TREATING WATER WITH CHLORINE AT POINT-OF-USE TO IMPROVE WATER QUALITY AND REDUCE CHILD DIARRHEA IN DEVELOPING COUNTRIES: A SYSTEMATIC REVIEW AND META-ANALYSIS

BENJAMIN F. ARNOLD AND JOHN M. COLFORD JR.*
Division of Epidemiology, School of Public Health, University of California, Berkeley, California

Abstract. We conducted a systematic review of all studies that measured diarrheal health impacts in children and the impact on water quality of point-of-use chlorine drinking water treatment. Twenty-one relevant studies were identified from > 856 screened abstracts. Data were extracted and combined using meta-analysis to provide summary estimates of the intervention effect. The intervention reduced the risk of child diarrhea (pooled relative risk: 0.71, 0.58–0.87) and it reduced the risk of stored water contamination with Escherichia coli (pooled relative risk: 0.20, 0.13–0.30). A major finding from this review is that nearly all trials on this topic have been short (median length was 30 weeks). Although not statistically significant, we observed an attenuation of the intervention’s reduction of child diarrhea in longer trials. Future studies with multi-year follow-up are required to assess the long-term acceptability and sustainability of health impacts shown by the shorter trials identified in this review.

INTRODUCTION

Diarrheal diseases cause a tremendous burden of illness in developing countries. The annual burden of diarrheal disease is estimated at 3.5 billion episodes and results in 1.8 million deaths in children worldwide.1,2 Prior reviews by Esrey and others3,4 and two systematic reviews5,6 suggest that between 20% and 35% of diarrhea episodes may be prevented by improved drinking water. Together, these findings have motivated the development and testing of numerous locally adapted water interventions and low-cost technologies to provide safe drinking water in developing countries. The systematic review of Fewtrell and others5 of water, sanitation, and hygiene interventions in developing countries found that point-of-use water treatment may be one of the most efficacious strategies to reduce diarrhea among people who lack access to safe water. The comprehensive review conducted by Clasen and others6 largely confirmed the findings by Fewtrell and others, reporting that point-of-use water treatment reduces diarrhea in people of all ages and in children < 5 years old.

A widely disseminated and tested method to improve the microbiologic quality of drinking water in the household is the use of dilute chlorine solution. Point-of-use chlorine disinfection uses one of two possible types of chlorine: sodium or calcium hypochlorite. Both compounds can be produced locally in developing countries. They are inexpensive, easy to distribute and use, and effective against most bacterial and viral pathogens.7 When added to water in tightly covered containers, volatilization is minimal, and chlorine disinfectants provide residual protection for many hours to days.7 The limitations of chlorine disinfection include limited effectiveness against parasites and reduced effectiveness, strong odor, and disagreeable taste in the presence of large quantities of organic material in the treated water.8 Chlorine disinfection is often combined with storage in a safe container (to prevent recontamination) and educational and motivational campaigns. This multi-faceted intervention, named the Safe Water System (SWS), was developed by the Centers for Disease Control and Prevention (CDC) and the Pan American Health Organization (PAHO).8–11 For the purpose of this study, we define point-of-use chlorine treatment as the addition of sodium or calcium hypochlorite to drinking water in the absence of a combined flocculant. We make this restriction because the treatment process differs sufficiently with the addition of a flocculant for us to consider the combined flocculant/disinfectant a separate intervention.

Here, we present the results of a systematic review and meta-analysis in which we evaluated the effects of point-of-use chlorine disinfection on the quality of water stored in the household and the evidence for health effects among children in less-developed countries. After the submission of the initial version of this manuscript, Clasen and others6 published their comprehensive systematic review that covers water quality interventions broadly and includes point-of-use chlorine treatment in children as subgroups in their analyses. Our review was conducted independently from the review by Clasen and others and, at points where the two analyses overlap, provides a rare opportunity to validate the two studies—an exercise that would be avoided, typically, to reduce duplication of research effort. The results presented here build on the review by Clasen and others by comparing health outcomes with water quality outcomes—a common proxy for health impacts. We also provide insight into the incremental effects of combining chlorine disinfection with safe storage and education and the effect of study length on the interventions’ impact on child diarrhea. Where appropriate, we use meta-analysis to derive summary effect estimates. We also use meta-analysis techniques to explore possible sources of heterogeneity in the effect estimates to better understand the varying results.

