Modelling the impact of prior immunity, case misclassification and bias on case-control studies in the investigation of outbreaks of cryptosporidiosis

P. R. HUNTER*

Public Health Laboratory, Countess of Chester Health Park, Liverpool Road, Chester CH2 1UL, UK

SUMMARY

Cryptosporidiosis is the most common cause of outbreaks of disease linked to mains water supply in the United Kingdom and the second commonest in the United States. Recent evidence has suggested that prior population immunity may have an impact on the epidemiology of waterborne outbreaks and in particular prior immunity may reduce the power of case-control studies for demonstrating association between disease and water consumption behaviour. However, the degree of impact of prior immunity on the power of epidemiological studies is not yet clear. This paper reports the results of some simple mathematical models of outbreaks of waterborne disease in populations with varying levels of immunity due to prior water and non-water exposure. The basic outbreak model was run on a spreadsheet. To determine the impact of prior immunity on case-control studies, further analysis was done using a Monte Carlo method to simulate sampling from cases and controls. It was found that moderate degrees of prior immunity due to water associated disease could markedly reduce the relative risk of water consumption on illness in waterborne outbreaks. In turn this would reduce the power of case-control studies. In addition, this model was used to demonstrate the impact of case misclassification and recall bias on case-control studies. Again it was found that within the model, the results of case-control studies could be significantly affected by both these sources of error. Anyone conducting epidemiological investigations of potentially waterborne outbreaks of disease should be aware of the epidemiological problems. Mistakes from case-control studies will be minimized if the outbreak team pays considerable attention to the descriptive phase of the investigation and if case-control studies are conducted as soon as possible after an outbreak is detected.

INTRODUCTION

Cryptosporidiosis is now known to be a frequent cause of gastroenteritis worldwide which can give rise to prolonged illness [1, 2]. In the United Kingdom, cryptosporidiosis has become the most commonly reported cause of outbreaks of waterborne disease [2, 3]. In the United States, cryptosporidia is now second only to giardia as a diagnosed pathogen in

waterborne outbreaks [4, 5]. However the largest recorded outbreak of a waterborne disease in the United States was of cryptosporidiosis which was responsible for an estimated 403 000 infections [6]. Clearly, waterborne cryptosporidiosis is the major infectious cause of waterborne disease in the western world.

Despite its importance there remains significant uncertainties about its epidemiology and transmission. One of the areas of outstanding uncertainty

* Author for correspondence.

is how does one determine that an outbreak of cryptosporidiosis is indeed waterborne. In both the United States and United Kingdom national disease surveillance organizations have produced guidance on categorizing levels of evidence in reported waterborne outbreaks [5, 7]. Both systems use a combination of epidemiological evidence and water quality data. In particular, considerable weight is given to the results of any analytical epidemiological investigation undertaken, which would usually mean a case-control study. However, we have recently expressed concern about the reliability of case-control studies for the investigation of waterborne outbreaks of cryptosporidiosis in certain circumstances [8].

One of the situations where case control studies may experience problems is when there is pre-existing immunity in the population. Whilst immunity to cryptosporidiosis is still not completely understood. Recent evidence, from human volunteer studies, suggest that resistance following a single repeat exposure is not complete, though oocyst shedding and severity of diarrhoea is significantly lower in people when challenged for the second time in 12 months [9, 10]. Other support for the hypothesis that infection can lead to immunity comes from the follow-up of an outbreak of cryptosporidiosis in Jackson County, Oregon [11]. During this outbreak it was noted that out-of-town guests at a wedding reception had a higher attack rate than local guests, presumably because of prior exposure. Indeed, in a recent United Kingdom outbreak of waterborne cryptosporidiosis in a river water supply area, a feature of the outbreak was the large number of cases in visitors (i.e. non-immunes) to the area [12].

