Epidemiology and Infection

cambridge.org/hyg

Symptom profiling for infectious intestinal disease (IID): a secondary data analysis of the IID2 study

A. L. Donaldson[1,2], H. E. Clough[1,2], S. J. O'Brien[1,2] and J. P. Harris[1,2]

[1]NIHR Health Protection Research Unit in Gastrointestinal Infections, University of Liverpool, Liverpool, UK and [2]Institute of Population Health Sciences, University of Liverpool, Liverpool, UK

Original Paper

Cite this article: Donaldson AL, Clough HE, O'Brien SJ, Harris JP (2019). Symptom profiling for infectious intestinal disease (IID): a secondary data analysis of the IID2 study. *Epidemiology and Infection* **147**, e229, 1–4. https://doi.org/10.1017/S0950268819001201

Received: 19 February 2019
Revised: 16 May 2019
Accepted: 4 June 2019

Key word:
Epidemiology; Gastroenteritis; Infectious intestinal disease; Surveillance; Symptomology

Author for correspondence: Anna Donaldson, E-mail: A.Donaldson2@liverpool.ac.uk

Abstract

Less than half of stool samples from people symptomatic with infectious intestinal disease (IID) will identify a causative organism. A secondary data analysis was undertaken to explore whether symptomology alone could be used to make inferences about causative organisms. Data were utilised from the Second Study of Infectious Intestinal Disease in the Community. A total of 844 cases were analysed. Few symptoms differentiated individual pathogens, but grouping pathogens together showed that viral IID was more likely when symptom onset was in winter (odds ratio (OR) 2.08, 95% confidence interval (CI) 1.16–3.75) or spring (OR 1.92, 95% CI 1.11–3.33), the patient was aged under 5 years (OR 3.63, 95% CI 2.24–6.03) and there was loss of appetite (OR 2.19, 95% CI 1.29–3.72). The odds of bacterial IID were higher with diarrhoea in the absence of vomiting (OR 3.54, 95% CI 2.37–5.32), diarrhoea which persisted for >3 days (OR 2.69, 95% CI 1.82–3.99), bloody diarrhoea (OR 4.17, 95% CI 1.63–11.83) and fever (OR 1.67, 95% CI 1.11–2.53). Symptom profiles could be of value to help guide clinicians and public health professionals in the management of IID, in the absence of microbiological confirmation.

Introduction

Infectious intestinal disease (IID) is characterised by the acute onset of diarrhoea and/or vomiting in otherwise healthy people caused by an infectious, transmissible organism [1]. In the UK, the surveillance of IID is based on statutory notifications, outbreak reports and syndromic surveillance from primary, secondary and remote health services [2, 3]. However, as the syndrome of diarrhoea and vomiting can have non-infectious causes, microbiological confirmation remains central to conclusive diagnosis of IID. Although microbiological testing is the gold standard, in some cohort studies, causative organisms have only been identified in 37–46% of samples from symptomatic individuals [1, 4, 5]. The likelihood of identifying a causative organism has been found to be affected by factors such as age, sex, occupation, the absence of specific symptoms such as vomiting, and the timing of the stool sample in relation to symptom onset [1]. Other factors such as the volume of the sample, the performance of the microbiological test and local organism testing policies may also impact on the isolation of organisms [6, 7].

This diagnostic gap means that for over half of symptomatic patients, the cause of illness will not be identified. Whilst the majority of IID cases are self-limiting, being aware of the underlying cause can be of value in both case and outbreak management. For outbreak situations, epidemiological criteria have been developed which utilise, among other factors, the proportion of people affected by given symptoms in order to make inferences as to the underlying organism. The most notable of these is the Kaplan criteria [8] which were developed in the 1980s in response to the lack of diagnostic tests available for isolating norovirus. Kaplan identified that, where no bacterial organism had been identified in stool cultures, outbreaks were more likely to be caused by norovirus when >50% of people were affected by vomiting; the incubation period was 24–48 h; and the mean duration of illness was 12–60 h. A subsequent re-evaluation of these criteria, once diagnostic tests became available, found them to be highly specific (99%) and moderately sensitive (68%) at distinguishing outbreaks of norovirus from bacterial IID outbreaks [9]. Other epidemiological criteria have also been proposed, including a greater fever–vomiting ratio [10] and a higher diarrhoea–vomiting ratio in bacterial outbreaks, suggesting that fever and diarrhoea are more indicative of a bacterial cause [11]. However, the basis of epidemiological criteria is the relative prevalence of symptoms occurring within a group of affected people, and as such they cannot be applied to individual cases. Seasonal outbreaks of IID may present as an increase in reporting of individual cases and therefore being able to ascribe likely cause to single cases of IID has public health and epidemiological value, as well as clinical application. This study uses data from a large community cohort and General Practice study to investigate whether

© The Author(s) 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE UNIVERSITY PRESS

symptoms alone can be used to make inferences as to the causative organisms for individual cases of IID.