MATERIALS AND METHODS

Search strategy. Database searches were conducted using the Cochrane Library, EMBASE, LILACS, Medline, and Web of Science with the keywords summarized in Table 1.
The search conducted for this review targeted publications that addressed two primary outcomes: diarrheal disease in children and stored water quality. Two organizational websites were also searched: the Centers for Disease Control and Prevention’s (CDC) Safe Water System website (www.cdc.gov/safewater) and the World Health Organization’s (WHO) household water treatment and safe storage website (http://www.who.int/household_water/en). The final literature search was conducted on March 30, 2006. This review is limited to articles indexed before that date and to articles in English, Spanish, or French (the languages spoken by the authors). All titles and abstracts (if available) were examined, and the relevant articles were retrieved for review. After the database search, the bibliographies of relevant articles were searched to identify additional studies. No restrictions were put on study design, but studies were only analyzed if they included a comparison group for the intervention.

Selection criteria and data extraction. Four selection criteria were applied to relevant studies: 1) the study was designed to evaluate point-of-use chlorine treatment of drinking water, defined as treatment with sodium or calcium hypochlorite in the absence of a combined flocculant; 2) the study took place in a less-developed country, defined as any country not within a Class A region in the 2002 World Health Report; 3) the study was published in a refereed journal (to maintain quality through peer review and to improve transparency); and 4) the study included an outcome related to diarrheal disease in children (< 18 years old) or stored water quality.

Data related to the intervention and outcomes were extracted, tabulated, and, if appropriate, pooled using meta-analysis. The authors independently conducted unblinded, duplicate data extraction for a random sample of 50% of the publications using standardized forms and definitions. The duplicate entries had perfect correspondence in the subset of publications, so duplicate entry for the remaining publications was not conducted. If multiple articles described the same study, we extracted design details and used the most recent results in the analyses. As a conservative measure, if a single intervention trial included multiple arms with chlorine treatment, we reported data from all arms, but only one arm was included in meta-analyses. In cases where a single study reported results for multiple chlorine intervention arms, the most complete intervention (defined by the education or safe storage components) was used in the analyses. If a study reported multiple water quality estimates, the average of the estimates was used in the analysis. If a single study provided age-stratified estimates, the estimate from the youngest age category was used in the analysis of child diarrhea. If investigators did not report risk measures, we extracted the original data from the article and calculated the appropriate measure of relative risk (odds ratio, incidence density ratio, or cumulative incidence ratio) with a 95% confidence interval using standard techniques. If authors published adjusted relative risk estimates, our analysis used the estimate that adjusted for the most covariates. For all analyses, relative risks are expressed such that values < 1.0 indicate a reduction in disease or improvement in water quality that results from point-of-use chlorine treatment of drinking water. The comparison group in all analyses is traditional water use (no intervention treatment).

We evaluated the quality of each study using a set of criteria suggested by Blum and Feachem and previously implemented by Fewtrell and others. Health outcome studies were deemed poor quality if they had any of the following design flaws: inadequate or inadequately designed control groups, no clear measurement or control of confounders, no specific definition of diarrhea or the gastrointestinal health outcome definition, or a recall period between the occurrence of illness and the recording of illness > 2 weeks. No study was excluded from the analysis on the basis of the quality criteria, but if possible, quality was examined as a source of heterogeneity between the results.

Meta-analysis. The meta-analysis portion of the review focuses on two primary outcomes from point-of-use drinking water treatment using chlorine: 1) the intervention’s impact on child diarrhea and 2) stored water quality, assessed using the proportion of stored water samples with no detectable *Escherichia coli*. Mean concentrations of indicator bacteria were not analyzed in the water quality meta-analyses because only one study reported variances for bacteria concentration effect estimates.

We pooled and analyzed risk estimates from selected stud-
ies using Stata software (version 9; Stata Corp., College Station, TX). We pooled health results across multiple measures of relative risk after determining that the outcomes were sufficiently rare in the study populations and assuming that exposure did not have a large effect on person-time at risk.\textsuperscript{13,16} We used fixed-effects and random-effects models with weights equal to the inverse of the variance to calculate a pooled effect of chlorine treatment on diarrhea incidence and water quality.\textsuperscript{17} We used a Mantel-Haenszel heterogeneity test using inverse variance weights to test for heterogeneity across study results. If the test for heterogeneity across study results was significant (defined conservatively as \( P < 0.20 \)), we used random-effects estimates to summarize the data. We evaluated publication bias using the Begg test, with \( P < 0.20 \) indicating the potential presence of bias.\textsuperscript{18} The influence of individual studies was evaluated by serially excluding each result and re-estimating the pooled effect.