Other situations where case-control studies may have problems are when there is case misclassification (cases of cryptosporidiosis are included in the analysis when they are not part of the outbreak or are secondary cases) and reporting bias (cases believe that drinking water is the likely source and artificially report that they are more likely to drink unboiled tap water).

In this paper I report the results of some relatively simple mathematical modelling of waterborne outbreaks of cryptosporidium which investigates the power of case-control studies in waterborne outbreaks where varying proportions of the population is immune, where case misclassification occurs and where there is reporting bias of water consumption behaviour. By power, I mean the ability of a case-control study to find a statistically significant as-sociation between drinking water and illness, if the outbreak is in fact due to drinking water.

METHODS

The basic model was run in on a personal computer using the Quatro Pro spreadsheet. A virtual population was categorized according to water-consumption behaviour based on that of the control population of an outbreak that we had previously reported [13]. Essentially 34·5 % of the population drank less than 1 glass of raw tap water each day, 43·1 % 1–3 glasses, 13·8 % 4–6 and 8·6 % 7 or more. These 4 groups were summarized as drinking 0·5, 2, 5 and 8 glasses per day.

Within each group there were susceptibles and immunes. Each cycle, nominally one year, susceptibles converted to immunes according to the following equation:

$$\text{Rate}_{\text{susceptible to immune}} = \text{susceptible} \\ \times \text{daily water consumption} \\ \times \text{water infectivity}.$$

This equation was defined from the assumptions that only susceptible individuals were at risk of infection and that the probability of infection (from drinking water) in a susceptible was directly related to the daily water consumption and a variable 'water infectivity'. This water infectivity variable was incorporated in the model to enable one to change the proportion of the population immune at equilibrium. However, the variable is directly related to the probability of a susceptible individual becoming infected in a year by drinking 1 glass of unboiled tap water per day.

In this model immunity was assumed to last for life and so immunes were replaced by susceptibles by death and replacement by newborns. Thus the population was assumed to be in balance with newborns exactly matching deaths. Therefore, the transfer from immune to susceptible was governed by the equation:

$$\text{Rate}_{\text{immune to susceptible}} = \text{immune}/70.$$

The variable for water infectivity was changed in the model so that various predetermined proportions of the population were created immune when the model had reached equilibrium. This was done by trial and error until the desired proportion was produced by the model.

Once the model reached equilibrium, relative risk of water consumption for water-related and non water-related infection were calculated.

In order to calculate the minimum number of cases required to give a statistically significant result in 95% of outbreaks a simple Monte Carlo model was developed in basic [14]. This was done by taking the proportion of each population group infected and creating n cases and $2 \times n$ controls. Cases were randomly assigned to the four water consumption groups with a probability equal to that of the proportion of each water consumption group amongst cases. Controls were allocated randomly in proportion to the distribution of all individuals in each water consumption group. Once all cases and controls were allocated to water consumption groups χ^2 for trend was calculated. For each value of n, 1000 trials were run and the number of trials giving a χ^2 with $P < 0.05$ calculated. The number of cases were increased by one until a statistically significant result occurred on more than 95% of occasions for two consecutive values of n. The lowest value was then taken as the number of cases.

To demonstrate the impact of reporting bias on case-control studies, I ran the Monte Carlo model as described in the previous paragraph except that cases and controls were allocated to the water-consumption groups with a probability equal to the proportion of the group in the total population. In other words, there was no difference in water-consumption behaviour between cases and controls. However, reported water-consumption behaviour for various proportions of cases was increased by 1–4 glasses to model the impact of reporting bias.

To demonstrate the impact of case-misclassification errors I made 20% of cases have the water-consumption behaviour of the control population, rather than the case population.

