that could be saved by providing safe drinking water to all rural households in India. The note received wide press coverage and was shared extensively in WASH forums. Further, Michael Kremer has been invited to participate in the LiFE Mission which aims to incentivize individual behavior changes to combat climate change.

In collaboration with DDSW and Evidence Action we have developed a plan for water testing and investigation of ILC systems, systems for chlorine refills, and scoping of states with the combination of high water contamination and baseline mortality and appropriateness of water infrastructure for ILC that might make them well-suited to both prioritization of ILC as a policy measure and for a mortality study.

Nigeria. November 2022-January 2023. Scoping activities to inform pilot and intervention design.

We are partnering with the Nigeria Institute of Medical Research (NIMR), and its Director General/CEO, Professor Babatunde Lawal Salako. Deputy Director for Research Dr. Babatunde Adewale is our contact person and leads a team of researchers in the institution including, a cellular microbiologist, a biostatistician, and a social scientist.

We secured University of Chicago and local IRB approvals, as well as state and local level stakeholders' permission for data collection. The team is working closely with NIMR to train local partner staff and enumerators to carry out the qualitative data collection effectively.

We have conducted focus group discussions (FGDs) and interviews with pregnant women and women with small children, health facility workers, shop keepers, and water point owners in Kano and Lagos states in Nigeria, and we will do the same in Ogun state in January. One goal is to understand chlorine acceptance, since this is key to take-up in a coupon program (ILC programs can still operate where chlorine acceptance is low, since ILC devices can be calibrated to give doses that are effective without leaving a taste). We plan to hand out chlorine to a small sample of women and follow-up with them in a few weeks to test their usage.

Another goal would be to assess water-treatment practices and existing water infrastructure to assess the feasibility of ILC. Team members already identified Society for Family Health, as a key NGO player in chlorination in Nigeria, who deliver water treatment in the country and could both procure chlorine and ILC devices. They are already beginning to use ILC devices for water storage tanks at a small scale. We believe that they are a highly promising partner, who we could work with to deliver ILC in both rural and urban areas. They could also support the delivery of a coupon program.

The team is also conducting field visits to understand the compatibility of ILC devices with existing water sources and assess whether sufficiently large populations in urban and rural settings could be served by ILC. We plan to partner with a survey firm or NGO to further map-out infrastructure, assess the feasibility, select and pilot ILC devices, in the context of Nigeria.

Phase I: Initial studies (Starting early 2023)

We are planning initial studies in up to three sites – Kenya, Nigeria, and India. In Kenya, we have a great opportunity to work with HDSS. Our team has been collaborating with them and will begin a study in Kenya in January 2023. We are conducting preliminary scoping to explore

the potential of initial studies in Nigeria and India. We would aim to launch an initial study in either India or Nigeria depending on which is more promising. This would be a full RCT with participants tracked for two years, and mortality data pooled with data from studies in other locations.

Kenya

February 2023. Launch initial study in 40 clinics (HDSS sites). This will be a full RCT, in which around 6,000 pregnant women - and 19,000 mothers of children under 5- will be enrolled over one year with implementation and data collection over the following two years. Mortality data from this study will be pooled with data from studies in other locations.

The HDSS is already regularly surveying these households so we can collect data on water quality, morbidity, mortality, and healthcare utilization at low cost. With our financial support, they will embed testing of water for chlorine in their survey starting in January 2023, allowing us to measure uptake.

We will also use this study to inform further work. Cross randomizing variants of the program, including redemption through clinics or kiosks, providing WaterGuard or Aquatabs, and using SMS information messages and reminders will allow optimization of design and to further refine budget and MDE estimates for the mortality study. In particular, we would be able to gauge: (1) the coupon redemption rate and chlorine usage rate, and whether they can be increased through additional information campaigns, product options, SMS reminders/nudges, allowing redemption at shops, etc.; (2) the characteristics of those reached by the program, in particular, how selective the sampling procedure of women coming for ANC visits is, and whether coupon redemption rate varies based on underlying need for water treatment and baseline child mortality.

