Estimating the Under-reporting Rate for Infectious Gastrointestinal Illness in Ontario

Shannon E. Majowicz, PhD^{1,2}

Victoria L. Edge, MSc^{1,2}

Aamir Fazil, MSc³

W. Bruce McNab, DVM, PhD⁴

Kathryn A. Doré, MHS^{1,2}

Paul N. Sockett, PhD^{1,2}

James A. Flint, MPH²

Dean Middleton, DVM, MSc⁵

Scott A. McEwen, DVM, DVSc¹

Jeffery B. Wilson, DVM, DVSc, PhD^{1,2}

ABSTRACT

Background: In Ontario, infectious gastrointestinal illness (IGI) reporting can be represented by a linear model of several sequential steps required for a case to be captured in the provincial reportable disease surveillance system. Since reportable enteric data are known to represent only a small fraction of the total IGI in the community, the objective of this study was to estimate the under-reporting rate for IGI in Ontario.

Methods: A distribution of plausible values for the under-reporting rate was estimated by specifying input distributions for the proportions reported at each step in the reporting chain, and multiplying these distributions together using simulation methods. Input distributions (type of distribution and parameters) for the proportion of cases reported at each step of the reporting chain were determined using data from the Public Health Agency of Canada's National Studies on Acute Gastrointestinal Illness (NSAGI) initiative.

Results: For each case of enteric illness reported to the province of Ontario, the estimated number of cases of IGI in the community ranged from 105 to 1,389, with a median of 285, and a mean and standard deviation of 313 and 128, respectively.

Conclusions: Each case of enteric illness reported to the province of Ontario represents an estimated several hundred cases of IGI in the community. Thus, reportable disease data should be used with caution when estimating the burden of such illness. Program planners and public health personnel may want to consider this fact when developing population-based interventions.

MeSH terms: Gastroenteritis; infectious disease reporting; surveillance; Ontario; Canada

La traduction du résumé se trouve à la fin de l'article.

1. Department of Population Medicine, University of Guelph, Guelph, ON

2. Foodborne, Waterborne and Zoonotic Infections Division, Public Health Agency of Canada, Guelph and Ottawa, ON

3. Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, Guelph

4. Innovation and Risk Management Branch, Ontario Ministry of Agriculture and Food, Guelph

5. Ontario Ministry of Health and Long-Term Care, Toronto, ON

Correspondence and reprint requests: Dr. Shannon Majowicz, Public Health Agency of Canada, 160 Research Lane, Suite 206, Guelph, ON N1G 5B2, Tel: 519-826-2201, Fax: 519-826-2244, E-mail: shannon_majowicz@phac-aspc.gc.ca

Acknowledgements: The authors thank the residents and physicians of the City of Hamilton, and the laboratories and local public health units of Ontario for their participation in the studies that enabled this analysis. Thanks also to Cara Bowman for her critical evaluation of the final drafts of the report. This project was funded by the Public Health Agency of Canada (formerly Health Canada).

Reportable disease data are a key information source for determining the epidemiology of various infectious gastrointestinal illnesses in Ontario.¹⁻⁵ However, such data capture only a fraction of the total reportable illnesses in the community, and only those illnesses caused by pathogens that are deemed reportable.^{6,7} In Ontario, infectious gastrointestinal illness (IGI) reporting can be represented by a linear model of several sequential steps required for a case to be reported in the provincial surveillance system (Figure 1). To determine the under-reporting rate, single value estimates of the proportion reported at each step could be multiplied together and the inverse taken. However, there is considerable uncertainty (i.e., our lack of knowledge about the true values of these proportions) and variability (i.e., inherent randomness in these values) that are not captured by single value estimates, but which may instead be represented by a probability distribution.⁸ Therefore, the objective of this study was to determine a distribution of plausible values for the overall under-reporting rate, defined as the number of cases of IGI in the community for each case of enteric illness reported to the province, in Ontario.

METHODS

Distributions for the under-reporting rates were estimated by specifying an input distribution for each proportion in the reporting chain (Appendix 1), and multiplying these distributions together, similar to a series of conditional probabilities, using simulation methods in @RISK 4.5.2 (Palisade Corporation, New York) as an add-in for Microsoft® Excel 2000 (Microsoft Corporation, Redmond, WA).

Proportion of cases reported at each step

Input distributions (type of distribution and parameters) for the proportion of cases reported at each step in the reporting chain were estimated using data from several studies from the Public Health Agency of Canada's (formerly Health Canada's) National Studies on Acute Gastrointestinal Illness (NSAGI) initiative (www.phac-aspc.gc.ca/nsagi-enmga/).

