Epidemiol. Infect. (2000), **125**, 35–45. Printed in the United Kingdom

Estimation of the under-reporting rate for the surveillance of *Escherichia coli* O157:H7 cases in Ontario, Canada

P. MICHEL[1]*, J. B. WILSON[1,3], S. WAYNE MARTIN[1], R. C. CLARKE[2], S. A. McEWEN[1] AND C. L. GYLES[4]

[1] *Department of Population Medicine, Ontario Veterinary College (OVC), University of Guelph, Guelph, Canada*
[2] *Guelph Laboratory, Health Canada, Guelph*
[3] *Laboratory Centre for Disease Control (LCDC), Health Canada, Guelph*
[4] *Department of Pathobiology, Ontario Veterinary College (OVC), University of Guelph, Guelph*

(*Accepted 3 February 2000*)

SUMMARY

Two models estimating the proportion of *Escherichia coli* O157:H7 cases not reported in the Ontario notifiable diseases surveillance system are described. The first model is a linear series of adjustments in which the total number of reported cases is corrected by successive under-reporting coefficients. The structure of the second model is based on a relative difference in the proportion of *E. coli* O157:H7 cases which are hospitalized between the surveillance database and the underlying population.

Based on this analysis, the rate of under-reporting of symptomatic cases of *E. coli* O157:H7 infection in Ontario ranges from 78 to 88% corresponding to a ratio of 1 reported case for approximately 4–8 symptomatic cases missed by the surveillance system. This study highlights the need to increase awareness among public health workers of the potential biases that may exist in the interpretation of routine surveillance data.

INTRODUCTION

In any surveillance system, the processes of detection, confirmation and reporting of cases are critical activities which greatly influence our capacity to accurately evaluate the impact of a given disease in a population. The ability of such a system to detect all possible cases occurring in a population is referred to as the sensitivity of the surveillance system. In the case of *E. coli* O157:H7 surveillance, this sensitivity is influenced by many factors, including the likelihood that an infected patient will seek medical attention, the proportion of those patients for which an appropriate laboratory test will be requested, the ability of the testing process to confirm true cases and the subsequent successful relaying of the patient's medical information through the reporting system.

* Author for correspondence.

Estimation of the under-reporting rate requires the collection of information external to the system to determine the most likely true occurrence of the disease in the population of interest. This can be achieved using two different strategies: (*a*) evaluating the disease frequency with an independent epidemiological study including a representative sample of the underlying population and (*b*) estimating correction factors for each element of the surveillance system to adjust observed frequencies for the likely under-reporting bias.

Estimation of a correction factor for each step in the surveillance process was central to a study conducted by Chalker and Blaser in 1988 to evaluate the under-reporting of salmonellosis in the United States [1]. Inspired by a surveillance model described for shigellosis [2], the surveillance artifact model proposed by these authors consisted of the evaluation

of seven steps involved in the process of detection and reporting of patients infected with salmonella. Based on a comprehensive review of the medical literature related to the parameters associated with each step, the true incidence of salmonellosis was estimated using a sequence of correction factors applied to each of these steps. The principal limitation of this approach lies in the amount of uncertainty attached to the value of the correction factors. In their calculation of the incidence of salmonellosis, Chalker and Blaser used a point estimate, the median, to calculate a correction factor for each step and the overall correction factor for the model. Their sequential model was deterministic in nature and no provision was made to calculate measures of variation around the point estimate.

Surveillance for *E. coli* O157:H7 infection in Ontario can also be modelled using linear sequential steps of detection and reporting. At present, there is no published information regarding the magnitude of under-reporting associated with surveillance for *E. coli* O157:H7 in the Province. Lacking such information, public health workers have often assumed the extent of *E. coli* O157:H7 under-reporting to be approximately the same as for salmonellosis or comparable to estimates derived for undifferentiated gastroenteritis [3–6].

The principal objective of the present study was to provide an estimate of the under-reporting rate for *E. coli* O157:H7 infection under the surveillance system in place for the Province of Ontario, Canada. To better appreciate the stochastic nature of variables describing the surveillance process, values for the correcting factors were obtained using simulation models which considered a probability distribution for each coefficient in the models and allowed the calculation of standard errors associated with the under-reporting estimates. In addition to a sequential model, a novel method, based on hospitalization rates, was used to validate the magnitude of the under-reporting estimate.

MATERIALS AND METHODS

Data sources

Information on the 2971 verocytotoxigenic *Escherichia coli* (VTEC) cases reported in Ontario between 1990 and 1995 was extracted from the Reportable Disease Information System database (RDIS) from the Ministry of Health of Ontario. Health services for residents of Ontario are paid for through a publicly funded universal health care system administered by the Province. The mean provincial population over the time period of interest was 10084885. Sporadic cases of VTEC were defined as persons with compatible clinical signs for which one or more of the following criteria applied: verocytotoxin was detected from stool specimens; one or more strains of verocytotoxigenic *Escherichia coli* was isolated from stool or blood. Outbreak related cases ($n = 94$) and VTEC cases other than *E. coli* O157:H7 ($n = 5$) were excluded from the study ($n = 2872$ selected cases).

