Community surveys of self-reported diarrhoea can dramatically overestimate the size of outbreaks of waterborne cryptosporidiosis

P.R. Hunter and Q. Syed
Communicable Disease Surveillance Centre - North West, Public Health Laboratory,
University Hospital Aintree, Longmoor Lane, Liverpool L9 7AL, UK (E-mail: phunter@nw.phls.nhs.uk)

Abstract This paper describes a community-based study undertaken to assess the size of a waterborne outbreak of cryptosporidiosis in the North West region of England. The outbreak was linked to a single reservoir in the English Lake District and provided drinking water to over 1.2 million people. There were some 308 laboratory confirmed cases. We conducted a community-based survey for self-reported diarrhoea in four towns within the outbreak area and four control towns. The rate of self-reported diarrhoea was higher in the control towns than in the outbreak towns. It would appear that retrospective community-based studies of diarrhoeal disease are subject to recall bias that would overestimate the incidence of illness, especially following reporting in the media. In the light of our findings, we reviewed the study undertaken during the Milwaukee outbreak that produced the estimated size of 405,000 cases. It is suggested that the estimate of the size of the Milwaukee outbreak is severely flawed, and the actual size of this outbreak was between 1% and 10% of that claimed.

Keywords Cryptosporidiosis; outbreak; waterborne; epidemiology; Milwaukee; bias

Introduction Cryptosporidiosis has the distinction of being the cause of the world’s largest reported outbreak of disease linked to drinking water (MacKenzie et al., 1994). During the three month period 1st March to 30th May 1993 there were some 739 laboratory confirmed cases of cryptosporidiosis in the Milwaukee area. However, a much larger figure of 403,000 cases was attributed to this outbreak. This latter figure was derived from a community-based survey where people were asked whether they had suffered from diarrhoea between 1st March and 28th April. A recent study in the UK suggested that retrospective surveys of subjective diarrhoeal disease may substantially over-estimate the true population level of illness (Wheeler et al., 1999). If this is the case, then the estimated size of the Milwaukee outbreak is also an overestimate. We took the opportunity provided by an outbreak of cryptosporidiosis in the North West of England to conduct a study into the validity of population-based surveys for estimating the true size of waterborne outbreaks.

The outbreak The outbreak, around which this study was done, occurred during April and May 1999 (CDSC - NW, 2000). A total of 308 laboratory confirmed cases were identified in just four health authority areas. The majority of the water supplied to the areas affected by this outbreak came from a single reservoir in the English Lake District, Thirlmere reservoir. The outbreak was deemed to be strongly associated with drinking water from this reservoir according to UK advice (Tillett et al., 1998). The decision to classify this outbreak as strongly associated with water was taken on the basis of the descriptive epidemiology which showed a strong correlation between attack rates in relation to amount of Thirlmere water supplied to each water zone and did not identify another plausible explanation. In addition, a 10 L grab sample of water from the implicated reservoir taken before the start of
the outbreak yielded a count of 34 oocysts. Water from the reservoir was treated by chlorination, but not filtration. Finally oocysts were detected in faecal samples from sheep grazing in the catchment area of the reservoir. Early in the investigation of the outbreak, it was clear that this was probably associated with drinking water and coverage in the media reflected this early conclusion. The drinking water distribution systems in the UK are more complex than those in many countries and many consumers have a mix of water derived from more than one source. The exact mix and source of an individual’s drinking water can vary from time to time. It was estimated that at the time of the outbreak some 1,260,160 people took at least 10% of their water from the implicated reservoir.

**Community survey**

Eight towns were identified for this study of which four were within the area supplied by the suspect reservoir and were known to be affected by the outbreak. Four control towns were chosen from close to but outside the affected area. A questionnaire was sent by post to 120 households in each town. The names and addresses were chosen at random from up-to-date telephone directories. Non-responders received a second questionnaire two weeks later and a subsequent telephone call if there was still no response a week later. The questionnaire asked only a very limited number of questions; the age and sex of each member of the household, whether they drank unboiled tapwater and if so how much each day, whether they had had diarrhoea since 15 April and, if so, when it start, how long had the illness lasted, had they visited their doctor and had they lost any time from work.