RESULTS

The impact of prior immunity

The impact of prior immunity due to infection acquired via the waterborne route on the relative risk associated with water consumption is shown in Figures 1 and 2. It can be seen from Figure 1 that, during a waterborne outbreak, when 20% of the population is immune, as a result of water associated infection, the relative risk associated with the highest water consumption halves from 16 to 8. When 50% of the population is immune then this halves again to only 4. For non water-associated outbreaks, the association becomes increasingly negative as water
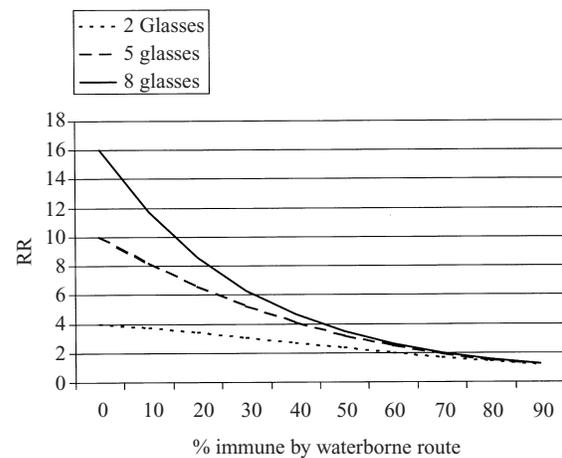

Fig. 1. Relative risk in a model of a waterborne outbreak according to daily water consumption and proportion of population immune due to water-associated disease.
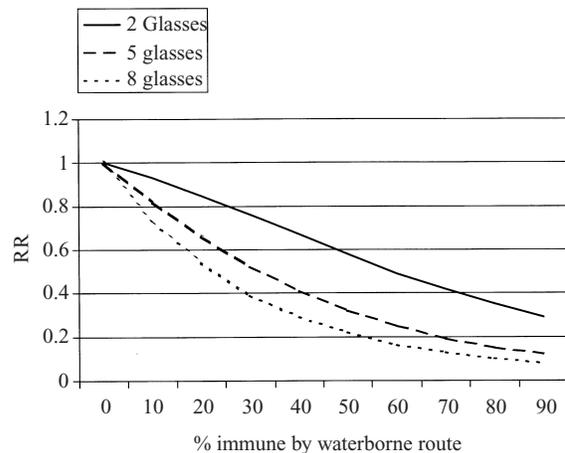

Fig. 2. Relative risk in a model of a non-waterborne outbreak according to daily water consumption and proportion of population immune due to water-associated disease.

associated immunity in the population increases (Fig. 2).

This decline in strength of association due to prior immunity has a very big impact on the cases required to give a case-control study sufficient power. Figure 3 demonstrates this by showing the number of cases required to give a 95% probability of achieving a statistically significant result. Above 20% immunity the number of cases required increases dramatically.

The impact of case misclassification

Case misclassification had a significant impact on the relative risk and on numbers required. Although any degree of misclassification reduced the power of case-
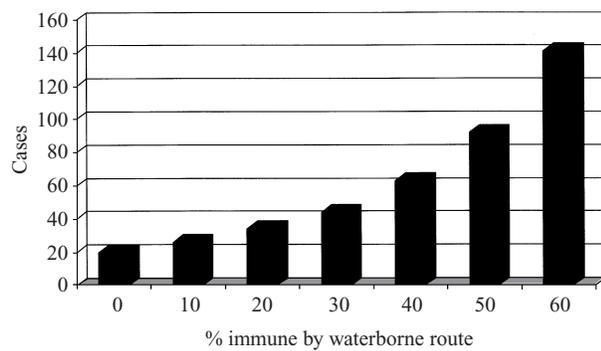
Fig. 3. Number of cases required for a case-control study of water consumption as a risk factor for cryptosporidiosis assuming varying levels of immunity due to previous exposure to cryptosporidium via the water route with and without 20% of cases being erroneously identified as primary.
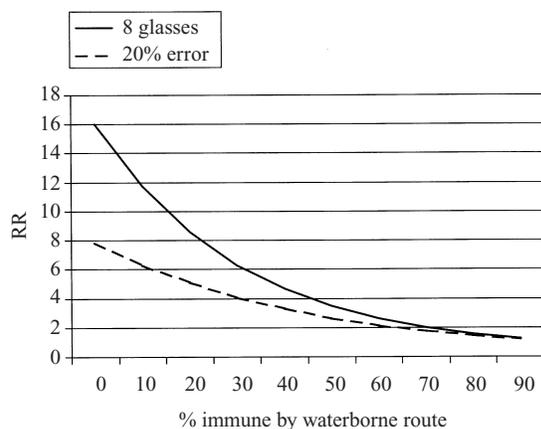
Fig. 4. Impact of 20% of identified cases being misclassification as being part of waterborne outbreak on calculated relative risk for population drinking 8 glasses of water a day compared to those drinking less than 1.