Methods

Data sources

A secondary data analysis was undertaken using data from the Second Study of Infectious Intestinal Disease in the Community (IID2 Study), the methodology of which is detailed elsewhere [1, 12]. This analysis included data from the two main components of the IID2 study: the General Practice (GP) presentation study, which was a 12-month prospective study of people consulting a GP with symptoms of IID; and the prospective population-based cohort study, which involved weekly follow-up of healthy volunteers in the community to identify any symptoms of IID. The case definition for IID that was used in the original study was loose stools or clinically significant vomiting lasting <2 weeks, in the absence of a known non-infectious cause. Both studies utilised symptom questionnaires and stool sample testing of symptomatic people who met the case definition. Cases were included in this analysis if they had completed a symptom questionnaire and submitted a stool sample. Cases with negative stool samples, where no pathogen was identified, were excluded. Data from dual and triple infections were included multiple times; once for each organism identified, as the primary cause of symptoms could not be determined.

Data analysis

Multivariable logistic regression was used to determine the odds of a case being caused by a given pathogen based on reported symptoms. The explanatory variables included the symptoms outlined in the IID2 study symptom questionnaire, along with the participant's age and date of symptom onset (Table 1). Continuous data, namely symptom duration, date of illness onset and age, were categorised before inclusion in the regression models. Given that diarrhoea and vomiting are the predominant symptoms of IID and many people will have both, variables were created to capture cases of diarrhoea in the absence of vomiting, and vomiting in the absence of diarrhoea. These variables were used to explore whether this is a symptom profile which offers discrimination between pathogens. Phi coefficients were used to identify any significant correlations between the explanatory variables which might lead to mathematical problems with model fitting.

The outcome variable was the presence of the infectious organism. Pathogens which accounted for >10% of the total number of cases were analysed independently, to identify symptoms which distinguished them from any other cause of IID. Below this threshold, case numbers were too small to generate meaningful output for a single organism. Grouped organism models were used to capture differences between the broader classes of pathogen; bacteria, viruses and protozoa, sequentially comparing one class against any other cause of IID.

Statistical analysis was undertaken in R 3.3.2 [13]. Odds ratios (OR) were calculated using binomial backward stepwise regression. Models were selected based on the Akaike information criterion (AIC). Upper and lower 95% confidence intervals (CI) were calculated around each estimate.

Table 1. Explanatory variables included in the multivariate analysis and their coding

Explanatory variables	Coding
Symptoms from IID2 questionnaire[a]	
Diarrhoea days	>3 days = 1 ⩽3 days = 0
Bloody diarrhoea	Yes = 1 No = 0
Vomiting days	>3 days = 1 ⩽3 days = 0
Nausea	Yes = 1 No = 0
Nausea days	>3 days = 1 ⩽3 days = 0
Abdominal pain	Yes = 1 No = 0
Loss of appetite	Yes = 1 No = 0
Fever	Yes = 1 No = 0
Headache	Yes = 1 No = 0
Cough/nose/throat	Yes = 1 No = 0
Combined symptom variables	
Diarrhoea but no vomiting	Yes = 1 No = 0
Vomiting but no diarrhoea	Yes = 1 No = 0
Participant characteristics	
Age[b]	>16 years = 0, 5–16 years = 1, <5 years = 2
Date of onset[b]	Autumn(0), Winter(1), Spring(2), Summer(3)[c]

[a]'Not sure' responses from the original questionnaires were left blank and treated as missing data.
[b]Coded as factors for analysis.
[c]Seasons defined by meteorological calendar.

Results

There was a total of 1657 cases identified from the IID2 study which met the IID2 case definition and had both completed a questionnaire and submitted a stool sample. Of these, 898 cases (54%) were excluded from the analysis as no organism was identified from their stool sample. The total sample size for analysis was 844; including 69 dual infections and eight triple infections.