The input distribution for the proportion of cases visiting a physician was estimated using data from the NSAGI Hamilton Population Study, a retrospec-

TABLE I

Input Distributions for the Proportions Reported at Each Step in the Reporting Chain, Used to Estimate the Under-reporting Rate for Infectious Gastrointestinal Illness in Ontario

Proportion	Input Distribution (parameters)*	Mean	5th percentile	Median	95th percentile
Reported to province	Normal (0.913, 0.168)	0.829	0.608	0.848	0.984
Reported to local health unit	Uniform (0.64, 0.97)	0.805	0.656	0.805	0.954
Tests positive	Uniform (0.074, 0.163)	0.118	0.078	0.118	0.158
Stool tested	Uniform (0.90, 1.00)	0.950	0.905	0.950	0.995
Stool submitted	Beta (7+1, 8-7+1)	0.800	0.571	0.820	0.960
Stool requested	Normal (0.260, 0.029)	0.260	0.213	0.260	0.308
Visits physician	Beta (61.368+1, 258.563-61.368+1)	0.235	0.194	0.235	0.280

* The type of distribution and its parameters, of the form: Normal (mean, standard error); Uniform (minimum, maximum); Beta (numerator + 1, denominator - numerator + 1)

tive survey of 3,496 randomly selected residents of Hamilton, Ontario, conducted from February 2001 to February 2002.⁹ The total number of cases and the number of cases who visited a physician reported in this study were weighted according to the age distribution of the 2001 Ontario census population. For the purpose of this analysis, a case of IGI was defined as having vomiting or diarrhea in the previous 28 days in the absence of a known non-infectious cause. Those reporting the following causes were excluded: pregnancy, menstruation, alcohol, drugs, chronic diseases, or medications. The weighted number of cases and the weighted number who visited a physician were used as parameters in a Beta distribution, a distribution commonly used to model uncertainty about a probability or proportion.^{10,11}

Of those who visited a doctor, the input distribution for the proportion of cases from whom a stool sample was requested was estimated using data from the NSAGI Hamilton Physician Study, a retrospective survey of family physicians and pediatricians in Hamilton, Ontario conducted in May 2001.¹² Each physician was asked to report the total number of cases of acute gastrointestinal illness seen in the previous 30 days and the number from whom a stool sample was requested. As sample means have the property of being normally distributed about the population mean,¹¹ a Normal distribution was used to model the variation in average stool request practices, with the mean proportion requested to submit a stool and its standard error used as parameters.

Of those requested to submit a stool sample, the input distribution for the proportion who actually submitted the requested sample was estimated using data from the NSAGI Hamilton Population Study. A Beta distribution was used to model our uncertainty about the true value of this proportion.

TABLE II

Estimated Cumulative Number of Cases of Infectious Gastrointestinal Illness Captured at Each Step in the Reporting Chain and Occurring in the Community, for Each Case of Enteric Illness Reported to the Province of Ontario

Reporting Chain Step	Mean	5th percentile	Median	95th percentile
Reported to province	1	1	1	1
Reported to local health unit	1.24	1.02	1.18	1.64
Tests positive	1.56	1.15	1.50	2.16
Stool tested	13.83	8.36	13.00	22.00
Stool submitted	14.57	8.82	13.68	23.22
Stool requested	18.73	10.52	17.22	32.05
Visits physician	72.80	39.19	66.70	126.71
Community cases	312.92	162.35	285.03	555.59

TABLE III

Correlation Between the Input Distributions for the Proportions in the Reporting Chain and the Distribution of the Overall Under-reporting Rate

Rank	Proportion	Correlation Coefficient
1	Tests positive	-0.595
2	Stool submitted	-0.385
3	Reported to province	-0.377
4	Reported to local health unit	-0.314
5	Visits physician	-0.276
6	Stool requested	-0.264
7	Stool tested	-0.066

Of those samples submitted to the laboratory, the input distribution for the proportion that were tested was estimated using Ontario data from the NSAGI National Laboratory Study, a retrospective survey of all laboratories in Canada licensed to process stool specimens.^{13,14} Laboratories were asked to describe testing and reporting practices, and proportions of tests positive for various organisms for the year 2000. The minimum and maximum proportions of stools tested as reported in the survey were used as parameters in a Uniform distribution, to model our uncertainty about the true value of this proportion. Uniform distributions are used when data are scarce or unavailable, since only a minimum and a maximum need to be specified.¹¹