We would also explore the possibility of conducting a study on ILC in Kenya. This would involve identifying suitable sites, and exploring whether people with access to water sources amenable to ILC use the same water sources consistently enough to allow randomization and tracking.

Nigeria

February - August 2023. Initial study preparation and launch. Based on the Based on the findings from Phase 0, we will prepare to launch an initial coupon and/or ILC study. This will include selecting the most promising villages or neighborhoods (rural and/or urban, in a particular state), based on child mortality and diarrhea rates, chlorine acceptance, ANC attendance and facility capacity (for coupons), presence of compatible water points (for ILC), presence of capable implementation partners and data collection firms.

We will then need to obtain permissions from officials and IRB approvals both at the federal and at the state level, contract with data collection partners in the relevant state(s), set up procurement and distribution systems for chlorine products and other materials, and get buy-in from health facilities, kiosks, and/or managers of water points.

If Nigeria is selected as an additional site for an initial study, we could launch an initial study of either coupons, ILC or both. As with Kenya, we would initially launch with a fairly modest sample, and monitor take-up carefully, possibly with slight variations in design. This would both inform decisions about whether to launch larger-scale work in Nigeria, and position us for a larger scale study. National elections will be held in February 2023, which could delay activities. Assuming all goes well with elections and partners in Nigeria, we would be ready for an initial launch in Summer 2023, and for larger scale work by the end of 2023.

India

Jan - May 2023. Scoping and preparation for initial study. Evidence Action is working on a technology assessment of ILC device(s) that could be used at scale within the JJM program. Once they have identified devices by early January, we will collaborate with the DDWS to conduct an implementation pilot of the device(s).

In a number of villages across a few states with high child mortality, we will conduct scoping activities to understand the water supply infrastructure, the different sources of water and potential compatibility with ILC devices, and the extent of water (re)contamination. We will map overhead water tanks, borewells, pipelines, etc. and collect information on whether and where the water is treated before being supplied to a village. Finally, we will test water samples at different points along the pipeline for chlorine, *E. coli*, and total dissolved solids to get a sense of the extent and pattern of (re)contamination of water. This step will allow us to figure out how to place ILC devices within the water system to ensure that water is free from microbiological contamination, without having so high a concentration of chlorine that there is a strong taste. We have been liaising with Department of Drinking Water and Sanitationofficials at each step and have solicited their support in connecting with state-level officials who are in charge of the JJM scheme.

July 2023 - Nov 2023. Initial study. If India were selected as a site for an initial study, we would work with a few state governments to install ILC devices, with cluster randomization. Given the uncertainty at this point about the types of water sources used in different parts and their compatibility with various types of ILC devices, we are planning to focus on installing and operating just one type of ILC device – one which can be retrofitted onto existing JJM pipes. We may update this plan after Evidence Action's technical review and infrastructure assessment if we feel it is an infeasible solution.

Over the first three to four months of the study, we will test different ways to identify and enroll pregnant women in a mortality study. One way would be to do a door-to-door census; this would be thorough but resource intensive. The second way would be to leverage the information collected by the village-level daycare workers (anganwadi sevikas) and/or the local health worker (ASHA worker) which would be more efficient but would risk excluding some pregnancies.

We have accounted for one month's time to receive government approval. Through the pilot, we will test out ways to operate and maintain the devices, refill chlorine, etc. We will conduct taste tests with households to gauge their level of chlorine acceptance.

Phase II: Allocation of remaining budget across sites and distribution methods

From mid to late 2023, we will have better information about a) whether it is plausible to run a large-scale study in each country, b) factors affecting the power of such a study (expected take-up, cost of implementation and data collection, intra-cluster correlation for ILC), and c) how quickly such a study could be launched (e.g, established partnerships, required permissions and processes, etc), and d) the likelihood of scale in each context (for example, if the government of Nigeria were enthusiastic about scaling a coupon program, that might raise the value of working in Nigeria).

Based on this, we would discuss the options with GiveWell. There may be trade-offs here between different objectives (estimating the local treatment effect, estimating the global treatment effect, positioning the program for scale-up). We would then make a collective judgment on how to spend the remaining $8.2m. This would include whether to run a single- or multi-country study, choosing countries and study sites, and deciding on the design of implementation and data collection.