**Estimation of outbreak size**

Of the 960 questionnaires sent out, 701 (73%), covering 1,613 individuals, were able to be included in the study. Table 1 shows the response rate for each of the affected towns, the number of people covered and the attack rates for self-reported diarrhoea. The attack rate for towns in the outbreak area was 12.8% (95% CI, 10.5 –15.2).

As mentioned above, the estimate of the size of the Milwaukee outbreak was based on a telephone survey of 840 households of which there were 613 (73%) included in the study. Within these 613 households there were 1,663 people of whom 26% reported diarrhoea. Extrapolating up to the 1,610,000 population this gave a total of 419,000. The investigators estimated a background rate of 0.5% per month, about 16,000 to give a total size of 405,000.

<table>
<thead>
<tr>
<th>Town</th>
<th>% response rate</th>
<th>Total responders</th>
<th>Total reporting diarrhoea</th>
<th>% attack rate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Affected towns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chorley</td>
<td>87.5</td>
<td>227</td>
<td>26</td>
<td>11.5 (7.6–16.3)</td>
</tr>
<tr>
<td>Preston</td>
<td>66.7</td>
<td>196</td>
<td>17</td>
<td>8.7 (5.1–13.5)</td>
</tr>
<tr>
<td>Salford</td>
<td>70.8</td>
<td>171</td>
<td>28</td>
<td>16.4 (11.2–22.8)</td>
</tr>
<tr>
<td>Wigan</td>
<td>70.8</td>
<td>221</td>
<td>33</td>
<td>14.9 (10.5–20.3)</td>
</tr>
<tr>
<td>All</td>
<td>71.3</td>
<td>15</td>
<td>104</td>
<td>12.8 (10.5–15.2)</td>
</tr>
<tr>
<td><strong>Control towns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blackburn</td>
<td>61.7</td>
<td>159</td>
<td>28</td>
<td>17.6 (12.0–24.4)</td>
</tr>
<tr>
<td>Ormskirk</td>
<td>87.5</td>
<td>255</td>
<td>22</td>
<td>8.6 (5.5–12.8)</td>
</tr>
<tr>
<td>Rochdale</td>
<td>72.5</td>
<td>193</td>
<td>20</td>
<td>10.4 (6.4–15.6)</td>
</tr>
<tr>
<td>Warrington</td>
<td>77.5</td>
<td>191</td>
<td>38</td>
<td>19.9 (14.5–26.3)</td>
</tr>
<tr>
<td>All</td>
<td>74.8</td>
<td>798</td>
<td>108</td>
<td>13.5 (11.2–16.1)</td>
</tr>
<tr>
<td><strong>All towns</strong></td>
<td><strong>73.1</strong></td>
<td><strong>1,613</strong></td>
<td><strong>212</strong></td>
<td><strong>13.1 (11.5–14.9)</strong></td>
</tr>
</tbody>
</table>

Table 1 Population covered by respondents and diarrhoeal attack rates for each town, for all towns within and without the outbreak area and for all towns.
Using the approach adopted by the investigators of the Milwaukee outbreak we can come up with an estimate of the size of the outbreak in the North West. Using this estimate in an outbreak would give an overall attack rate of 12.8%. Using the Milwaukee approach, this would give a total size of 12.8% of 1,260,160, i.e. 161,300 cases. Taking the same background rate of 0.5% per month would give a figure of 9,500 to give 151,800 cases.

We suggest that there are two inter-related sources of error in this approach to estimating the size of outbreaks of diarrhoeal disease. The first is the estimate of the background incidence of disease and the second is the potential for recall-bias.