Subgroup analyses and meta-regression were used to explore possible sources of heterogeneity across study results. Specifically, study quality, the length of the intervention, the setting of the intervention (rural, peri-urban, or urban), baseline diarrhea rates in the control group, use of a safe storage container in the intervention, inclusion of an educational or motivational component in the intervention, and proportion of stored water samples in the intervention group with detectable free chlorine (a measure of compliance) were defined \textit{a priori} as potential sources of heterogeneity and, where possible, explored in the analysis. All significance tests were evaluated using a conservative Bonferroni correction for multiple hypothesis tests (dividing the \( \alpha \) level of 0.05 by the number of comparisons).

**RESULTS**

**Search results.** The broad search criteria identified 17,058 publications. An initial title screen yielded 856 publications that were screened by abstract (if available). A total of 50 publications were reviewed in full, and 22 publications (representing 20 studies) met the inclusion criteria. Publications were excluded because they were reviews or editorials (13),\textsuperscript{5,8,10,11,19–27} did not have diarrhea or water quality outcomes (6),\textsuperscript{28–32} were not published in refereed journals (4),\textsuperscript{34–37} did not have child health outcomes (2),\textsuperscript{38,39} or did not test point-of-use chlorine treatment (3).\textsuperscript{40–42} The final studies were conducted in 13 different countries and a variety of settings (Tables 2 and 3). Of the final 22 publications, 12, representing 10 studies, included child health outcomes and diarrhea was the only health endpoint (Table 2).\textsuperscript{43–54} We validated our search results by applying our selection criteria to the population of studies identified by Clasen and others\textsuperscript{6} and found perfect correspondence. The definition of children varied across studies, with the youngest age definition < 1 year and the oldest age definition < 15 years. With one exception,\textsuperscript{45} all studies showed a protective effect of point-of-use chlorine treatment on child diarrhea (Table 2). A single study measured diarrhea outcomes but did not provide child-specific estimates.\textsuperscript{55} It was used in the water quality analysis, but it was not included in the child health outcome meta-analysis. A single study provided estimates stratified by HIV status, and the estimate for HIV negative children was used in the meta-analysis.\textsuperscript{46}

Eight studies included both child health and water quality outcomes,\textsuperscript{43,45,48–53} and 10 additional studies included only water quality outcomes.\textsuperscript{15,55–60} By all measures of indicator bacteria (\textit{E. coli}, fecal coliforms, total coliforms), point-of-use chlorine treatment dramatically improved stored water quality in the 18 studies (Table 3).

Two studies included multiple intervention arms with point-of-use chlorine treatment; in both cases, only one arm from each trial was included in the pooled effect estimates. Reller and others\textsuperscript{45} included chlorine treatment arms with and without safe storage, and the treatment arm with safe storage (the more extensive intervention) was included in the analysis. Luby and others\textsuperscript{46} included chlorine treatment arms with safe storage in an imported vessel and safe storage with a local vessel. The treatment arm with the imported vessel was included in the analysis because the length of study using that intervention arm was longer (34 versus 26 weeks), and the local vessel was not part of the originally planned trial. We also excluded from our analysis the second year of follow-up in the original intervention arm of Luby and others because the intervention population during the second year of the study was a mixed set of new and long-term participants; this mixed level of experience in the intervention group during the second year was inconsistent with other studies included in this review.

**Quality of publications.** Two studies did not have a specific definition of diarrhea and instead had mothers in the study define episodes of diarrhea for their children.\textsuperscript{55,51} Of the remaining nine studies deemed to have adequate outcome definitions, eight defined an episode of diarrhea as the occurrence of either three or more loose or watery stools or one bloody stool, in a 24 hour period, and one defined an episode as a significant change in bowel habits marked by an increased frequency or decreased consistency.\textsuperscript{55} In all studies, outcomes were reported by the mother, and there were no data with which the possibility of resulting measurement error of the outcome could be assessed. With regard to our second quality criteria, Sobsey and others\textsuperscript{53} described control for potential confounders during the analysis phase but only presented unadjusted risk estimates in the published article. However, the paper published by Sobsey and others summarizes two intervention trials, one that was included separately in this review\textsuperscript{50} and a second that was not published in a refereed journal but was readily available online. The second study, which took place in Bangladesh, was also published by Handzel as a PhD dissertation with detailed methods and adjusted risk estimates.\textsuperscript{44} We chose a conservative approach by using the adjusted risk estimate from Handzel in the analysis because its point estimate was similar to the unadjusted estimate (0.67 versus 0.78), but its variance was three times larger, likely reflecting a proper adjustment for repeated measurements on individuals.\textsuperscript{44}