We would then launch larger scale studies as early as possible. We understand that GiveWell puts substantial value on having answers as soon as possible, as well as on conducting the best possible study.

We're fairly confident that we could launch a larger scale study in Kenya by mid 2023. Three to four months after the launch in HDSS sites, we should have sufficient information to settle on a final version of the program (clinic vs kiosk redemption, water-guard vs aqua-tabs, etc.), and launch at a larger scale.

We can provide estimates for other countries, but they're subject to some uncertainty. In both India and Nigeria, we should be able to conduct the research activities described above and be ready to launch larger studies between October 2023 and January 2024. However, timelines will depend on local partners, including state and national governments and implementing partners. We are continuing to liaise with government partners to try to address this, for example, Elisa Maffioli just traveled to Nigeria and Michael Kremer is going to India in January. They both met (or plan to meet with) the relevant ministers and officials. By Fall 2023, we will have a much better understanding of timelines in each country, and this can be taken into account when choosing where to conduct larger-scale research.

4. Precision of estimates of mortality effects

Consider multiple studies of water interventions, indexed by i. If the true effects of in each study is denoted by θ_i, an RCT estimates the unknown effect, θ_i, with some standard error:

$\hat{\theta}_i \sim N(\theta_i, se_i^2)$. In our case by effect we mean magnitude of reduction in (log) odds of child mortality in treatment vs control.

This section presents several different estimators of the effect of water treatment on child mortality.

1. Estimates of the average global effect of water treatment based on a meta-analysis of all available experimental studies. We will focus on an estimate based on a random-effects meta-analysis and the corresponding Bayesian meta-analysis.

    In the context of the notation above, this corresponds to assuming that effects in each context are drawn from a common normal distribution, $\theta_i \sim N(\mu, \sigma^2)$, as is done in random-effects meta-analysis (e.g. Kremer *et al.* 2022), we can estimate the average global effect, μ, and its heterogeneity parameter, σ.

2. An estimate of the effect of water treatment in a particular context, based on data only from that context. (From a frequentist standpoint this is a single study estimate without any meta-analysis; from a Bayesian standpoint it can be considered a no-pooling estimate.)

    This means simply using $\hat{\theta}_i$ as the decision-makers context-specific estimate of true treatment effect in a given setting.

3. An estimate of the effect of water treatment in a particular context, based on a frequentist random effects model (and the corresponding Bayesian partial pooling model). Under the

random effects model above, this is an optimally weighted average of the global average treatment effect (calculated in 1) and the treatment effect in the particular study (calculated in 2).[2]

In terms of notation we introduced above, the estimate in each context, θ^*, is a weighted averages of $\hat{\theta}_i$ and μ, that is:

$$\theta^*_i = \omega_i\mu + (1 - \omega_i)\,\hat{\theta}_i, \text{ where } \omega_i = 1 - \sigma^2/(\sigma^2 + se_i^2).$$

This allows for the estimate of treatment effect in a particular context to be informed by data from other contexts.

We call ω a pooling factor; setting it to 1 coincides with fixed-effects meta-analysis (homogeneous study effect), setting it to 0 assumes each study of water is completely unrelated and yields no information about effectiveness of intervention in other settings. In practice $0 < \omega < 1$ and we can estimate ω from data to provide optimal weights, which balance context-specific effects with borrowing of information from other contexts. (We will also report a fixed effects meta-analysis and the corresponding Bayesian full pooling meta-analysis for comparability with GiveWell's current approach, and a test of the hypothesis of no heterogeneity across contexts).

As discussed below, this third estimator is likely to be a better predictor of effects from scaling up treatment in a particular context that has been studied than either estimate 1 or 2.

We propose to produce each of these estimates, and with a total budget of $10m ($8.2m for Phase II) we expect fairly good precision.

Below we first discuss some methodological considerations. We then share estimates of the minimum detectable effects (MDE) on mortality, expressed in percentage reductions from baseline, for a variety of single-country and multi-country study scenarios.