Systematic literature review and coefficient estimation

Two simulation models, referred to as sequential and hospital models, were used for the estimation of the rate of under-reporting of *E. coli* O157:H7 in Ontario. Estimation of the coefficients entering the models were derived from a comprehensive literature review of outbreak reports and surveillance studies on *E. coli* O157 infection. All North-American and European reports published between 1980 and 1995 and describing the clinical and/or epidemiological characteristics of *E. coli* O157 cases for outbreaks or sporadic cases were considered for inclusion in the study. From these, studies and reports for which case definitions of *E. coli* O157 were ill-defined, for which the total population exposed was not estimated, or for which the underlying population was not defined were excluded. Scientific articles which reported information on under-reporting associated with undifferentiated gastroenteritis were also searched. The literature search was conducted using Medline (National Library of Medicine) and Current-Contents (Institute for Scientific Information, Inc. 1993) and by consulting with recognized experts in the field. The bibliographies of articles identified in this manner were systematically reviewed to identify additional relevant references. Relevant information was extracted from the selected articles and compiled into a spreadsheet program (Lotus 123; Lotus Development Corporation, Georgia).

The appropriate distributions were selected for the coefficients entering the models based on the values reported in the literature as well as on the amount of uncertainty attached to each of these values.

Sequential model

In the sequential model, the series of events necessary for *E. coli* O157:H7 infection to be detected and

https://doi.org/10.1017/S095026889900374X Published online by Cambridge University Press

reported was modelled using six variables. The first variable (α) was an estimate of the proportion of the total number of infected people who are symptomatic. The second variable (β) estimated the proportion of symptomatic cases who seek medical attention. The third variable (γ) estimated the total proportion of symptomatic patients for whom a stool sample is requested for confirmation of *E. coli* O157:H7. The fourth variable (δ) estimated the proportion of these samples which are laboratory-confirmed. The fifth variable (ϵ) estimated the proportion of laboratory-confirmed cases which are reported to the appropriate health authority. The last variable 'n' consisted of the annual average number of *E. coli* O157:H7 cases recorded in the RDIS database over the time period of interest. The proportion of the total population infected with *E. coli* O157:H7 not reported in the surveillance database ($UR_{population}$), and the proportion of the symptomatic cases of *E. coli* O157:H7 not reported in the surveillance database (UR_{sympt}) were calculated using equations (1) and (2) in Appendix 1.

Hospitalization model

Another approach based on the proportion of cases hospitalized was developed to estimate the under-reporting rate of *E. coli* O157:H7 in Ontario. The hospitalization model relied on the assumptions that hospitalized cases recorded by the surveillance database were symptomatic, were seen by a physician and a stool sample was taken for laboratory identification and confirmation. Under these circumstances, the under-estimation of symptomatic cases of *E. coli* O157:H7 is reflected by the unknown fraction of two groups; the non-hospitalized and hospitalized *E. coli* O157:H7 cases. As in the sequential model, the number of reported hospitalized cases was assumed to be dependent on the proportion of the cases which are laboratory confirmed (δ coefficient) and the extent to which laboratory-confirmed *E. coli* O157:H7 cases were reported by laboratories and hospitals to the Ontario Ministry of Health (ϵ coefficient). The under-reporting rate (UR_{sympt}) was calculated by taking the ratio of the hospitalization rate based on the RDIS surveillance database (X) to the hospitalization rate occurring in the general symptomatic population of cases (Y). The estimated general hospitalization rate was derived from a comprehensive literature review of outbreak reports and surveillance studies on *E. coli* O157 infection. The under-reporting rate (UR_{sympt}) based on that model was calculated using equation (3) in Appendix 2.

Analyses

Computations were performed using a risk analysis software (@RISK, Palisade Corporation, New York). Probability distributions for the expected number of people with *E. coli* O157:H7 infection, the expected number of symptomatic cases, under-reporting rates and ratios of reported to unreported cases for both models were generated based on 10000 iterations in a Monte-Carlo re-sampling procedure.

RESULTS

Sequential model

Proportion of E. coli *O157:H7 infected people who are symptomatic* (α)

Prior investigations have demonstrated that *E. coli* O157:H7 can cause a spectrum of illnesses which includes non-bloody diarrhoea, HUS and asymptomatic carriage [7–10]. One estimate of the overall proportion of asymptomatic infections for a given population can be derived from outbreak investigation data. In a small community outbreak in south-east Scotland, six cases were identified on the basis of stool culture, from which one (16·6%) was reported asymptomatic [11]. In 1986, Duncan and collaborators [12] described an outbreak in a kindergarten class in Ontario. They reported 43 out of 62 children had symptoms and 10 out of the remaining 19 were asymptomatic but showed laboratory evidence of infection, giving an approximate asymptomatic proportion of 10/53 (18·8%). In an outbreak of haemorrhagic colitis where 17 persons had confirmed infection, 4 (23·5%) were asymptomatic [13]. More recently in a Minnesota child day-care outbreak, 6 of 38 individuals (15·8%) met the case definition but were reported to have no symptoms [14]. From 20 cases of confirmed *E. coli* O157 detected in a 3-year survey conducted in one laboratory in Brussels, absence of any gastro-intestinal symptoms was reported in two (11·1%) patients [15]. During a laboratory survey in Wales (1900–3), 147 cases of *E. coli* O157 were detected, from which 18% were reported to be asymptomatic [16]. Considering the above information, a triangular distribution with a minimum value of 11%, a most likely value of 18% and a maximum value of 24% was chosen for the

https://doi.org/10.1017/S095026889900374X Published online by Cambridge University Press

Table 1. *Proportion of* E. coli *O157 cases with bloody diarrhoea in selected studies of* E. coli *O157 infection*

		Bloody diarrhoea	
Outbreaks	Total	No. (%)	Ref.
1982, Ontario	31	20 (64·52)	[10]
1984, Nebraska	34	19 (55·88)	[27]
1984 North Carolina	36	11 (30·56)	[28]
1985, Ontario	18	5 (27·78)	[29]
1985, Ontario	55	41 (74·55)	[29]
1988, Minnesota	38	22 (57·89)	[14]
1988, Minnesota	54	31 (57·41)	[30]
1990, Missouri	243	86 (35·39)	[17]
1992, Germany	39	11 (28·21)	[31]
1994, Virginia	20	7 (35·00)	[32]

proportion of asymptomatic people infected with *E. coli* O157:H7 (α'). The proportion of *E. coli* O157:H7 infected people who are symptomatic (α) was calculated as $1-\alpha'$.