Norovirus was the most commonly identified cause of IID, and campylobacter was the commonest bacterial cause (Table 2). Only four pathogens met the criteria for organism-specific analysis; norovirus, campylobacter, rotavirus and sapovirus. The total number of protozoal infections was <10% of the total number of cases and consequently grouped organism models were only generated for bacterial and viral IID. To capture any important differences in symptoms, protozoa were included in the comparison group for both the bacterial and viral models.

The grouped organism models (Table 3) showed that the odds of the causative organism being bacterial were higher with diarrhoea in the absence of vomiting (OR 3.54, 95% CI 2.37–5.32), diarrhoea which persisted for >3 days (OR 2.69, 95% CI 1.82–3.99), bloody diarrhoea (OR 4.17, 95% CI 1.63–11.83) and fever (OR 1.67, 95% CI 1.11–2.53). The odds of a viral cause of illness were higher when symptom onset was in winter (OR 2.08, 95% CI 1.16–3.75) or spring (OR 1.92, 95% CI 1.11–3.33), the patient was under 5 years of age (OR 3.63, 95% CI 2.24–6.03) and there was loss of appetite (OR 2.19, 95% CI 1.29–3.72). Given protozoa have a similar aetiology to bacterial IID, as contrasted to viral IID, the analysis was repeated with protozoa assigned to the bacterial group to explore what impact this would have on the symptom profiling. The resulting viral and bacterial/protozoal models did not differ significantly from the above

https://doi.org/10.1017/S0950268819001201 Published online by Cambridge University Press

Table 2. Organisms and the associated number of cases, as included in the analysis

Pathogen	No. cases identified
Bacteria	238
C. difficile	11
C. perfringens	25
Campylobacter sp.	150
E.coli VTEC	15
Enteroaggregative *E.coli*	27
Salmonella sp.	9
Yersinia sp.	1
Viruses	576
Adenovirus	58
Astrovirus	36
Norovirus	237
Rotavirus	96
Sapovirus	149
Protozoa	30
Cryptosporidium sp.	15
Giardia sp.	15
Total organisms identified	844

Table 3. Grouped organism multivariate model outputs (OR with 95% confidence intervals) for bacterial and viral pathogens, as compared to any other pathogen

	Odds ratio (95% confidence intervals)	
Explanatory variable	Bacterial	Viral
Aged <5 years	0.25 (0.14–0.41)	3.63 (2.24–6.03)
Onset in winter	0.65 (0.35–1.19)	2.08 (1.16–3.75)
Onset in spring	0.70 (0.40–1.23)	1.92 (1.11–3.33)
Diarrhoea but no vomiting	3.54 (2.37–5.32)	0.27 (0.18–0.39)
Bloody diarrhoea	4.17 (1.63–11.83)	0.30 (0.11–0.79)
Diarrhoea lasting >3 days	2.69 (1.82–3.99)	0.33 (0.22–0.48)
Loss of appetite	0.44 (0.26–0.75)	2.19 (1.29–3.72)
Fever	1.67 (1.11–2.53)	0.59 (0.39–0.88)

models; the same explanatory variables were identified, but the significance of winter and spring in the bacteria/protozoa model was increased.

The organism-specific modelling generated less meaningful outputs. The campylobacter model largely mirrored the grouped bacterial model and did not provide any further discriminatory information. The virus-specific analysis for norovirus, rotavirus and sapovirus was sensitive to changes in the parameters of the models which led to inconsistent symptom profiles. Phi coefficients were used to identify any significant correlations between the binary explanatory variables which could impact on the model fitting. There was no evidence of significant co-linearity which would affect the modelling, although there were some mild-to-moderate correlations (phi coefficient <0.5) between some symptoms such as nausea and loss of appetite.