It is not clear how the investigators in the Milwaukee outbreak decided on the figure of 0.5% per month in relation to the background levels of diarrhoeal disease. In the US the reported estimates of annual diarrhoeal risk are about 1.4 episodes per person per year (Table 2). This gives a monthly level of about 11.7%, substantially greater (23,400%) than that estimated by the Milwaukee group. Using the US FoodNet figure (Foodnet, 1997) the background number of cases over the two months would be 376,740 to give the actual size of the Milwaukee outbreak as 42,260 only 10.4% of the original estimate.

In the UK the annual reported incidence of diarrhoeal disease derived from retrospective studies is lower than in the US at about 0.8 episodes/person/year or about 6.7% per month (Table 2). Using this estimate for the North West outbreak we have a background of 126,650 to give a total size of 34,650, 22.8% of the original estimate.

A further problem with using retrospective surveys, like the ones described here, is that such surveys seem to over-estimate the true level of diarrhoeal illness in a community compared to prospective studies by a factor of about 2.8 (Wheeler et al., 1999). This would further reduce the estimated size of these two outbreaks to about 15,090 for the Milwaukee outbreak and 12,375 for the NW England outbreak.

Unfortunately this approach to estimating outbreak size is undermined by the finding that the reported incidence of diarrhoea during the same period in the control towns is actually higher at 13.5% (95% CI, 11.2 –16.1). This severely limits the potential to calculate outbreak size from such a community-based survey. A recent study (Wheeler et al., 1999) of infectious intestinal disease in the UK compared the incidence of a variety of intestinal pathogens in a community-based study with those being diagnosed by the family doctor. The ratio of GP diagnosis to community incidence was 1.9:1 (95% CI 0.6 – 6.1). Using this ratio to calculate the size of the two outbreaks we have a result of some 585 cases for the UK outbreak and 1,400 for the Milwaukee outbreak.

**Discussion**

The primary aim of this study was to validate community surveys as a means of estimating the total size of outbreaks of cryptosporidiosis. The study reported here is essentially the same as
that done in Milwaukee although it differed in one respect. In our study we used a postal questionnaire and telephoned respondents only if they had not replied to our initial two mailings. In Milwaukee, respondents were initially contacted by phone. In our study we also did not specify that diarrhoea should be "watery". The reason for this was that a high proportion of our laboratory confirmed cases reported having had bloody diarrhoea (CDSC NW, 2000).

We have demonstrated a greater prevalence of diarrhoeal illness in both outbreak and control towns than would be considered normal in the UK from results of retrospective studies (13.5% in the control towns compared to 10.1%). As already discussed, the control towns were chosen because of their close proximity to the outbreak area. Although we cannot be certain, we suggest that many of the respondents in the control areas will have believed themselves to live in areas that were at risk from waterborne illness. This would, in our view increase the risk of recall bias in the control populations. Recall bias is a well-known problem in epidemiology (Greenland, 1996; Hennekens and Buring, 1987) but one that seems to have been ignored in infectious disease epidemiology.

In the context used here, recall bias will mean that people are more likely to say they have suffered from diarrhoea in the past few weeks if they believe there has been an outbreak of diarrhoeal disease in the area. We would further argue that this effect of recall bias would be greater the more media attention that is given to the outbreak. It is our belief that the media coverage of the Milwaukee outbreak was far greater than was experienced during our outbreak and that this could have had a greater impact in Milwaukee than seen in our study. Consequently, the previously estimated size of the Milwaukee outbreak would be an even greater overestimate than our findings would suggest.

Conclusions

We can conclude from these observations that retrospective community-based surveys, like the ones done here and in Milwaukee, are very poor for estimating size of outbreaks of waterborne disease. Such surveys will overestimate the size of the outbreak many-fold, probably by a factor of between 10 and 100. The implications of these findings are, however, much wider. One of the successes of quantitative microbial risk assessment was to "correctly" predict the estimated size of the Milwaukee outbreak (Haas et al., 1999). If the size of the Milwaukee outbreak was substantially less then this, our findings must also call into question the validity of quantitative microbial risk assessment as a tool for estimating the burden of microbial disease in the community.

References