4.1 Methodology: MDEs and other concepts

This section focuses on MDEs. This is in part because we had started by discussing an individual RCT with GiveWell and it is still possible that the full budget will wind up being allocated to an individual RCT, and an MDE is a key metric for a single RCT; and in part because GiveWell

---

[2] Again, this could incorporate information on how effects vary with variables such as baseline mortality rates and rates of water treatment.

requested information on MDEs. These MDE estimates are useful in that they demonstrate that we can achieve reasonable power within the budget.

However, in the meta-analysis context, the MDE concept should be supplemented with other measures to best inform the ultimate decisions of how to allocate the remaining budget based on the data revealed in piloting. We think that the concepts of *precision* of estimates may also be useful as well as the chance (and ideally expected cost) of a false negative and of a false positive. Ideally these would form part of a value of information analysis, but they are also informative on their own. After piloting we will share not only updated estimates of MDEs, but also a fuller analysis with a broader range of concepts, which can factor into the decision about how to allocate the budget within the set of contexts which are feasible.

There are two reasons why we think a broader range of concepts will be useful. First, it is not completely clear how to define the MDE in a multiple study context (see 4.1.1). Second, a lower MDE is not a reliable indicator of a more valuable study. For a single RCT, increasing the sample size of an RCT lowers the MDE, increases precision, reduces the chance of a false negative, and produces more valuable information. However, in a meta-analysis context, this is not necessarily the case. Some decisions (such as splitting the budget across multiple studies) could produce a host of benefits as indicated by the value other concepts (see 4.1.2), but actually make some MDE measures appear worse .

4.1.1 Defining MDEs

MDEs are not commonly used for meta-analysis, so to calculate one such that it is analogous to the MDEs for an individual-study, we define the MDE in a meta-analysis context as follows: "if each new study found an effect of X, what is the smallest X such that the meta-analysis would have an 80% chance of finding a statistically significant on mortality?". So when calculating an MDE for a meta-analysis, we assume that each new study has an identical effect. One downside of this approach is that this does not fit in well with the assumed random effects model, in which the true effect is different in every context. However, alternative approaches would require making assumptions on the extent of heterogeneity and would be less comparable across scenarios with different numbers of studies.

4.1.2 Value of other concepts

To illustrate why these additional concepts are useful in a meta-analysis context, consider the choice between concentrating the Phase II budget on a large single-site study, or splitting it between two or more studies in different contexts.

The multi-site study would yield much more precise estimates of the average global effect. It would also increase precision in understanding how the impact of water treatment depends on

variables such as the fraction of the population who treat their water or the overall mortality rate. Finally, it could reduce the chance of a false negative or false positive result.

However, focusing only on MDE would miss these effects. In fact, splitting the study across contexts worsens some of the MDEs we present below. This is not only because of the fixed costs of working in more settings, but due to limitations of the MDE concept which focuses on identifying the lowest effect such that a single summary statistic of program impact would be statistically significant.

4.1.3 Adjusting MDEs for comparability across delivery mechanisms.

We calculate higher MDEs for ILC than for coupon studies. This is because ILC requires cluster randomization. Assuming positive intra-cluster correlation, this increases the estimated MDE within a budget constraint.

However, the benefit of ILC is that it is likely to have near 100% take-up. This does not reduce the MDE directly, but it means that we would expect a larger mortality effect, reducing the overall probability of a false negative result. We therefore rescale the MDEs for ILC studies to be comparable to the MDE for coupon studies, by multiplying it by the ratio of the expected take-up between ILC and coupons.

4.1.4 Age of children

Age distributions vary between studies of water treatment. This study would focus on children from pregnancy until the age of two, although we may also obtain some data on older children (e.g. older siblings in a coupon study or children who were already alive at the time ILC devices were installed in ILC study.). As discussed above, our proposed meta-analysis framework deals with this by allowing for heterogeneity in effects between studies. It also allows us to examine if mortality effects vary with age, since this is consequential for program design. We are therefore working on a specific model to estimate effects by age group. We have obtained micro-data on age of death from 11 out of 15 studies in the meta-analysis. However, at the moment we have limited power to estimate how effects vary by age. New studies would help address this.