**Diarrhea in children.** Point-of-use chlorine treatment of drinking water reduced diarrhea in children in 9 of 10 studies identified, and the effect was statistically significant in 5 of the studies (Figure 1). The Mantel-Haenszel test for heterogeneity indicated that there was significant heterogeneity across studies (\( X^2_{(5)} = 38.10, P < 0.001 \)). In addition, the \( I^2 \) statistic, which estimates the fraction of the variation in the effect estimate caused by heterogeneity, indicated that heterogeneity accounted for 76% of the variation in these data.\textsuperscript{64} As a result of this observed heterogeneity, a random effects model
was used to calculate a pooled effect estimate. The pooled effect estimate across the 10 studies indicated that point-of-use treatment of drinking water with chlorine reduces diarrhea in children by 29% (relative risk, 0.71; 0.58–0.87). There was no evidence of publication bias based on the Begg test (Z = −0.63, P = 0.53), and sensitivity analysis indicated that no single study had a disproportionate impact on the summary effect estimate (results available from the authors).

There were a sufficient number of studies to evaluate possible sources of heterogeneity in the effect using subgroup and meta-regression analyses. Inclusion of a safe-storage vessel and inclusion of an educational or motivational component in the intervention were perfectly collinear in the data: every study that included safe storage also included education or motivation. It is impossible to disentangle their independent effects using these data, but it was still possible to con-
### TABLE 3
Studies of point of use chlorine water treatment with water quality outcomes