Of those tested, the input distribution for the proportion of cases whose stool tested positive was estimated using Ontario data from the NSAGI National Laboratory Study. Laboratories were asked to report the total number of stool tests completed and the number of tests positive for (a) bacteria,

excluding *C. difficile*, (b) *C. difficile*, (c) ova and parasites, and (d) viruses. Since these data referred to *tests*, the minimum possible proportion of *cases* positive was estimated assuming each case had only one of the four tests done. The maximum possible proportion of *cases* positive was estimated assuming each case had all four tests done but was only positive for a maximum of one test. The minimum and maximum proportions were used as parameters in a Uniform distribution, to model our uncertainty about the true value of this proportion.

Of those who tested positive (for all organisms, both reportable and non-reportable), the input distribution for the proportion of cases reported to the local health unit by the laboratory was estimated using Ontario data from the NSAGI National Laboratory Study. The minimum and maximum proportions reported in this survey were used as parameters in a Uniform distribution, to model our uncertainty about the true value of this proportion.

Figure 1. Distribution of the estimated overall under-reporting rate of infectious gastrointestinal illness in Ontario, showing the number of cases in the community for each case reported to the province

Of those reported to the local health unit, the input distribution for the proportion of cases reported to the province was estimated using Ontario data from the NSAGI Public Health Reporting Study, a retrospective survey of all health units in Ontario and British Columbia.¹³ Health units were asked to report the total number of cases received in the year 2001, and the total number reported to the province. To model the variation in average reporting practices, the mean proportion reported to the province and its standard error were used as parameters in a Normal distribution.

Under-reporting rates

Distributions of values for the under-reporting rates at each step in the reporting chain, and the overall under-reporting rate, were generated in @RISK using simulation modelling with 10,000 iterations and Latin Hypercube sampling. In each iteration (where an iteration is a realization of the model using randomly selected values from each of the input distributions), the overall under-reporting rate and each of the step-specific under-reporting rates was calculated using randomly selected values from the proportions' input distributions. The overall under-reporting rate, that is the number of cases of IGI in the community for each case reported to the province, was obtained by multiplying together the series of proportions and taking the inverse. The under-reporting rates at each step in the reporting chain were obtained by multiplying together any proportion estimates subsequent in the reporting chain (i.e., those at higher steps) and taking the inverse. A sensitivity analysis was conducted to determine which input distribution(s) had the most influence on the generated overall under-reporting

rate distribution, by calculating and ranking correlation coefficients between each of the input distributions and the distribution for the overall under-reporting rate.

RESULTS

Proportion of cases reported at each step

Input distributions for each of the proportions in the reporting chain used in this analysis are shown (Table I), with the mean, median, 5th and 95th percentile values generated during the simulation.

Under-reporting rates

Mean, median, 5th and 95th percentile values of the distributions generated for the under-reporting rates at each step in the reporting chain, that is, the cumulative number of cases at each step for each case reported to the province, are shown in Table II. The distribution of the overall under-reporting rate is shown in Figure 1; for each case of enteric illness reported to the province of Ontario, the estimated number of cases of IGI in the community ranged from 105 to 1,389, with a median of 285, and a mean and standard deviation of 313 and 128, respectively. The input distribution with the most influence on the distribution of the overall under-reporting rate was the proportion who tested positive (Table III); that is, our uncertainty about the proportion who tested positive made the largest contribution to the uncertainty of the overall under-reporting estimate.

DISCUSSION

For each case of enteric illness reported to the province of Ontario, an estimated sever-

al hundred cases of IGI occur in the community. Multiplying a given number of reported enteric illnesses by the under-reporting rate presented here yields an estimate of the relative number of IGI cases in the community, under the assumption that the occurrence of IGI caused by non-reportable organisms is consistently related to the occurrence of IGI caused by reportable organisms. For example, 44,451 enteric illnesses (caused by 8 of the 12 reportable enteric pathogens) were reported in Ontario from 1997 to 2001,⁴ translating to an annual number of approximately 78 cases of reportable enteric illness per 100,000 population. Thus, these 78 reported cases per 100,000 population represented a mean of approximately 24,388 community cases of IGI per 100,000 population (caused by all pathogens, both reportable and not reportable, via all routes of transmission). Note that the estimates presented here do not account for any differential reporting by pathogen, severity of illness, risk factors, demographic or other factors.