Symptomatic cases self-reporting to a physician (β)

Information regarding *E. coli* O157:H7 patients seeking medical attention is very limited. To our knowledge, only the study by Swerdlow and colleagues [17] describing a waterborne outbreak of *E. coli* O157:H7 in Missouri contained an explicit assessment of the proportion of affected patients who sought medical attention. In this article, the authors reported that 40 out of 55 *E. coli* O157:H7 patients with bloody stool (73%) reported to the medical authorities, and 11 out of 50 with non-bloody stools (22%) responded the same way (overall proportion 49%). In an analysis of the costs associated with *E. coli* O157:H7 infection by Marks and Roberts [18], the authors assumed that about half of all cases seek medical attention. Their choice of this coefficient (higher than for salmonella infection) was supported by the increased likelihood of patients to see a doctor in the presence of bloody diarrhoea which occurred in approximately 45% of all *E. coli* O157:H7 cases (Table 1).

Studies containing self-reporting estimates in cases of unspecified enteric illness or food borne disease are also available. In Great Britain, the results of a survey by the Ministry of Agriculture, Fisheries and Food (MAFF) revealed that only 17% of respondents who experienced suspected food poisoning reported the incident to medical authorities [19]. Following an episode of water contamination with sewage in Ireland, only 22% of 340 cases with clinical signs of enteric illness (diarrhoea, vomiting or abdominal cramps) visited their general practitioner [5]. In another study, approximately 20% of patients with gastrointestinal diseases had consulted a general practitioner, independent of the degree of severity of the symptoms [4]. Considering this information, a triangular distribution with a lower limit of 17%, an upper limit of 73% and a most likely value of 50% was chosen for this coefficient (β).

Proportion of symptomatic cases for whom stool samples are obtained (γ)

We found no report in the literature presenting direct information regarding the proportion of symptomatic cases of *E. coli* O157:H7 infection for which a stool sample was requested by a medical authority. In their report concerning the sporadic occurrence of haemorrhagic colitis associated with *E. coli* O157:H7 in Newfoundland, Ratnam and March [20] wrote 'All seven patients whose specimens were positive for *E. coli* O157:H7 had clinical manifestations typical of haemorrhagic colitis, but the syndrome was clinically suspected and a specific test requested in only two cases.' In this case, the relevant estimate would be equal to 29% (2/7).

Similar estimates have been derived for other enteric illnesses. In their review, Chalker and Blaser [1] gave an overall estimate of 41·66% (median) calculated from five salmonella and shigella investigations in which 24, 32, 42, 66, and 86% of patients had stools requested for examination. In Great Britain a survey of people reporting to a physician with diarrhoea showed only 5·4% as having stool requested for

https://doi.org/10.1017/S095026889900374X Published online by Cambridge University Press

Table 2. *Sensitivity of isolation and confirmation procedures in selected studies of* E. coli *O157 infection*

Outbreak	Total	Percent confirmed	Ref.
1982, Michigan	21	42·86	[33]
1982, Ontario	31	58·06	[10]
1982, Oregon	26	50·00	[33]
1984, Nebraska	34	17·65	[27]
1984, North Carolina	36	26·67	[28]
1985, Ontario	73	42·86	[29]
1985, UK	89	35·96	[34]
1986, Ontario	42	67·44	[12]
1986, Washington	37	37·84	[35]
1987, Alberta	17	58·82	[36]
1987, Ontario	15	60·00	[36]
1987, UK	26	56·52	[13]
1988, Minnesota	54	53·57	[30]
1988, Minnesota	38	72·22	[14]
1988, Ontario	25	56·00	[37]
1988, Alberta	63	61·90	[37]
1988, UK	49	65·63	[38]
1990, Michigan	243	58·33	[17]
1991, Massachusetts	23	17·39	[39]
1992, Germany	39	20·51	[31]
1994, Virginia	20	77·78	[32]
1995, Georgia	10	87·50	[40]
1995, Illinois	12	66·67	[41]

examination [21]. Considering the above information, a triangular distribution with a lower limit of 5% an upper limit of 85% and a most likely value of 42% was chosen for this parameter.

Sensitivity of the laboratory procedures (δ)

In 1988, Kleanthous and colleagues published a study evaluating the use of sorbitol MacConkey agar (SMAC) in conjunction with serotyping when compared to a DNA probe specific for the genes encoding for verocytotoxins 1 and 2 [22]. The results of this study indicated a sensitivity of 62 and 56% for the identification of *E. coli* O157 cases using SMAC and serotyping in patients with bloody diarrhoea and non-bloody diarrhoea, respectively. The specificity of the routine test procedure using SMAC and serotyping was 100% for both groups. However, similar data (sensitivity and specificity) concerning the use of SMAC and serotyping for groups of cases with all levels of disease severity, including ones with no symptoms, are not currently available. Table 2 summarizes the proportion of symptomatic cases of *E. coli* O157 infection from whom stool samples were obtained and laboratory-confirmed in 21 selected outbreak investigations and one surveillance study. Since the majority of licensed diagnostic laboratories in Ontario routinely screen stool samples for *E. coli* O157:H7 using Sorbitol MacConkey agar, and since there is a high level of awareness regarding this condition in the Province, we assumed that all submitted samples are tested. Considering the above information, a normal distribution with a mean of 51·8% and a standard deviation of 10·4% was chosen for this coefficient.

