REPORT ON INDEPENDENT REVIEW

Kremer M, Luby S, Maertens R, Tan B. Water treatment and child survival: a systematic review and meta-analysis

Thomas Clasen, JD, PhD

Commission

This report was commissioned by GiveWell in connection with its assessment of the cost-effectiveness of a water treatment intervention consisting of either chlorine dispensers at water collection points or in-line water dosing systems on water storage tanks. The commission requested that I review and comment on the above-referenced manuscript which GiveWell provided to me. As perhaps the primary source of data on the mortality outcome, the conclusions from the paper may weigh heavily in estimating the cost-effectiveness of the intervention. As a result, my aim was to identify potential issues that may impact the strength of its conclusions.

Disclosure of Interests

I accepted this commission after discussing with GiveWell representatives the possible conflicts of interest presented. I was a co-investigator of the WASH Benefits trials in Kenya and Bangladesh that were led by Drs. Kremer and Luby (and Dr. J. Collford from UC Berkeley). I currently am a lead investigator with Dr. Luby on RISE trials led by Monash University, and I've been on several other papers with him. This includes systematic reviews of WASH interventions led by the WHO and used to estimate the effects of interventions on diarrheal disease morbidity as part of its estimates of the global burden of disease associated with poor WASH conditions. I have led many trials of WASH interventions, including some of the studies included in the Kremer review (Boisson 2013, Peletz 2012, Kirby 2019). These and several other of my studies have been funded by private sector companies, including Aquatabs--makers of the NaDCC tablets used in several of the studies included in this review--as well Vestergaard who manufacture and sell point of use water filters used in the Boisson 2010, Petletz 2012 and Kirby 2019, and Unilever. I have also led my own Cochrane reviews of interventions to improve water quality to prevent diarrhea (2006, 2015) from which Kremer et al. seem to have identified most of their included papers. I am the Rose Salamone Gangarosa Professor of Sanitation and Safe Water at the Rollins School of Public Health, Emory University, but am providing this report as an independent consultant and not as an employee of Emory. While I consider myself agnostic about household water treatment and storage (HWTS), I believe others might regard me as an advocate.

Threshold Issues

1. I note that the review has not yet been published. I understand that it was submitted to one journal where it was rejected, and has now be resubmitted elsewhere. In any case, I have not seen any comments by editors or peer reviewers. I strongly suggest that before relying on this paper, GiveWell wait for it to be published, so that it has the benefit of any comments or suggestions from peer reviewers and editors.

2. While systematic reviews are not more complicated methodically than the work for which these authors are well known, rigorous systematic reviews are protocol-driven and follow a carefully prescribed methodology designed to minimize potential bias (e.g., Cochrane Collaboration Handbook). This review does not always do so.

3. There is always a risk of bias in systematic reviews, even disciplined ones, that include papers by one or more of the reviewers. Here, a majority of the 13 papers included in the review were led by Luby (6) or Kremer (2). I cannot tell how much weight these papers contributed to the overall results, or whether their omission would have influenced the pooled estimate of effect; it almost certainly would widen the confidence interval. This heavy reliance is not fatal; there are not many studies overall, and Luby is a leading researcher in this area. Kremer has been an advocate of chlorine dispensers at water collection points, the intervention that is promoted here even though none of the interventions were exclusively dispensers (Null 2018 and the follow-up Haushofter 2020 study used a combination of dispensers and bottled chlorine). While none of this establishes bias, it might be helpful to at least acknowledge the authors' experience in this area.

4. I understand that this review is being used by GiveWell to inform a CEA on chlorine dispensers or in-line dosing systems in order to advise potential donors regarding these interventions. While I understand that the cost-effectiveness of health interventions must often be based on limited and indirect evidence on health effects, I am not clear how the necessary inferences can be drawn from this review. This is not only due to the methodological shortcomings noted below; it is due to the fact that the interventions included in this review--though problematically heterogeneous--are almost all different from the dispensers and dosing systems. I am advised by GiveWell that the CEA will somehow adjust for some of these differences, though I do not know how such adjustments can yield reliable estimates of effect on the interventions GiveWell has under consideration. In my judgement, the effect estimates that can reasonably be drawn from this review must be limited to the sub-groups of interventions that are actually included in the review. Inferences about other interventions, such as dispensers or chlorine dosing systems, would be highly speculative.