<table>
<thead>
<tr>
<th>Reference, by indicator bacteria</th>
<th>Country (setting)</th>
<th>Safe storage</th>
<th>Measure</th>
<th>CFU per 100 mL water*</th>
<th>Percentage of intervention samples with detectible free chlorine</th>
<th>Number (%) of stored water samples with detectible E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eschericia coli</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahfouz and others 1995&lt;sup&gt;49&lt;/sup&gt;</td>
<td>Saudi Arabia (rural)</td>
<td>None</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR (100)</td>
</tr>
<tr>
<td>Quick and others 1996&lt;sup&gt;15&lt;/sup&gt;</td>
<td>Bolivia (peri-urban)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>28</td>
<td>2.2</td>
<td>93</td>
</tr>
<tr>
<td>Daniels and others 1999&lt;sup&gt;52&lt;/sup&gt;</td>
<td>Guinea-Bissau (urban)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>6,200</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Quick and others 1999&lt;sup&gt;50&lt;/sup&gt;</td>
<td>Bolivia (peri-urban)</td>
<td>Imported vessel</td>
<td>Median</td>
<td>9,200</td>
<td>0</td>
<td>77</td>
</tr>
<tr>
<td>Sobsey and others 2003&lt;sup&gt;53&lt;/sup&gt;</td>
<td>Pakistan (urban)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>7</td>
<td>0.5</td>
<td>91</td>
</tr>
<tr>
<td>Mong and others 2001&lt;sup&gt;59&lt;/sup&gt;</td>
<td>Madagascar (rural)</td>
<td>Local vessel</td>
<td>Median</td>
<td>13</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>Ogutu and others 2001&lt;sup&gt;60&lt;/sup&gt;</td>
<td>Kenya (rural)</td>
<td>None</td>
<td>Median</td>
<td>100</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Quick and others 2002&lt;sup&gt;55&lt;/sup&gt;</td>
<td>Zambia (peri-urban)</td>
<td>Imported vessel</td>
<td>Median</td>
<td>44</td>
<td>0</td>
<td>78</td>
</tr>
<tr>
<td>Rangel and others 2003&lt;sup&gt;54&lt;/sup&gt;</td>
<td>Guatemala (rural)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>324</td>
<td>6</td>
<td>83</td>
</tr>
<tr>
<td>Reller and others 2003&lt;sup&gt;51&lt;/sup&gt;</td>
<td>Guatemala (rural)</td>
<td>Imported vessel</td>
<td>Median</td>
<td>73</td>
<td>NR</td>
<td>44</td>
</tr>
<tr>
<td>Reller and others 2003&lt;sup&gt;51&lt;/sup&gt;</td>
<td>Guatemala (rural)</td>
<td>None</td>
<td>Median</td>
<td>53</td>
<td>NR</td>
<td>36</td>
</tr>
<tr>
<td>Sobsey and others 2003&lt;sup&gt;53&lt;/sup&gt;</td>
<td>Bangladesh (urban)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>39</td>
<td>1</td>
<td>89</td>
</tr>
<tr>
<td>Crump and others 2004&lt;sup&gt;56&lt;/sup&gt;</td>
<td>Kenya (urban)</td>
<td>None</td>
<td>Mean</td>
<td>3,938</td>
<td>0.5</td>
<td>NA</td>
</tr>
<tr>
<td>Crump and others 2005&lt;sup&gt;53&lt;/sup&gt;</td>
<td>Kenya (rural)</td>
<td>None</td>
<td>Mean</td>
<td>116</td>
<td>NR</td>
<td>61</td>
</tr>
<tr>
<td>Lule and others 2005&lt;sup&gt;58&lt;/sup&gt;</td>
<td>Uganda (rural)</td>
<td>Imported vessel</td>
<td>Median</td>
<td>248</td>
<td>5,117</td>
<td>23</td>
</tr>
<tr>
<td><strong>Fecal Coliforms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kirchhoff and others 1985&lt;sup&gt;45&lt;/sup&gt;</td>
<td>Brazil (rural)</td>
<td>None</td>
<td>Mean</td>
<td>1,600,000</td>
<td>7,000</td>
<td>NR</td>
</tr>
<tr>
<td>Quick and others 1996&lt;sup&gt;15&lt;/sup&gt;</td>
<td>Bolivia (peri-urban)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>34</td>
<td>2.2</td>
<td>93</td>
</tr>
<tr>
<td>Semenza and others 1998&lt;sup&gt;52&lt;/sup&gt;</td>
<td>Uzbekistan (unstated)</td>
<td>Imported vessel</td>
<td>Mean</td>
<td>54</td>
<td>52</td>
<td>73</td>
</tr>
<tr>
<td>Sobel and others 1998&lt;sup&gt;63&lt;/sup&gt;</td>
<td>Guatemala (urban)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>7</td>
<td>2</td>
<td>NR</td>
</tr>
<tr>
<td>Roberts and others 2001&lt;sup&gt;52&lt;/sup&gt;</td>
<td>Malawi (refugee camp)</td>
<td>None</td>
<td>Mean</td>
<td>625</td>
<td>250</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Total Coliforms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sobel and others 1998&lt;sup&gt;63&lt;/sup&gt;</td>
<td>Guatemala (urban)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>54</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Daniels and others 1999&lt;sup&gt;57&lt;/sup&gt;</td>
<td>Guinea-Bissau (urban)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>34,000,000</td>
<td>360</td>
<td>NA</td>
</tr>
<tr>
<td>Luby and others 2001&lt;sup&gt;58&lt;/sup&gt;</td>
<td>Pakistan (urban)</td>
<td>Imported vessel</td>
<td>Geometric Mean</td>
<td>9,397</td>
<td>47</td>
<td>91</td>
</tr>
<tr>
<td>Crump and others 2004&lt;sup&gt;56&lt;/sup&gt;</td>
<td>Kenya (rural)</td>
<td>None</td>
<td>Mean</td>
<td>25,553</td>
<td>12</td>
<td>NA</td>
</tr>
</tbody>
</table>

* CFU, colony-forming units. Samples from stored water except the baseline estimates from Crump and others 2004, which are measurements from source water.
† Proportion of samples with >1 E. coli per 100 mL of drinking water.
‡ Excluded from meta-analysis because number of samples in control group not reported.
§ Included in over-all meta-analysis.
¶ Omitted from meta-analysis due to multiple trial arms in a single study.
RCT, randomized controlled trial; EXP, controlled experiment; CS, cross-sectional; NA, not applicable; NR, not reported.
duct analysis on their joint effect. Similar near-collinearity between safe storage, education, and rural setting was also present in the data: of the five studies conducted in a rural setting, three did not include safe storage or education. Because of this lack of variability in the study designs, we recommend that results from the subgroup analyses of setting and safe storage/education should be used for hypothesis generation only.