Success in reporting to the Ontario Ministry of Health (ε)

Failure to transmit information on confirmed cases to the Ontario Ministry of Health is estimated to be less than 1% in the present surveillance system (H. Lior, National Laboratory for Enteric Pathogen, personal communication). Considering this information, a uniform distribution with a lower limit of 99% and an upper limit of 100% was chosen for this parameter.

Average number of E. coli *O157:H7 cases recorded in RDIS during one year* (*n*)

The average number of *E. coli* O157:H7 cases reported yearly by the surveillance system was estimated by the mean annual number of cases

https://doi.org/10.1017/S095026889900374X Published online by Cambridge University Press

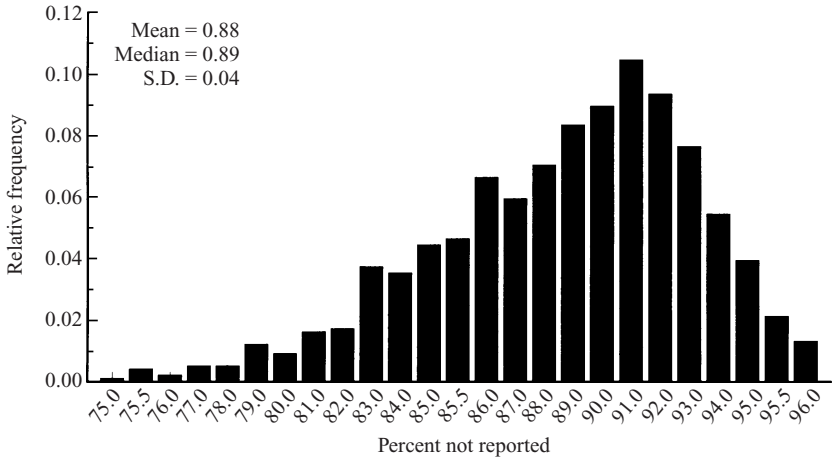

Fig. 1. Predicted proportion of symptomatic *E. coli* O157:H7 cases which are not reported based on the sequential model, Ontario (1990–5).

Table 3. *Hospitalization rates in selected studies of* E. coli *O157 infection*

Outbreak	Number hospitalized	Hospitalization rate (%)	Ref.
1982, Michigan	14	66·7	[33]
1982, Ontario	4	12·9	[10]
1982, Oregon	19	73·1	[33]
1984, Nebraska	14	41·2	[27]
1984, North Carolina	3	8·3	[28]
1986, Ontario	3	7·1	[12]
1986, Washington	17	45·9	[35]
1987, Ontario	4	26·7	[36]
1987, UK	6	23·1	[13]
1987, Utah	8	15·7	[45]
1988, Minnesota	4	7·4	[30]
1988, UK	19	38·8	[38]
1988, WI	2	3·3	[46]
1990, Michigan	32	13·2	[17]
1990, North Dakota	16	24·6	[46]
1991, Massachusetts	6	26·1	[39]
1993, California	14	41·2	[43]
1993, Idaho	4	28·6	[43]
1993, Nebraska	9	15·5	[43]
1993, Washington	144	30·2	[43]
1994, Virginia	3	15·0	[32]
1994, Washington	3	15·0	[44]
1995, Illinois	3	25·0	[41]

recorded in the RDIS surveillance database between 1990 and 1995. A normal distribution with a mean of 478·7 (years per 100000 population) and a standard deviation of 71·8 was calculated for this variable.

Simulation results for the sequential model

The distribution of the expected total number of people with *E. coli* O157:H7 infection for Ontario was slightly skewed to the right with a mean of 5791 (53·9 per 100000 population) cases per year. Under this model, the expected number of people with *E. coli* O157:H7 infection in Ontario would range from 1296 to more than 33564 per year. The distribution of the proportion of all people with *E. coli* O157:H7 infection who are unreported had a mean of 90·0% while the mean of the distribution of the ratio of reported to unreported cases was approximately 1 to

https://doi.org/10.1017/S095026889900374X Published online by Cambridge University Press

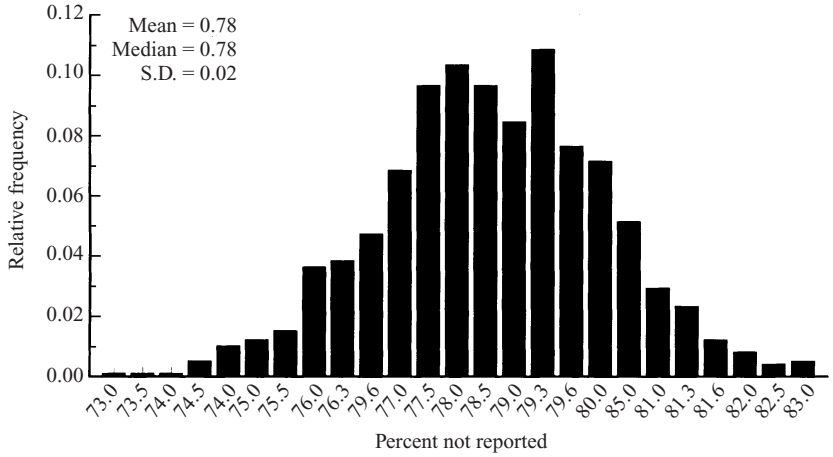

Fig. 2. Predicted proportion of symptomatic *E. coli* O157:H7 cases which are not reported based upon the hospital model, Ontario (1990–5).