The stochastic methods used here have also been used to estimate the under-reporting of *E. coli* O157:H7 in Ontario, where each case reported to the province corresponds to approximately four to eight cases missed by the surveillance system.⁸ Stochastic methods offer an advantage over the deterministic methods typically used in these types of analyses¹⁵ because they allow our uncertainty about the true input values to be accounted for in the final estimates. The distribution of estimated values for the under-reporting rate presented here reflects our uncertainty about the true values of the proportions reported at various steps in the reporting chain; future estimates can be improved by obtaining more accurate estimates of these proportions.

Assessing the relative impact of each input distribution on the estimated under-reporting distribution showed that our uncertainty about the proportion testing positive was most influential in determining the under-reporting distribution. It is important to note that the data used to estimate this proportion originally referred to the proportion of tests positive, from which the proportion of cases positive was estimated. Obtaining a more accurate estimate for this proportion is necessary, and is the most effective way to improve the accuracy and precision of the estimated under-reporting rate.

The range of values for the overall under-reporting rate presented here was higher than the English estimate of 136 (95% CI: 93-197) community cases of infectious intestinal disease for each case reported nationally.⁷ However, the higher under-reporting observed for Ontario is likely due to multiple factors, including differences in the reportable disease systems and data used to estimate under-reporting. Reasons for IGI under-reporting are numerous, discussed in detail elsewhere, and thus not presented here.^{7,13,14,16,17}

Considering that under-reporting may be affected by factors that vary by time and place, these results may not be valid for populations outside Ontario or for other time periods. As well, data specific to the City of Hamilton were used to estimate several proportions in the reporting chain. If Hamilton differs from the rest of Ontario with respect to these proportions (e.g., health care-seeking behaviours, stool request practices), these results will be biased.

In conclusion, enteric illnesses reported to the province of Ontario represent only a fraction of the total IGI occurring in the community, with an estimated 313 cases of IGI occurring in the community for every case of enteric illness reported to the province. Although reportable disease data are one source of information useful in describing the epidemiology of IGI, these data should be interpreted with caution, since few cases visit a physician and few physicians request stool samples. Program planners and public health personnel should consider this fact when developing population-based interventions. Additionally, future studies should address differential reporting by severity of illness, risk factors, routes of transmission, demographics and other factors, which was not addressed here. Future studies should also address pathogen-specific under-reporting rates, especially given that different pathogens appear to be reported at different rates.^{7,18-20}

REFERENCES

1. Majowicz SE, Michel P, Aramini JJ, McEwen SA, Wilson JB. Descriptive analysis of endemic cryptosporidiosis cases reported in Ontario, 1996-1997. *Can J Public Health* 2001;92(1):62-66.
2. Greig JD, Michel P, Wilson JB, Lammerding AM, Majowicz SE, Stratton J, et al. A descriptive analysis of giardiasis cases reported in Ontario, 1990-1998. *Can J Public Health* 2001;92(5):361-65.
3. Michel P, Wilson JB, Martin SW, Clarke RC, McEwen SA, Gyles CL. A descriptive study of verocytotoxigenic *Escherichia coli* (VTEC) cases

Appendix 1

Reporting chain for infectious gastrointestinal illness in Ontario, showing the sequential steps necessary for case capture in the provincial surveillance system

- 1) Case visits a physician
- 2) Stool sample is requested
- 3) Stool sample is submitted
- 4) Stool sample is tested at laboratory
- 5) Stool sample tests positive
- 6) Case is reported to local health unit
- 7) Case is reported to province

Reporting sequence may differ in certain situations, e.g., outbreaks, hospital in-patients