control studies we have chosen to present only the data for a 20% rate. The impact on relative risk is shown in Figure 4. Whatever the proportion of the population immune from prior water exposure, if 20% of cases included in the study are not primary cases, then approximately 50% more cases are required to maintain the power of any case-control study.

The impact of reporting bias

The impact of bias on giving falsely significant results is shown in Table 1. Without bias, only 5% of studies should give significant results. However, even when a minority of cases give biased results, this can have a major impact on the likelihood of falsely reporting a significant study. Indeed, even with 30% of cases reporting that they drink 2 glasses of water more each

day then they do, this would give a falsely significant association between illness and water consumption in more than 50% of studies with 100 cases and 200 controls.

DISCUSSION

Even within affluent countries, with otherwise satisfactory water supplies, some communities have significantly greater exposure, and so immunity, to cryptosporidium, through their water supply, than do others. This can be seen in the reported variation in the results of seroepidemiological surveys which report antibody positivity ranging from about 20 to 60% [9, 15, 16]. Based on our experience of investigating two outbreaks of waterborne cryptosporidiosis, we recently suggested that a major factor in the geographical variation in seroprevalence was source of drinking water [8, 13]. More specifically those populations drinking treated river water were more likely to be exposed to infection then those drinking groundwater [8, 17]. Such repeat exposure in populations taking river water would interfere with the power of case-control studies to demonstrate significant associations. We found evidence for this hypothesis in a review of the literature where we demonstrated that the relative risk of disease was always higher in those living in a groundwater supply area during a waterborne outbreak than in those living in river water supply areas during an outbreak [8].

Subsequently, further evidence to support this hypothesis has come from several sources. Frost reported serological studies following two outbreaks in Jackson County, Oregon [11, 18]. They noted that residents of Talent, a river water supply area, had higher antibody levels soon after the outbreak and also 2 years later. Their suggestion that 'long-term residents of communities such as Talent, who may receive recurrent exposure to oocysts, may eventually experience minimal risk of illness from re-infection' is in line with our explanation. The work of McLauchlin and colleagues showing higher levels of antibody in populations consuming river water also provides further confirmatory evidence [16]. Perz and colleagues devised a model to predict the number of cases of cryptosporidiosis in a population and applied this model to New York [19]. They found that their model predicted many more cases then were seen in practice. Perhaps, this discrepancy is due to high levels of immunity in the New York population.

Table 1. *Proportion of runs giving a significant response when outbreak not water-related assuming varying levels of misclassification and reporting bias for 50 and 100 cases*

No. of extra glasses reported due to bias	Proportion of runs giving a statistically significant response ($P < 0.05$)/%					
	Study with 50 cases and 100 controls			Study with 100 cases and 200 controls		
	10% cases misclassified	20% cases misclassified	30% cases misclassified	10% cases misclassified	20% cases misclassified	30% cases misclassified
1	7·9	10·8	19·5	8·9	17·6	34·2
2	9·3	16·7	29·9	13·9	29·2	58·4
3	10·5	22·5	46·8	15·5	42·5	77·6
4	16·1	48·9	81·6	28·8	79·8	98·3