11. The distribution of the total number of symptomatic *E. coli* O157:H7 cases had a similar shape to that for the total number of *E. coli* O157:H7 infected people. The annual mean number of symptomatic cases expected under the model was 4764 (44·3 per 100000 population). The mean of the distribution of the proportion of symptomatic cases not reported was 87·9% and that of the distribution of the ratio of reported to unreported cases was approximately 1 to 2 (Fig. 1).

Hospitalization model

Hospitalization rate

The annual hospitalization rate for *E. coli* O157:H7 cases reported in RDIS (X) was assumed to be normally distributed with a mean annual rate of 53·9% (S.D. = 1·6%) which corresponded to the overall proportion of *E. coli* O157:H7 cases that necessitated hospitalization between 1990 and 1995. Values obtained from the literature were used to estimate the overall population hospitalization rate for *E. coli* O157:H7 cases. An overall hospitalization rate (Y) of 24·2% (S.D. = 1·1%) was calculated from the hospitalization rates of 23 outbreak investigations from 1982 to 1995 (Table 3). Based on this information, a normal distribution with a mean of 24% and a standard deviation of 1% was chosen for this parameter.

Simulation results for the hospital model

Under the input conditions specified for the coefficients and their distributions, the resulting distribution of expected total *E. coli* O157:H7 cases for Ontario was approximately normal with a mean of 2201 (20·5 per 100000 population) per year. The proportion of symptomatic *E. coli* O157:H7 cases which are not reported was also approximately normally distributed with a mean of 78% (Fig. 2). The corresponding mean of the distribution of the ratio of reported to unreported cases was approximately 1 to 4.

DISCUSSION

Despite the two distinctive methodological approaches used, the results of the present study suggest comparable estimates of the proportion of *E. coli* O157:H7 cases which are not reported in the Ontario surveillance system. According to our analysis, under-reporting of symptomatic cases of *E. coli* O157:H7 infections in Ontario ranges between 78 and 88%. This means that, on average, for each *E. coli* O157:H7 case reported to the Ontario Ministry of Health, approximately 4–8 other symptomatic cases are missed by the surveillance system. We also estimated that 90% of the total number of people infected with *E. coli* O157:H7, including those who were asymptomatic, were not reported in the provincial surveillance database. These estimates were lower than the 95–99% under-reporting rates previously estimated for salmonella infection [1], or for gastroenteritis in general [3, 21]. This lower percentage of unreported *E. coli* O157:H7 infection can be explained in part by the relatively high proportion (45%) of *E. coli* O157:H7 infections associated with bloody diarrhoea, which influences the level of reporting in two ways. First, the

https://doi.org/10.1017/S095026889900374X Published online by Cambridge University Press

presence of blood in the stool may be a major reason for individuals to report to a physician [9]. This was shown by the study of Swerdlow and colleagues (17) in which the proportion of *E. coli* O157:H7 patients with bloody diarrhoea reporting to medical authorities was more than three times greater than for patients with non-bloody diarrhoea. Secondly, bloody stool is one of the principal reasons for physicians to collect stool samples and submit them for microbiological analysis [6]. The occurrence of asymptomatic infection is likely to be influenced, in part, by the level of immunity in the population under study. For this, and other reasons, our estimates of the under-reporting of asymptomatic *E. coli* O157:H7 infection may not be directly applicable to populations outside Ontario.

The main limitation associated with the sequential reporting model relates to the uncertainty around the coefficients used. Estimates of the proportion of symptomatic *E. coli* O157:H7 patients self-reporting to medical authorities and the proportion of those cases for which a stool sample was requested were particularly problematic in this regard. Similarly, differences in case definitions between reported outbreaks introduces variability into the estimates produced by the simulation models. A value of the Monte Carlo simulation approach is that it allows for incorporation of uncertainty around the input variables into the modelling process. It is important to note that the outputs of the models (e.g. under-reporting rates) are also uncertain and are thus expressed as probability distributions, not point estimates (Figs 1 and 2).

Another factor found to have considerable influence on the proportion of *E. coli* O157:H7 cases reported was the overall sensitivity of the laboratory procedures to identify and confirm true *E. coli* O157:H7 cases. In spite of considerable improvement in the laboratory methods for detection of *E. coli* O157:H7 and other VTEC since the early 1980s, approximately 50% of suspected cases could not be confirmed either by isolation of the organism or the detection of verocytotoxin activity in the stool specimen submitted. The brief period of time that *E. coli* O157:H7 organisms are shed as well as the low number of organisms excreted have been suggested as reasons to explain the difficulty in recovering and identifying *E. coli* O157:H7 from stool specimens [8, 9, 23]. Hence, further development and utilization of sensitive, rapid and accurate tests for the detection and confirmation of *E. coli* O157:H7 infection combined with early testing of suspicious cases would appear to be important elements in improving the discovery/reporting rate for *E. coli* O157:H7 cases.