Discussion

This study has identified that people with IID who reported symptoms of diarrhoea in the absence of vomiting, diarrhoea lasting for more 3 days, bloody diarrhoea and fever were at increased odds of having a bacterial pathogen. Young age (<5 years), onset in spring or winter and loss of appetite were associated with increased odds of viral cause. These findings are consistent with other studies which have found associations between bacterial pathogens and symptoms of fever, bloody diarrhoea and prolonged illness, whilst vomiting and a short duration of symptoms have been associated with viral causes [14–17]. Epidemiological criteria, utilised in outbreak situations, have similarly highlighted the importance of vomiting and short duration of illness as indicative of norovirus, whilst symptoms such as fever and diarrhoea have been associated with bacterial outbreaks [8–11]. This study is largely consistent with these criteria, identifying similar associations between these symptoms and the class of the underlying pathogen. However, our analysis would suggest that vomiting and a short duration of symptoms are better ascribed to viral IID than any single viral pathogen. Furthermore, the duration of norovirus symptoms is known to be affected by individual risk factors, such as hospitalisation and age [18, 19]. Therefore, the 12–60 h duration used in the Kaplan criteria may be less applicable when considering individual cases of norovirus illness. This study used ⩽3 days to categorise symptom duration, which provided good discrimination between bacterial and viral causes of IID.

This analysis did not identify symptoms which could be used to adequately differentiate individual IID pathogens. This should act as a caution against making assumptions about the underlying organism on the basis of symptoms alone. However, this dataset did not contain sufficient numbers of some organisms to generate the statistical power necessary to model at the level of individual pathogens. Furthermore, the mild-to-moderate correlations identified between certain symptoms could make it harder for statistical models to distinguish individual pathogens on the basis of these symptoms alone.

The findings of this study have application for clinicians, public health professionals and epidemiologists, who use symptoms to generate hypotheses regarding causative organisms when managing cases and outbreaks of IID. This analysis would suggest that assumptions should not be made as to the individual pathogen in the absence of microbiological confirmation. However, given the different transmission patterns and natural histories of bacterial and viral IID [20], using symptom profiles to indicate a likely bacterial or viral cause could assist the early stages of outbreak investigations when microbiology is not yet available. This could help guide infection prevention and control; for example, viral causes are more likely to be spread person to person, whereas bacterial IID would raise suspicion of a food or animal contact. Given the large diagnostic gap for IID, the role of symptoms is still of vital importance to guide clinical and public health action. These findings could have further application for syndromic surveillance systems, enabling symptomatic cases to be categorised as either suspected bacterial or viral IID. However, the benefits of this would have to be weighed against the practical challenges of developing sensitive and specific case definitions that would

https://doi.org/10.1017/S0950268819001201 Published online by Cambridge University Press

be compatible with the level of symptom detail gathered and recorded by syndromic surveillance systems [21].

Strengths and limitations

This study utilised data from a large prospective cohort study [12], removing some of the reporting biases inherent within national surveillance data [22]. However, given that the severity and duration of illness is known to affect health-seeking behaviour and stool sample submission [22], mild short-lived illness is still likely to be underrepresented in these data. Despite the large size of the dataset, the total numbers of some organisms were too low to allow organism-specific models to be developed for all but the four most common causes. Furthermore, protozoa could not be examined as a separate class of pathogen due to small numbers. In this analysis, protozoa were included in the comparison group for both the bacterial and viral models. To explore the impact this could have had on the modelling, the analysis was repeated with protozoa assigned to the bacterial group. The resulting viral and bacterial/protozoal models did not differ significantly from the original models indicating that the group allocation of the protozoa had little impact on the findings of this analysis.

It should be considered that the grouped organism profiles will be naturally weighted by the relative prevalence of different organisms within each class; campylobacter accounted for almost two-thirds of all the bacterial cases and norovirus accounted for over 40% of viral cases. Consequently, the bacterial and viral models will disproportionality reflect the symptoms associated with these pathogens. However, this reflects real-life diagnostics where certain symptoms or organisms are more likely simply because they occur more commonly. Whilst this analysis could not identify symptom profiles which discriminated individual pathogens, this is an area that warrants further exploration. Future studies could also consider the role of co-infections, as co-infections have been found to affect the pathogenicity of organisms [23].

Conclusion

Symptom profiles could be used to help dissociate between bacterial and viral causes of IID however, symptoms do not allow further discrimination of individual organisms. Microbiology remains the gold standard and where possible, microbiological confirmation is recommended. However, in situations where microbiology is not available or results are inconclusive, symptom profiling could be of value for clinicians, public health professionals and epidemiologists to distinguish likely bacterial and viral pathogens to guide the management of cases and outbreaks of IID.