Methodological Issues

1. At the outset, the review does not seem to follow methods that help minimize bias and help ensure rigor. There is no indication that it was executed in accordance with a pre-specified protocol--a basic and important step in systematic reviews. Protocols are not always peer-reviewed or published (as they are, for example, by Cochrane) but they should at least be fixed in advance and time-stamped to hold the reviewers to it and to explain any deviation. The authors registered a protocol in July 2020 (AEA RCT Registry number: AEARCTR-0005977), but on inquiry with GiveWell, it appears to have been registered after initial versions of the meta-analysis were conducted. Systematic reviews should be registered in advance, like trials.

2. The scope of the review is confusing. The title--"Water quality and child survival" implies a vast range of possible interventions, including conventional municipal systems that combine various methods. In the Summary, this is narrowed to "the effect of chlorination". However, in at least three included papers (Peletz 2012, Kirby 2019, Kremer 2011) the intervention used filters or spring protection and no chlorine at all. Moreover, in at least four other papers (Reller 2003, Luby 2006, Crump 2005, and Chiller 2006), investigators tested a product that included both chlorine and a flocculant--an essential antimicrobial agent that not only enhances chlorination but is necessary to address the chlorine-resistant microbes such as *Cryptosporidium* that is a common cause of moderate to severe diarrhea (Kotloff et al.2013). Even the type of chlorine is different, ranging from NaDCC tablets to calcium hypochlorite to different dilutions of sodium hypochlorite. It would be misleading to infer an "effect of chlorination" by treating this disparate group of studies as a homogeneous intervention.

3. The confusion about scope is not made more clear in the Methods section of the paper itself where it uses the term "water interventions"--a term that would suggest improvements in water supplies (e.g.,

household connections, boreholes, etc.) that address water quantity and access, independent contributors to health beyond water treatment alone. The review should clearly define eligibility in terms of populations, interventions and outcomes, and to determine eligibility on those criteria.

4. The other issue concerning scope is the use of the terms "child survival" and "child mortality". These are not synonymous, and neither are defined in the paper to understand what was meant here. The review needs to clearly define the outcome. I believe it is all cause child mortality reported by the paper or investigators but determined in a variety not especially rigorous ways since none of the papers had death as a primary outcome. Survival connotes a survival analysis (time to event), which is not what was reported in the eligible studies or analyzed here.

4. While we all understand that powering an RCT on mortality is likely impossible, trying to pool estimates of effect on non-primary outcomes is fraught. In all of these included studies, mortality was at best a secondary outcome; in many studies it was probably not a pre-specified outcome at all but something that researchers collected incidentally or to address the adverse event reporting required by ethics authorities. This is perhaps evidenced by the fact that of the nearly 60 RCTs of water quality interventions to prevent diarrhea in our Cochrane review from five years ago (Clasen 2015)--and many more conducted since--the authors included found mortality data on only 13. They note in Table S2 that 24 of 48 otherwise eligible studies did not make the cut because no mortality data was collected. This limited data from the large pool of studies that undertook drinking water quality interventions is a significant source of possible bias.

5. Although mortality is presented here as an objective outcome, there is considerable heterogeneity in the way it is collected especially in studies, like these, where it is not a primary outcome. Curiously, while Table 1 describes extracted study data such as contamination levels and diarrhea rates that are at best indirectly relevant to the aims of the review, it does not describe how mortality was ascertained. Best practices would require reviewing death certificates or hospital records or conducting a verbal autopsy. However, it's unlikely that any of that was done in the included studies given the secondary need for these data and the expense of doing so. More likely it was reported by the field workers based on the householder reports. This is another source of measurement bias and limitation that the review should acknowledge.

6. Some of the review's methods are also inconsistent with best practices for systematic reviews. For example, the review excludes studies (Boisson 2010, du Preez 2011) for reasons other than failing to meet eligibility criteria. The reasons given are not convincing; most of the studies had issues that impact results. At the same time, the review includes one unpublished study by Haushofer (2020) despite its falling outside the stated search period for eligibility and presenting risks of sampling error, selection bias and measurements errors that may have contributed to its highly protective effect on mortality (63% reduction) and large confidence level.