Few important assumptions influence the validity of the hospitalization-based method for estimating the overall under-reporting rate of *E. coli* O157:H7 infection. The first relied on an accurate estimation of the proportion of *E. coli* O157:H7 cases which were hospitalized. We derived this estimate from the RDIS database; however, the descriptive analysis of this database revealed a high percentage of missing values associated with the hospitalization status (missing values = 59%) [24]. In the present study, estimation of the hospitalization rate from the RDIS surveillance data was calculated under the assumption that the likelihood of observing a missing value in the hospitalization field was independent of the true underlying hospitalization status. Nonetheless, the level of hospitalization calculated from the Ontario surveillance data under this assumption coincides with the estimated proportion of *E. coli* O157:H7 cases hospitalized from other surveillance and hospital-based studies thus supporting the validity of this assumption [9, 25, 26]. Secondly, differential estimation of the proportion of cases hospitalized from the RDIS data and the general *E. coli* O157:H7 population is the central concept of the hospitalization model. The proportion of patients hospitalized in the general *E. coli* O157:H7 population was assumed to be equal to the same parameter estimated from outbreak investigations in which, in most cases, the total number of symptomatic people can be evaluated with reasonable accuracy. Patients suffering from milder enteric symptoms and those with more severe clinical signs were all considered in the estimation of the overall hospitalization rate calculated from outbreak situations. In this context, the average of 24% of outbreak-origin *E. coli* O157:H7 cases admitted to the hospital should reflect closely the true proportion of patients hospitalized in the general population of cases of *E. coli* O157:H7 infection. Finally, this estimate is derived under the assumption that all hospitalized *E. coli* O157:H7 cases had symptoms of gasteroenteritis which prompted the collection of a stool specimen for laboratory analysis. This assumption would result in the exclusion of hospitalized *E. coli* O157:H7 cases having other diagnoses (e.g. stroke, tromboembolic purpura), and for which no such specimen was collected. The term 'symptomatic' as it relates to this under-reporting estimate should be interpreted accordingly.

Under-reporting reflects a practical limitation of a surveillance system to detect and report all disease events occurring in the target population. Nonetheless, a surveillance system that has less than perfect sensitivity is still useful in evaluating patient, temporal and geographical characteristics of a disease, provided that its sensitivity remains reasonably constant over time. Factors such as an increased awareness of the disease, the use of new diagnostic procedures, or the modification of surveillance methods can change the sensitivity of the system and must be considered when investigating long-term trends in disease occurrence.

Routine surveillance for notifiable diseases can be invaluable in the detection of disease outbreaks and sporadic cases, the identification of high risk populations and trends in disease incidence, and in the identification of disease risk factors. However, modelling exercises such as this illustrate the practical limitations that under-reporting of cases imposes on the usefulness and interpretation of surveillance data. These studies not only provide estimates of the extent of under-reporting, but also can provide insight into the mechanisms by which under-reporting can occur. To enhance the validity of interpretation of surveillance data, public health workers should be made aware of the potential biases associated with these data, as well as their likely magnitude and direction.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the participation Dr C. Leber and Dr J. Carlson of the Ontario Ministry of Health and the Medical Officers of Health of the regional health units of Ontario for their collaboration in data collection, Emma Hamilton for reviewing the manuscript and the Laboratory Centre for Disease Control (LCDC), Health Canada and the Fond Canadien d'Aide á la Recherche (FCAR) for financial support.

Appendix 1. *Calculation of the rate of under-reporting in the sequential model*

N total number of *E. coli* O157:H7 infected people expected in the population
N' total number of symptomatic cases of *E. coli* O157:H7 infection expected in the population
n average yearly number of symptomatic and asymptomatic cases of *E. coli* O157:H7 infection reported in the Reportable Disease Information System (RDIS)
n' average yearly number of symptomatic cases of *E. coli* O157:H7 infection reported in RDIS
α proportion of *E. coli* O157:H7 infected people who are symptomatic
β proportion of symptomatic cases which self-report to medical authorities
γ proportion of cases seen by physician which are asked for stool sample
δ proportion of cases with stool sample which are confirmed by laboratory procedures
ϵ proportion of laboratory-confirmed cases which are transmitted to the Ontario Ministry of Health

Let $N = \dfrac{n}{\alpha*\beta*\gamma*\delta*\epsilon}$ and $N' = \dfrac{n'}{\beta*\gamma*\delta*\epsilon}$

and the proportion of the total number of people infected with *E. coli* O157:H7 which are reported in RDIS is estimated as: $P = \{n/N'\}$.

The proportion of the total number of symptomatic *E. coli* O157:H7 cases which are reported in RDIS is estimated as: $P' = \{n'/N'\}$

From which

$$\text{Under-reporting rate} = \text{UR}_{\text{population}} = \left(1 - \frac{n}{N}\right) = 1 - (\alpha*\beta*\gamma*\delta*\epsilon). \quad (1)$$

$$\text{Under-reporting rate} = \text{UR}_{\text{sympt}} = \left(1 - \frac{n'}{N'}\right) = 1 - (\beta*\gamma*\delta*\epsilon). \quad (2)$$

https://doi.org/10.1017/S095026889900374X Published online by Cambridge University Press

Appendix 2. *Calculation of the rate of under-reporting in the hospitalization-based model*

N' number of symptomatic cases of *E. coli* O157:H7 infection in the Ontario population
a number of hospitalized cases of *E. coli* O157:H7 in the Ontario population
b number of hospitalized cases in Reportable Disease Information System (RDIS) database
n' average yearly number of symptomatic cases of *E. coli* O157:H7 infection reported in RDIS
δ proportion of cases with stool sample which are confirmed by laboratory procedures
ϵ proportion of those laboratory-confirmed cases which are transmitted to the Ontario Ministry of Health
X hospitalization rate in RDIS database $= b/n'$
Y hospitalization rate estimated for the population $= a/N'$.