What then are the implications of this model? Outbreaks of waterborne cryptosporidiosis in river water areas will be more difficult to detect as illness will be less severe and cases will be fewer. Furthermore case-control studies in such areas will also lack power unless the case numbers are large. From the model, at least 60 cases will be needed if 40% of the population is immune through prior water exposure, even assuming all cases included in the study are primary cases. The number of cases will be much higher if secondary cases are included in the analysis, or if a greater proportion of the population is immune by the water route. We do not know exactly what the levels of effective immunity are in the population. It would appear that antibody response does not always occur following a primary or secondary infection [10, 20]. It could well be the case that immunity is much greater than indicated by seroprevalence surveys. By contrast one would expect to see a negative dose-response relationship for many non-waterborne outbreaks.

Further, reducing the power of any case-control study is the problem of including cases even when they have not acquired infection directly from the water supply. These could be secondary cases or they could be totally unrelated to the outbreak under investigation. The problem of secondary cases in a household is usually solved by excluding the second case if illness developed more than a certain number of days after the first. However, secondary cases may still be included if the primary case was asymptomatic. I am not aware of any studies that have adequately estimated the number of secondary cases of cryptosporidiosis that were erroneously included in epidemiological investigations of outbreaks.

Case-control studies, however, can also give false positive (type I errors) when cases' reporting of water consumption is biased. Indeed, as can be seen in this report, bias can have a dramatic impact on the probability of generating a type I error, even when only a minority of cases' reporting is biased. Unfortunately we can not be totally sure about the level of bias in an epidemiological study, though such bias may be more marked than is generally thought and can have significant impact on the results of case-control studies [21–23]. Factors leading to increased recall bias are not fully understood, though a knowledge of the hypothesis under evaluation and potential for financial gain probably play a significant role [24, 25]. Such factors may artificially increase cases' recall of unboiled tap water consumed in an outbreak where the media have already blamed the water supply and individuals are hoping to make compensation claims.

Finally, are case-control studies so flawed that they should not be used to investigate outbreaks of waterborne cryptosporidiosis? Despite their potential problems, case-control studies do have an important role, as long as investigators are aware of their potential problems. Assuming that most populations have some immunity from water exposure then, in a non-waterborne outbreak, one would expect to see a negative dose–response relationship between disease and water consumption. The finding of a negative association would be taken as good evidence that the outbreak was not waterborne. Similarly, if the case-control study was not too restricted in the exposures enquired about, an outbreak due to another factor, such as farm visits or milk consumption, should, when studied, show an association with these other exposures. This approach reinforces the importance of good descriptive epidemiology, undertaken as early in the investigation as possible. This may be even more important for waterborne disease than for other causes of outbreaks.

REFERENCES

1. Meinhardt PL, Casemore DP, Miller KB. Epidemiologic aspects of human cryptosporidiosis and the role of waterborne transmission. Epidemiol Rev 1996; **18**: 118–36.

2. Hunter PR. Waterborne disease – epidemiology and ecology. Chichester: Wiley, 1997.

3. Furtado C, Adak GK, Stuart JM, Wall PG, Evans HS, Casemore DP. Outbreaks of waterborne infectious intestinal disease in England and Wales, 1992–5. Epidemiol Infect 1998; **121**: 109–19.

4. Moore AC, Herwaldt BL, Craun GF, Calderon RL, Highsmith AK, Juranek DD. Surveillance for waterborne disease outbreaks – United States, 1991–1992. MMWR 1993; **42** (SS-5): 1–22.

5. Kramer MH, Herwaldt BL, Craun GF, Calderon RL, Juranek DD. Surveillance for waterborne-disease outbreaks – United States, 1993–1994. MMWR 1996; **45** (SS-1): 1–33.

6. MacKenzie WR, Hoxie NJ, Proctor ME, et al. A massive outbreak in Milwaukee of cryptosporidium infection transmitted through the public water supply. N Engl J Med 1994; **331**: 161–7.