7. The Results section leads with a summary of diarrhea morbidity and present a distribution of same. However, this is not "results" from this study at all but from IHME's work. Moreover, the focus on diarrhea morbidity here is curious, not only because it is not described as an aim of the review at all (though it could have easily been extracted from the included papers), and because the authors go out of their way in the Discussion section to dismiss the relevance of diarrhea morbidity in estimating mortality. What might have been helpful is to report in Table 1 not the prevalence of diarrhea (and contamination level) but the effect that the study had on diarrhea (and exposure), since these might help support (or not) the inference that the intervention actually reduced mortality by improving health through reductions in

exposure--presumably the theory of change for this review. If they had, it would be evident that the two studies that contribute most weight to the Peto-ORs were Null 2018 and Humphrey 2019, studies that reported no effect from the intervention on diarrhea.

8. Many of the studies included in the review, including some of the large trials with large weights, had multiple intervention arms such as sanitation, handwashing and nutrition or a combination of these (Null 2018, Luby 2018, Humphrey 2019). The water treatment only arms represented a relatively small population of the overall study. It is not clear from the paper that the mortality effect included from the multi-arm trials was from those chlorination only arms given the large weight assigned to the studies. This should be checked.

9. Notably, the Results section reports no data on cost or cost-effectiveness. Fair enough, this was not an aim of the review. Nevertheless, the Supplemental Material has two pages describing cost-effectiveness and advocating for it based on other research. This is carried through in the Discussion section of the paper. This might belong in a commentary or advocacy piece, or in communications with GiveWell pitching the intervention. There, it would not have to be supported by data and its purpose would be more transparent.

10. The paper does not include any analysis of study quality (e.g., LQAT) or of strength of evidence (GRADE), both of which are standard in systematic reviews. It is especially important to identify shortcomings in study methods in order to better understand the inferences that can reasonably be drawn in the meta-analysis. The authors might also consider at least acknowledging the investigators that contributed data to the review.

Discussion/Interpretation

1. The methodological shortcomings of the review substantially limit the inferences that can be drawn. Unlike other reviews, however, this paper makes little mention of either the limitations of systematic reviews and meta-analyses generally, or the additional limitations of this review. If it did, it would have emphasized the heterogeneity in interventions and the methods of the papers included (not just the I^2, which only measures heterogeneity across the measures of effect), and the appropriateness of meta-analyses under these conditions. It would have also noted the limited number of studies that contributed data out of the total pool of eligible interventions. It would also have emphasized that none of the included studies was actually designed to investigate mortality so that this outcome was not always rigorously ascertained.

2. The Discussion section is also unusual in that there is no attempt to position the results in the context of other research on child mortality and water. While the focus is obviously on RCTs, there is a large body of non-randomized controlled interventions as well as non-experimental studies that address this subject. A Discussion section would normally cite the more rigorous work (such as their references 4-7), placing it in that context but also studies that have found contrary results beyond GBD estimates.

3. Instead, the Discussion focuses on two subjects, both of which are actually beyond the scope of the results. The first is a long list of reasons why models of mortality informed by diarrhea morbidity may underestimate deaths. There are some useful points here, but it is speculation that is beyond the findings reported here. The second is a pivot away from the results--addressing cost-effectiveness and advocating for the chlorine dispenser or a coupon intervention for which this review provides limited or no direct evidence. These departures from the data reported in the paper may be customary and entirely appropriate in economics; that is not my field. As a reviewer coming from an environmental health

perspective, however, I would suggest that the paper restrict itself to the effectiveness analysis, and leave the cost-effectiveness to be assessed in another study designed specifically for that purpose.

Conclusion

The authors deserve credit for their efforts to ascertain the effects of water treatment on child mortality, the key driver to determine its cost effectiveness. However, due mainly to its methodological shortcomings and potential for bias, I would caution against relying on this paper. I believe the paper can be improved, and perhaps will be as it moves through a serious peer review process. However, some of the basic weaknesses--such as the failure to start with a comprehensive protocol that specifies clearly the scope and methods of the review before the work is undertaken--cannot be fixed retroactively. The paper may prove to be a useful contribution by pointing out the signal the authors found but emphasizing the limited inferences that can be drawn.