Then the proportion of the total number of symptomatic *E. coli* O157:H7 cases reported in RDIS can be estimated as: $P = n'/N'$

$$\text{since } P = \frac{n'/b}{N'/b} \quad \text{and} \quad b = a*\delta*\epsilon, \quad \text{then} \quad P = \frac{n'/b}{N'/(a*\delta*\epsilon)} = \frac{1/X}{(1/Y)*(1/(\delta*\epsilon))} = \left(\frac{Y}{X}\right)*\delta*\epsilon.$$

From which

$$\text{Under-reporting}_{\text{sympt}} = 1 - P = 1 - \left(\frac{Y}{X}*\delta*\epsilon\right). \tag{3}$$

REFERENCES

1. Chalker RB, Blaser MJ. A review of human salmonellosis: III. Magnitude of *Salmonella* infection in the United States. Rev Infect Dis 1988; **10**: 111–24.
2. Rosenberg ML, Gangarosa EJ, Pollard RA, Wallace M, Brolnitsky O. *Shigella* surveillance in the United States, 1975. J Infect Dis 1977; **136**: 458–9.
3. Notermans S, Hoogenboom-Verdegaal A. Existing and emerging food borne diseases. Inter J Food Microbiol 1992; **25**: 197–205.
4. Hoogenboom-Verdegaal A, DeJong J. Community-based study of the incidence of gastrointestinal diseases in the Netherlands. Epidemiol Infect 1994; **112**: 481–7.
5. Fogarty J, Thornton L, Hayes C, et al. Illness in a community associated with an episode of water contamination with sewage. Epidemiol Infect 1995; **114**: 289–95.
6. Sarfati D, Bates M, Garrett N, Baker M. Survey of gastroenteritis diagnostic practices of New Zealand general practitioners. Ministry of Health. Epidemiology group. ESR: Communicable Disease Centre, 1996.
7. Karmali M. Infection by verocytotoxin producing *Escherichia coli*. Clin Microbiol Rev 1989; **2**: 15–38.
8. Pai CH, Gordon R, Simms HV, Byan LE. Sporadic cases of haemorrhagic colitis associated with *Escherichia coli* O157:H7; clinical, epidemiologic, and bacteriologic features. Ann Intern Med 1984; **101**: 738–42.
9. Pai CH, Ahmed N, Lior H, Johnson WM, Sims HV, Woods DE. Epidemiology of sporadic diarrhoea due to verocytotoxin-producing *Escherichia coli*: a two-year prospective study. J Infect Dis 1988; **157**: 1054–7.
10. Stewart P, Desromeaux W, Chene J. Hemorrhagic colitis in a home for the aged – Ontario. Can Dis Wkly Rep 1983; **9**: 29–32.
11. Brewster D, Brown M, Robertson D. An outbreak of *E. coli* O157 associated with a children's paddling pool. Epidemiol Infect 1994; **112**: 444–7.
12. Duncan L, Mai V, Carter A, Carlson JAK, Borczyk A, Karmali MA. Outbreak of gastrointestinal disease – Ontario. Can Dis Wkly Rep 1987; **13**: 5–8.
13. Salmon R, Farrell I, Hutchison J, et al. A christening party outbreak of haemorrhagic colitis and haemolytic uraemic syndrome associated with *Escherichia coli* O157:H7. Epidemiol Infect 1989; **103**: 249–54.
14. Belongia E, Osterholm M, Soler J. Transmission of *E. coli* O157:H7 infection in Minnesota child day-care facilities. JAMA 1993; **269**: 883–8.
15. Pierard D, Stevens D, Moriau L, Lior H, Lauwers S. Three years PCR screening for *E. coli* O157 in human stools in Brussels. In: Karmali MA, Goglio AG, eds. Recent advances in verocytotoxin-producing *Escherichia coli* infections: 2nd International Conference on *E. coli* O157, Bergamo, Italy, 1994. Elsevier Science BV, 1994.
16. Salmon RL, Smith RMM. How common is *Escherichia coli* O157 and where is it coming from? Total population surveillance in Wales 1990–1993. In: Karmali MA, Goglio AG, eds. Recent advances in verocytotoxin-producing *Escherichia coli* infections; 2nd International Conference on *E. coli* O157, Bergamo, Italy, 1994. Elsevier Science BV, 1994.
17. Swerdlow DL, Bradley AW, Brady RC, et al. A waterborne outbreak in Missouri of *Escherichia coli* O157:H7 associated with bloody diarrhea and death. Ann Inter Med 1992; **117**: 812–9.
18. Marks S, Roberts T. *E. coli* O157:H7 ranks as the fourth most costly food borne disease. Food Rev 1993; **16**: 51–9.
19. Ministry of Agriculture Fisheries and Food. Food Hygiene; Report on a consumer survey. London: Her Majesty's Stationery Office, 1988.
20. Ratnam S, March S. Sporadic occurrence of hemorrhagic colitis associated with *E. coli* O157:H7 in Newfoundland. Can Med Assoc J 1986; **134**: 43–5.

https://doi.org/10.1017/S095026889900374X Published online by Cambridge University Press