Restricting the analysis to only high-quality studies (adequate control group, control for confounders, clear outcome definition, and < 2-week recall period) resulted in a relative risk estimate of 0.66 (0.51–0.87). Additional subgroup analyses suggest an enhanced effect of safe storage and education and urban or peri-urban settings. The pooled random effects estimate from the seven studies that included safe storage and education was 0.65 (0.46–0.80), which was larger than the pooled estimate for chlorine treatment alone (without safe storage or education; relative risk, 0.87; 0.62–1.22). The three studies conducted in urban or peri-urban settings showed a larger effect (relative risk, 0.63; 0.50–0.80). Subgroup analyses of the effect by child age definition yielded similar results to the pooled analysis but with less precision because of fewer studies: < 5 years (relative risk, 0.71; 0.62–1.22), < 3 years (relative risk, 0.81; 0.61–1.06), < 2 years (relative risk, 0.80; 0.59–1.08), and < 1 year (relative risk, 0.69; 0.44–1.09). We did not evaluate underlying risk as a possible source of heterogeneity because the variance in the log baseline rates of diarrhea in control groups across studies was sufficiently small (μ = 1.12, σ = 1.53) to bias their use as a measure of underlying risk caused by regression to the mean. Although Bayesian techniques exist to address this issue, they are beyond the scope of this analysis, given its relatively small number of studies.

In this analysis, there are 10 studies with child diarrhea outcomes and a large degree of heterogeneity between estimates ($I^2 = 0.76$), a scenario that may reduce the validity of meta-regression models with multiple covariates. Consequently, we conducted only bivariate analyses using the continuous covariates of intervention length and proportion of water samples with detectable free chlorine with the understanding that they should be used mainly for hypothesis generation. This is consistent with the overall limitation of observational analyses, which in this context cannot evaluate causal relationships between trial level attributes.

Figure 2 plots the effect size against study length and the fraction of water samples with detectable free chlorine (a measure of compliance). The meta-regression of the bivariate association between intervention length (weeks) and child diarrhea identified an attenuation of the reduction of child diarrhea among studies of longer length (relative risk of a 10-week increase, 1.07; 0.96–1.19; N = 10 studies, $\tau^2 = 0.0661$; proportion of heterogeneity explained = 0.23). The intervention’s effectiveness was enhanced among studies with a larger fraction of water samples with detectable free chlorine, a measure of compliance (relative risk of a 10% increase, 0.90; 0.77–1.05; N = 6 studies, $\tau^2 = 0.0767$; proportion of heterogeneity explained = 0.20). Neither result is statistically significant at the 95% confidence level after Bonferroni correction.

Water quality. Consistent with its effect on child diarrhea, point-of-use treatment of drinking water with chlorine improved water quality in nearly all studies and all measurements reviewed (Table 3). There exists significant heterogeneity in the effect across water quality studies ($X^2_{9} = 98.48, P < 0.001; F^2 = 0.91$) and so we used a random-effects model to estimate a pooled effect of the intervention on the probability of testing positive for E. coli contamination in stored water samples (Figure 3). Across the 10 studies,
the intervention led to an 80% reduction in the proportion of stored water samples with detectable *E. coli* (relative risk, 0.20; 0.13–0.30). The results from the Begg test indicated that there was no evidence of publication bias in water quality studies ($Z = -0.98$, $P = 0.33$), and sensitivity analysis indicated that no single study had a disproportionate impact on the summary effect estimate (results available from the authors).

Subgroup analyses produced highly consistent effect estimates across all study subgroups. Restricting the analysis to just randomized controlled field trials by excluding the two controlled experimental studies resulted in a relative risk of 0.21 (0.14–0.32). The eight studies that included safe storage and education in addition to chlorine disinfectant had a pooled relative risk of 0.19 (0.12–0.31), and the two studies that did not include safe storage or education (chlorine alone) had a pooled relative risk of 0.25 (0.19–0.32). The observed differences between rural and urban/peri-urban settings in the intervention’s effect on child diarrhea were not apparent in the *E. coli* contamination outcome. The pooled relative risk of *E. coli* contamination among the five rural studies was 0.21 (0.12–0.38). Among the five urban and peri-urban studies, the relative risk was 0.18 (0.10–0.33). There was minimal difference in the effect estimates and confidence intervals between studies that measured the proportion of samples with detectable *E. coli*.