4.2 Process to refine estimates

The MDEs, precision, and chance of a false negative that can be achieved for a given budget depend on variables such as the cost per data point, the intra cluster correlation for inline chlorination, mortality, etc. This proposal gives MDE estimates based on our current expectations on these parameters, but during piloting, we will gather more data on these parameters, and explore ways to maximize power within the budget constraint, such that by summer 2023 we would have updated estimates. We think there is a good chance that we would be able to find substantial gains during piloting, for example by finding ways of increasing

take-up, randomizing within a group with higher take-up, focusing on higher mortality areas, or finding ways to reduce data-collection costs. Another reason to look for improvements is to reduce the risk that negative shocks prevent us from securing the MDEs presented below.[3]

We have created a model to estimate how much statistical power we would have depending on the number and location of studies, the implementation method, mortality rates, cost of data collection, etc. As discussed above, piloting will provide information to update the parameters of the model and thus will feed into decisions on how to allocate Phase II expenditure across potential studies.

4.3 Estimated MDEs

4.3.1 Estimating the global average treatment effect

The new study will have a much larger sample than the 15 studies in Kremer et al. combined (depending on how the study is designed it could have between 120,000 and 400,000 participants, compared to around 25,000 in the existing meta-analysis), and thus they would have the potential to substantially improve the precision with which the mean effect of water interventions can be estimated and to shed much more light on how effects vary with factors such as the death rate and the chlorination rate. The ability to do this would likely be maximized by allocating the budget across three or more different studies, although this would reduce sample size and thus the precision of the zero pooling estimate for each study.

GiveWell had asked about a scenario with studies in Kenya and Nigeria. We estimate that combining a multi-site study in Kenya and Nigeria with the 15 studies in the Kremer et al. (2022) meta-analysis would yield an MDE of the global average mortality reduction between 0% (in which the overall global average estimated treatment effect through the meta-analysis would be significant with any non-negative effect in the new studies) and 2.5%.[4]

It is likely that we will be able to improve on this by optimizing the split of the budget across settings in order to maximize statistical power.

4.3.2 Estimates of context-specific treatment effects using only data from that context

---

[3] Each of these estimates relies on standard statistical design assumptions of 80% power at 5% significance level. Each estimate assumes a three year study (one year of enrollment followed by two years observing each participant). Each estimate assumes a budget of $8.2m.

[4] We could conduct this calculation with some subset of the studies in Kremer et al. if that would be helpful.

Concentrating the Phase II budget in a single country would minimize the MDE for the context-specific effect in that country, albeit at the cost of lower precision to estimate the global effect or to characterize heterogeneity. We estimate that spending the entire Phase II budget on a single coupon study in Kenya would lead to an MDE of 5.9%.[5]

If we were to conduct a two-country study, we could also estimate an average treatment effect in the two contexts by pooling the data from these two contexts. For example, if we conducted coupon studies in Nigeria and Kenya (as in the above example), we would have an MDE of 6.2%. This is higher than the MDE for either single-country study because there is a fixed cost of working in each new country, so the sample for a two-country study is slightly smaller. However, a multi-site study would yield other benefits, such as providing more quantitative evidence to predict context-specific effects.

We currently estimate that a single-country study of ILC in Nigeria would have an adjusted MDE of 3.6% (raw MDE 6.1%) and a single-country study of ILC in India would have an adjusted MDE of 2.4% (raw MDE 9%).[6]

4.3.3 Estimates of effects in a specific context using random effects/Bayesian partial pooling

We will calculate frequentist random effects and analogous Bayesian partially pooled estimates for every context where studies are available. These are estimates of the effect in a particular context, which are a weighted average of the estimated global average treatment effect from a meta-analysis (as in 4.1.1), and the estimate obtained from data in that study (as in 4.1.2). Weights are chosen optimally under the model, with higher weighting of the study-specific estimate if it has a larger sample size. This will provide more precise estimates of context-specific treatment effects and will likely lead to better predictions of effects in future implementations in that context than using data from the study alone or focusing only on the global average treatment effect. As such, we prefer this measure to those outlined in sections 4.1.1 and 4.1.2.