21. Feldman R, Banatvala N. The frequency of culturing stools from adults with diarrhoea in Great Britain. Epidemiol Infect 1994; **113**: 41–4.
22. Kleanthous H, Fry N, Smith H, Gross R, Rowe B. The use of sorbitol MacConkey agar in conjunction with a specific antiserum for the detection of Vero cytotoxin producing strains of *Escherichia coli* O157. Epidemiol Infect 1988; **101**: 327–35.
23. Karmali MA, Petric M, Lim C, Freming PC, Arbus GS, Lior H. The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing *Escherichia coli*. J Infect Dis 1985; **151**: 775–82.
24. Michel P, Wilson JB, Martin SW, Clarke RC, McEwen SA, Gyles CL. A descriptive study of verocytotoxigenic *Escherichia coli* (VTEC) cases reported in Ontario 1990–1994. Can J Publ Hlth 1997; **89**: 253–7.
25. Waters J, Sharp J, Dev V. Infection caused by *Escherichia coli* O157:H7 in Alberta, Canada, and in Scotland: A five-year review, 1987–1991. Clin Infect Dis 1994; **19**: 34–43.
26. Ostroff SM, Kobayashi JM, Lewis J. Infections with *Escherichia coli* O157:H7 in Washington state. JAMA 1989; **262**: 355–9.
27. Ryan CA, Tauxe RV, Hosek GW, et al. *Escherichia coli* O157:H7 diarrhea in a nursing home: clinical, epidemiological, and pathological findings. J Infect Dis 1986; **154**: 631–8.
28. Spika JS, Parsons JE, Nordenberg D, Wells JG, Gunn RA, Blake PA. Hemolytic uremic syndrome and diarrhea associated with *Escherichia coli* O157:H7 in a day care centre. J Pediatr 1986; **109**: 287–91.
29. Carter A, Borczyk A, Carlson J, et al. A severe outbreak of *Escherichia coli* O157:H7 associated hemorrhagic colitis in a nursing home. N Engl J Med 1987; **317**: 1496–500.
30. Belongia E, MacDonald K. An outbreak of *E. coli* O157:H7 colitis associated with consumption of precooked meat patties. J Infect Dis 1991; **164**: 338–43.
31. Reida P, Wolff M, Pohls H, et al. An outbreak due to enterohaemorrhagic *Escherichia coli* O157:H7 in a children day care centre characterized by person-to-person transmission and environmental contamination. Zbl Bakt 1994; **281**: 534–43.
32. Frost B, Chaos C, Ladaga L et al. *Escherichia coli* O157:H7 outbreak at a summer camp – Virginia, 1994. MMWR 1995; **44**: 419–21.
33. Riley L, Remis R, Helgerson S, et al. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N Engl J Med 1983; **308**: 681–5.
34. Smith HR, Rowe B, Gross RJ, Fry NK, Scotland S. M. Haemorrhagic colitis and vero-cytotoxin-producing *Escherichia coli* in England and Wales. Lancet 1987; **i**: 1062–5.
35. Ostroff S, Griffin P, Tauxe R, et al. A statewide outbreak of *Escherichia coli* O157:H7 infections in Washington State. Am J Epidemiol. 1990; **132**: 239–47.
36. Hockin J, Lior H. Hemorrhagic colitis and hemolytic uremic syndrome caused by *Escherichia coli* O157:H7 in Canada. Can Dis Weekly Rep 1987; **13**: 203–4.
37. Hockin J, Lior H, Stratton F, Ratnam S, April N, Remis R. Hemorrhagic colitis due to *Escherichia coli* (verotoxigenic) in Canada. Can Dis Weekly Rep 1988; **14**: 147–8.
38. Morgan GM, Newman C, Palmer SR, et al. First recognized community outbreak of haemorrhagic colitis due to verotoxin-producing *Escherichia coli* O157:H7 in the UK. Epidemiol Infect 1988; **101**: 83–91.
39. Besser R, Lett S, Weer JT. An outbreak of diarrhoea and hemolytic uremic syndrome for *E. coli* O157:H7 in fresh-pressed apple cider. JAMA 1993; **269**: 2217–20.
40. CDC. Outbreak of *Escherichia coli* O157:H7 infection – Georgia and Tennessee, June 1995. MMWR 1996; **45**: 249–50.
41. Warrner M, Kuo K, Wiliams L. Adam Beal. Lake-associated outbreak of *Escherichia coli* O157:H7 – Illinois, 1995. MMWR 1996; **45**: 437–9.
42. Griffin PM, Tauxe RV. The epidemiology of infection caused by *E. coli* O157:H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. Epidemiol Rev 1991; **13**: 60–97.
43. CDC. Update: multistate outbreak of *E. coli* O157:H7 infections from hamburgers – Western United States, 1992–1993. MMWR 1993; **42**: 258–63.
44. Alexander ER, Boase J, Davis M, et al. *Escherichia coli* O157:H7 outbreak linked to commercially distributed dry-cured salami – Washington and California, 1994. JAMA 1995; **273**: 985–6.
45. Pavia AT, Nichols CR, Green DP, et al. Hemolytic-uremic syndrome during an outbreak of *Escherichia coli* O157:H7 infections in institutions for mentally retarded persons: clinical and epidemiologic observations. J Pediatr 1990; **116**: 544–51.
46. CDC. Food borne outbreak of gastroenteritis caused by *Escherichia coli* O157:H7 North Dakota. MMWR 1991; **40**: 265–7.