**Figure 2.** The effect of point-of-use chlorine treatment on child diarrhea by length of intervention (A) and proportion of intervention households with detectable free chlorine (B). The area of each circle reflects the study weight in the random effects meta-analysis.

![Figure 2](image1.png)

**Figure 3.** Random effects meta-analysis of point-of-use chlorine drinking water treatment and *E. coli* contamination of stored water.

<table>
<thead>
<tr>
<th>Study</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniels et al, 1999</td>
<td>0.01 (0.00–0.22)</td>
</tr>
<tr>
<td>Quick et al, 1999</td>
<td>0.30 (0.23–0.51)</td>
</tr>
<tr>
<td>Luby et al, 2004</td>
<td>0.04 (0.02–0.09)</td>
</tr>
<tr>
<td>Quick et al, 2002</td>
<td>0.32 (0.20–0.52)</td>
</tr>
<tr>
<td>Rangel et al, 2003</td>
<td>0.09 (0.04–0.20)</td>
</tr>
<tr>
<td>Reller et al, 2003</td>
<td>0.50 (0.43–0.57)</td>
</tr>
<tr>
<td>Sobsey et al, 2003</td>
<td>0.23 (0.19–0.28)</td>
</tr>
<tr>
<td>Crump et al, 2004</td>
<td>0.18 (0.08–0.40)</td>
</tr>
<tr>
<td>Crump et al, 2005</td>
<td>0.26 (0.20–0.33)</td>
</tr>
<tr>
<td>Lule et al, 2005</td>
<td>0.13 (0.05–0.33)</td>
</tr>
<tr>
<td>Overall</td>
<td>0.20 (0.13–0.30)</td>
</tr>
</tbody>
</table>
and studies that measured the proportion of samples with > 1 \(E. coli\) per 100 mL (results available from the authors).

DISCUSSION

One difficulty of systematic reviews and meta-analyses is their attempt to synthesize results from data that were collected under widely varying conditions. In this review, the intervention of interest—point-of-use chlorine treatment of drinking water to improve water quality and reduce diarrhea in children—was implemented using different strategies in a large number of developing countries around the world (Tables 2 and 3). Although it is important to be aware of this potential pitfall in meta-analysis, our finding of a near-universal protective effect across very different conditions, populations, and investigators suggests that the intervention effectively reduces diarrhea in children. Similarly strong evidence of a beneficial effect was also seen when we used \(E. coli\) contamination as the outcome.

The pooled effect of point-of-use chlorine treatment on child diarrhea indicates a 29% reduction in risk (relative risk, 0.71; 0.58–0.87) compared with traditional practices (Figure 1). The results from this review are consistent with findings from two broader reviews of water interventions in developing countries. Fewtrell and others found a pooled effect for household water treatment of 0.65 (0.48–0.88). This consistency is probably because of important overlap: 5 of the 12 studies included in the pooled estimate calculated by Fewtrell and others were also included in this analysis. Our search results and findings are also consistent with those of Clasen and others, although direct comparison is difficult because they report results stratified by measure of relative risk and include multiple arms from the same trials in some of their comparisons. Clasen and others reported pooled effects in children < 5 years that range between 0.48 (0.33–0.68) and 0.91 (0.82–1.02). Our pooled estimate for children < 5 years (relative risk, 0.71; 0.56–0.89) falls near the center of this distribution, which would be predicted because the estimates are derived from the same population of studies. Taken together, these findings support a growing consensus that household water treatment methods may be a key transitional precursor to large capital water projects for the large fraction of the world’s population who currently lacks access to clean water. At the very least, such treatment methods offer an evidence-supported choice for interim protection while millions of families continue to await large-scale system-wide projects.

The difference between the intervention’s impact on water quality (relative risk, 0.20; 0.13–0.30) and child diarrhea (relative risk, 0.71; 0.58–0.97) likely arises from imperfect compliance and competing, non–water-borne causes of diarrheal disease. Indeed, we observed larger reductions in child diarrhea with better compliance (Figure 3B) as did Clasen and others. It is also possible that alternate pathways such as person-to-person or environment-to-person may be important for infants and school aged children. Even among water-borne pathogens, \(E. coli\) is just one of many pathogens that cause diarrhea. Point-of-use chlorine treatment is not highly effective against parasites, and if parasites are an important cause of gastrointestinal infection, the reductions in \(E. coli\) contamination would not translate directly into health effects. The large difference in the intervention’s impact between water quality and child diarrhea suggests that water quality is an insufficient surrogate for child health improvements in this context.