- reported in Ontario, 1990-1994. *Can J Public Health* 1998;89(4):253-57.
4. Lee M, Middleton D. Enteric illness in Ontario, Canada, from 1997 to 2001. *J Food Prot* 2003;66(6):953-61.
5. Ford MW, Odoi A, Majowicz SE, Michel P, Middleton D, Cieben B, et al. A descriptive study of human *Salmonella* serotype typhimurium infections reported in Ontario from 1990 to 1998. *Can J Infect Dis* 2003;14(5):267-73.
6. Palmer S, Houston H, Lervy B, Ribeiro D, Thomas P. Problems in the diagnosis of food-borne infection in general practice. *Epidemiol Infect* 1996;117:479-84.
7. Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, Tompkins DS, et al. Study of infectious intestinal disease in England: Rates in the community, presenting to general practice, and reported to national surveillance. *Br Med J* 1999;318:1046-50.
8. Michel P, Wilson JB, Martin SW, Clarke RC, McEwen SA, Gyles CL. Estimation of under-reporting rate for the surveillance of *Escherichia coli* O157:H7 cases in Ontario, Canada. *Epidemiol Infect* 2000;125:34-45.
9. Majowicz SE, Doré K, Flint J, Edge VL, Read S, Buffett MC, et al. Magnitude and distribution of acute, self-reported gastrointestinal illness in a Canadian community. *Epidemiol Infect* 2004;132(4):607-17.
10. Phillips CV, LaPole LM. Quantifying errors without random sampling. *BMC Med Res Methodol* 2003;3(9):1-10.
11. Vose D. *Risk Analysis: A Quantitative Guide*, 2nd ed. West Sussex, England: John Wiley & Sons, Ltd., 2000.
12. Health Canada. *Results of a Physician Study Pilot in the New City of Hamilton Region*. 2002. Available on-line at: www.hc-sc.gc.ca/pphb-dgsp/psagi-enmga/pub_e.html (Accessed on April 29, 2004).
13. Flint JA, Doré K, Majowicz SE, Edge VL, Sockett P. From stool to statistics: Reporting of acute gastrointestinal illnesses in Canada. *Can J Public Health* 2004;95(4):309-13.
14. Health Canada. *National Studies on Acute Gastrointestinal Illness Report of the 2001 Canadian Laboratory Survey*. 2002. Available on-line at: www.hc-sc.gc.ca/pphb-dgsp/psagi-enmga/pub_e.html (Accessed on April 29, 2004).
15. Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, et al. Food-related illness and death in the United States. *Emerging Infect Dis* 1999;5(5):607-25.
16. Tam CC, Rodrigues LC, O'Brien SJ. The study of infectious intestinal disease in England: What risk factors for presentation to general practice tell us about potential for selection bias in case-control studies of reported cases of diarrhea. *Int J Epidemiol* 2003;32:99-105.
17. de Wit MA, Kortbeek LM, Koopmans MP, de Jager CJ, Wannet WJ, Bartelds AI, et al. A comparison of gastroenteritis in a general practice-based study and a community-based study. *Epidemiol Infect* 2001;127(3):389-97.
18. Qamruddin AO, Keaney MGL, McCann R, Chadwick PR. Increased stool sampling during a waterborne outbreak of cryptosporidiosis does not increase the detection of other faecal pathogens. *J Clin Pathol* 2002; 55:271-74.
19. de Wit MAS, Koopmans MPG, Kortbeek LM, Wannet WJB, Vinjé J, van Leusden F, et al. Sensor, a population-based cohort study on gastroenteritis in the Netherlands, incidence and etiology. *Am J Epidemiol* 2001;154(7):666-74.
20. Gerba CP. Calicivirus: A major cause of food-borne illness. *J Food Sci* 2003;68(4):1136-41.

Received: June 1, 2004

Accepted: December 16, 2004

RÉSUMÉ

Contexte : En Ontario, la notification des maladies gastrointestinales infectieuses (MGI) peut être représentée par un modèle linéaire illustrant la marche à suivre pour saisir chaque cas dans le système provincial de surveillance des maladies à déclaration obligatoire. Comme on sait que les données sur les entéropathogènes à déclaration obligatoire ne représentent qu'une petite fraction de toutes les MGI présentes dans la collectivité, notre étude visait à estimer le taux de sous-notification des MGI en Ontario.

Méthode : Nous avons estimé la répartition des valeurs plausibles du taux de sous-notification en spécifiant les répartitions des paramètres d'entrée selon les proportions déclarées à chaque étape de la chaîne de notification, puis en multipliant ces répartitions entre elles à l'aide de méthodes de simulation. Grâce aux données de l'Étude nationale des maladies gastro-intestinales aiguës (ENMGA) de l'Agence de santé publique du Canada, nous avons déterminé les répartitions d'entrée (genre de répartition et paramètres) pour la proportion de cas déclarés à chaque étape de la chaîne de notification.

Résultats : Pour chaque cas de maladie entérique déclaré en Ontario, le nombre estimatif de cas de MGI dans la collectivité se situe entre 105 et 1 389, la médiane étant de 285, avec une moyenne de 313 et un écart-type de 128.

Conclusions : Selon nos estimations, chaque cas de maladie entérique déclaré en Ontario correspond à plusieurs centaines de cas de MGI présents dans la collectivité. Les données sur les maladies à déclaration obligatoire doivent donc être utilisées avec prudence lorsqu'on estime le fardeau de ces maladies. Les planificateurs de programmes et les intervenants en santé publique devraient en tenir compte lorsqu'ils élaborent des interventions pour l'ensemble de la population.