We estimate that if we concentrated the Phase II budget in a single study in Kenya, a Bayesian partial pooling estimate would have an MDE of 5%. If we split the study between two contexts, the MDE for each context would be somewhat greater, as we have a smaller sample in each study. For example, if we split the budget between a coupon study in Kenya and a coupon study in Nigeria, we estimate that the partially pooled estimates would have MDEs of 7.8% and 6.1%

---

[5] Note that this MDE is not directly comparable to the one in the previous section, because it is estimating something different (the context-specific treatment effect, rather than the underlying average global treatment effect).

[6] There is more uncertainty in our estimate for ILC in Nigeria than in India. Power in a cluster-randomized evaluation is sensitive to the cluster size and the intra-cluster correlation. We are using an estimate of ICC that comes from Kenya, and we have estimates of cluster size from Uttar Pradesh. For Nigeria, we are assuming the same cluster size as India for now, but this could vary substantially.

respectively. As discussed, MDEs for ILC are substantially lower than for coupon studies, if we adjust for the higher take-up.

While decisions on whether to undertake a single or multi-site study can only be made based on the results of Phase I piloting, currently we see a multi-site study as the most promising option. This would enable us to estimate the global average treatment effect more precisely and could allow GiveWell and others to incorporate more empirical data into efforts to adjust estimates of treatment effects for context. It would also produce strong statistical power for estimating a context specific treatment effect, with an MDE only slightly higher than a single country study.

<u>4.3.4 Predictions of treatment effect in a given context, using a meta-regression</u>

Under the relatively simple random effects model described above, effects in different contexts vary, but there is no attempt to model the causes of these differences. This means that the difference between the global average effect and the local context specific effect is taken as statistically independent and hence estimates of context-specific effects are a weighted average of the estimated global average treatment effect from a meta-analysis and the estimate from that study.

However, we may be able to consider models in which effects from similar interventions in similar environments are more likely to have similar effects and would thus be more heavily weighted. One way to do this is in a meta-regression which would estimate how various characteristics (e.g. mortality and effective take-up rate, etc.) affect treatment effects. When estimating the effect in each study, it might then be possible to give other studies with similar characteristics more weight. This would serve the same purpose as (some of) the 'internal validity adjustments' and 'external validity adjustments' in GiveWell's cost-effectiveness analysis.

For example, we estimate that a small initial HDSS study in Kenya with 6,000 pregnant women would have an MDE in a random effects model of 20.5%. (Of course this study is designed not primarily to shed light on mortality impacts but to inform the design of larger studies and to yield low cost information on morbidity, health care utilization and expenditure, etc.). A meta analysis which more heavily weighted studies in the same geography and with similar mortality rates would likely yield a lower MDE.

How to treat study characteristics is ultimately a subjective decision. As such, we would pre-register our planned design. We would be happy to discuss that with GiveWell. GiveWell would also be able to input its own parameters into this model if you have different views on how to take into account study characteristics.

Table 1: MDEs for selected scenarios

Scenario	Adjusted MDEs		
	No pooling (estimate for only new study)	**Meta-analysis estimate of global average treatment effect**	**Partial pooling (weighted average of meta-analysis estimate and individual study effect)**
Kenya coupons only	5.9%	0%	5%
Kenya coupons + Nigeria coupons	6.1%	2.5%	KEN: 7.8% NIG: 6.1%
Kenya coupons + Nigeria ILC	4.4%	1.4%	KEN: 7.8% NIG: 3.2%
Kenya coupons + India ILC	4.5%	1.5%	KEN: 7.8% IND: 3.4%

Row 1 (Kenya coupons only), has around 300,000 participants. Each subsequent scenario has 100,000 participants in Kenya, with the number of participants in Nigeria and India calculated such that the total budget is $8.2m. For all of the calculations we assumed an attrition rate of 5%.