An important limitation of nearly all the health outcome studies in this review is that the outcomes were reported by participants, and in all but one study, the control groups were not blinded to the intervention. This combination of factors can potentially lead to differential misclassification of the outcome if participants or assessors realize that the reduced diarrhea is a “desirable outcome” of the intervention. If differential misclassification occurs, it would likely bias the effect away from the null, and the estimated effect sizes may be larger than their true values. The single-blinded study in our analysis found no effect of the intervention, which is consistent with this potential problem. A second blinded trial from Gambia that was not published in refereed journals but was reported in the review by Clasen and others also found no effect of point-of-use chlorine treatment in children (relative risk, 1.05; 0.25–4.35). The challenge in this context is that it is impossible to blind participants to the use of improved vessels or educational materials—potentially important complementary interventions for chlorine treatment. Where possible, participants should be blinded to their treatment assignments, and if blinding is impossible, trials would be improved by validation studies to estimate the degree of bias resulting from the use of unblinded, self-reported outcomes.

We found heterogeneity across studies in both diarrhea and water quality outcome measures. Although our subgroup analyses and meta-regression results explain some of the variability between health outcome studies, much of the heterogeneity likely arises from site-specific differences in cultural practices, pre-intervention conditions, and underlying risk from water exposure. We view the results of the heterogeneity analyses as primarily hypothesis-generating because of the small number of studies and the relatively small amount of variability in the design across studies. Some of this homogeneity in design is an artifact of our analysis groupings. For example, there is large variability in the form of education included with the chlorine water treatment intervention: Semenza and others provided hygiene education to intervention households (no additional information specified), but Luby and others had an extensive education and motivational component, which combined twice-weekly visits to intervention households with community education using slides, videos, and pamphlets. Ideally our analysis would account for this type of variability, but it is impossible to do so without a larger number of studies.

An important question raised by our analysis is whether the large health impacts observed during shorter trials persist over longer periods. With two exceptions, nearly all of the trials to date have been relatively short—the longest was 20 months and the median length was 6 months (Table 2). Figure 3 shows an attenuated effect in the reduction of child diarrhea among longer trials, but this finding lacks a causal interpretation and could be confounded by other trial attributes. This trial-level data contrasts with the hypothesis that technical mastery and habituation of the interventions increases with time, so increased effectiveness should be observed with
longer follow-up; a hypothesis with some support from individual-level data. However, an alternate hypothesis is that the diminished effectiveness of the intervention over longer time periods could result from gradual lack of interest among participants leading to lower compliance, or ultimately, abandonment of the intervention. It could also result from variable effectiveness of the intervention across different seasons. For example, the longest study in this review documented negligible differences between intervention and control group diarrhea prevalence during the peak of the rainy season but found large differences during other seasons. This type of common seasonal trend could diminish the measured health effects observed in longer trials compared with shorter trials if the shorter trials were conducted during seasons in which the differences were greatest. Only long-term follow-up of health and behavioral outcomes among individuals can resolve this important question. Further advances in this field will also require validation of self-reported health outcomes and assessment of long-term acceptability and sustainability of health impacts shown by the shorter trials identified in this review. Ultimately the development of strategies for large-scale programmatic dissemination of this water treatment method must incorporate these findings into their design and evaluation.

Received June 28, 2006. Accepted for publication October 30, 2006.

Acknowledgments: We are grateful for the useful comments that Catherine Wright provided during preparation of the manuscript. We also appreciate the detailed and constructive feedback provided by the editors and referees of this journal.

Authors’ addresses: Benjamin F. Arnold, Division of Epidemiology, School of Public Health, University of California, Berkeley, 140 Warren Hall 7360, Berkeley, CA 94720–7360, E-mail: benarnold@berkeley.edu. John M. Colford Jr., Division of Epidemiology, School of Public Health, University of California, Berkeley, 140 Warren Hall 7360, Berkeley, CA 94720–7360, E-mail: jcolford@berkeley.edu.

REFERENCES