7. Tillett HE, de Louvois J, Wall PG. Surveillance of outbreaks of waterborne infectious disease: categorizing levels of evidence. Epidemiol Infect 1998; **120**: 37–42.

8. Hunter PR, Quigley C. Investigation of an outbreak of cryptosporidiosis associated with treated surface water finds limits to the value of case control studies. Comm Dis Publ Hlth 1998; **1**: 234–8.

9. DuPont HL, Chappell CL, Sterling CR, Okhuysen PC, Rose JB, Jakubowski W. The infectivity of *Cryptosporidium parvum* in health volunteers. N Engl J Med 1995; **332**: 855–9.

10. Okhuysen PC, Chappell CL, Sterling CR, Jakubowski W, DuPont HL. Susceptibility and serologic response to healthy adults to reinfection with *Cryptosporidium parvum*. Infect Immun 1988; **66**: 441–3.

11. Frost FJ, Calderon RL, Muller TB, Curry M, Rodman JS, Moss DM, de la Cruz AA. A two-year follow-up of antibody to *Cryptosporidium* in Jackson County, Oregon following and outbreak of waterborne disease. Epidemiol Infect 1998; **121**: 213–7.

12. Anonymous. An outbreak of cryptosporidiosis in South and West Devon, August to September 1995. South and West Devon Health Authority, 1995.

13. Bridgman SA, Robertson RMP, Syed Q, Speed N, Andrews N, Hunter PR. Outbreak of cryptosporidiosis associated with a disinfected groundwater supply. Epidemiol Infect 1995; **115**: 555–66.

14. Raeside DE. Monte Carlo principles and applications. Physics Med Biol 1976; **21**: 181–97.

15. Kuhls T, Moiser DA, Crawford D, Griffs J. Seroprevalence of cryptosporidial antibodies during infancy, childhood and adolescence. Clin Infect Dis 1994; **5**: 731–5.

16. McLauchlin J, Casemore DP, Moran S, Patel S. The epidemiology of cryptosporidiosis: application of experimental sub-typing and antibody detection systems to the investigation of water-borne outbreaks. Folia Parasitol 1998; **45**: 83–92.

17. Craun GF, Hubbs SA, Frost F, Calderon RL, Via SH. Waterborne outbreaks of cryptosporidiosis. J Am Water Works Ass 1998; **90**: 81–91.

18. Frost FJ, de la Cruz AA, Moss DM, Curry M, Calderon RL. Comparisons of ELISA and Western blot assays for detection of *Cryptosporidium* antibody. Epidemiol Infect 1998; **121**: 205–11.

19. Perz JF, Ennever FK, Le Blancq SM. Cryptosporidium in tap water: comparison of predicted risks with observed levels of disease. Am J Epidemiol 1998; **147**: 289–301.

20. Moss DM, Bennett SN, Arrowodd MJ, Wahlquist SP, Lammie PJ. Enzyme linked immunoelectrotransfer blot analysis of a cryptosporidiosis outbreak on a United States Coast Guard cutter. Am J Trop Med Hyg 1998; **58**: 110–8.

21. Chouinard E, Walter S. Recall bias in case-control studies: an empirical analysis and theoretical framework. J Clin Epidemiol 1995; **48**: 245–54.

22. Neutra RR, Swan SH, Hertz-Picciotto I, et al. Potential sources of bias and confounding in environmental epidemiologic studies of pregnancy outcomes. Epidemiol 1992; **3**: 134–42.

23. Hertz-Picciotto I, Swan SH, Neutra RR. Reporting bias and mode of interview in a study of adverse pregnancy outcomes and water consumption. Epidemiol 1992; **3**: 104–12.

24. Kerkhofs M, Lindeboom M. Subjective health measures and state dependent reporting errors. Health Econ 1995; **4**: 221–35.

25. Werler MM, Shapiro S, Mitchell AA. Periconceptional folic acid exposure and risk of occurrent neural tube defects. JAMA 1993; **269**: 1257–61.