Financial support. This research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at the University of Liverpool in partnership with Public Health England (PHE), in collaboration with the University of East Anglia, University of Oxford and the Quadram Institute (Grant number NIHR HPRU 2012-10038).

Disclaimer. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or Public Health England.

Author ORCIDs. A. L. Donaldson, 0000-0002-5760-2753; J. P. Harris, 0000-0001-9606-9480

Conflict of interest. None.

References

1. **Tam CC *et al.*** (2012) The second study of infectious intestinal disease in the community (IID2 study). UK Food Standards Agency; project no B18021.
2. **Wall PG, *et al.*** (1996) Food poisoning: notifications, laboratory reports, and outbreaks – where do the statistics come from and what do they mean? *Communicable disease report. CDR review* **6**, R93–100.
3. **Elliot AJ, *et al.*** (2013) Syndromic surveillance – a public health legacy of the London 2012 Olympic and Paralympic Games. *Public Health* **127**, 777–781.
4. **UK Food Standards Agency** (2000) *A Report of the Study of Infectious Intestinal Disease in England*. London: Stationery Office.
5. **De Wit MA, *et al.*** (2001) Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. *American Journal of Epidemiology* **154**, 666–674.
6. **Atchison CJ** (2009) Clinical laboratory practices for the detection of rotavirus in England and wales: can surveillance based on routine laboratory testing data be used to evaluate the impact of vaccination? *Eurosurveillance* **14**(20). doi: 10.2807/ese.14.20.19217-en.
7. **Humphries RM and Linscott AJ** (2015) Laboratory diagnosis of bacterial gastroenteritis. *Clinical Microbiology Reviews* **28**, 3–31.
8. **Kaplan JE** (1982) Epidemiology of Norwalk gastroenteritis and the role of Norwalk virus in outbreaks of acute nonbacterial gastroenteritis. *Annals of Internal Medicine* **96**, 756–761.
9. **Turcios RM, *et al.*** (2006) Reevaluation of epidemiological criteria for identifying outbreaks of acute gastroenteritis due to norovirus: united States, 1998–2000. *Clinical Infectious Diseases* **42**, 964–969.
10. **Hedberg CW and Osterholm MT** (1993) Outbreaks of food-borne and waterborne viral gastroenteritis. *Clinical Microbiology Reviews* **6**, 199–210.
11. **Dalton CB, *et al.*** (1999) Outbreaks of enterotoxigenic Escherichia coli infection in American adults: a clinical and epidemiologic profile. *Epidemiology & Infection* **123**, 9–16.
12. **Tam CC** (2012) Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut* **61**, 69–77. doi: 10.1136/gut.2011.238386.
13. **The R Project for Statistical Computing**. Available at https://www.r-project.org/ (Accessed 14 June 2018).
14. **Liu L-J, *et al.*** (2005) Diagnostic value of bacterial stool cultures and viral antigen tests based on clinical manifestations of acute gastroenteritis in pediatric patients. *European Journal of Clinical Microbiology and Infectious Diseases* **24**, 559–561.
15. **Uhnoo I, Olding-Stenkvist E and Kreuger A** (1986) Clinical features of acute gastroenteritis associated with rotavirus, enteric adenoviruses, and bacteria. *Archives of Disease in Childhood* **61**, 732–738.
16. **De Wit MA, *et al.*** (2001) Etiology of gastroenteritis in sentinel general practices in the Netherlands. *Clinical Infectious Diseases* **33**, 280–288.
17. **Wiegering V, *et al.*** (2011) Gastroenteritis in childhood: a retrospective study of 650 hospitalized pediatric patients. *International Journal of Infectious Diseases* **15**, e401–e407.
18. **Chen SY, *et al.*** (2007) Molecular epidemiology and clinical manifestations of viral gastroenteritis in hospitalized pediatric patients in northern Taiwan. *Journal of Clinical Microbiology* **45**, 2054–2057.
19. **Lopman BA, *et al.*** (2004) Clinical manifestation of norovirus gastroenteritis in health care settings. *Clinical Infectious Diseases* **39**, 318–324.
20. **Hall AJ, *et al.*** (2013) Acute gastroenteritis surveillance through the National Outbreak Reporting System, United States. *Emerging Infectious Diseases* **19**, 1305–1309.
21. **Guasticchi G, *et al.*** (2009) Syndromic surveillance: sensitivity and positive predictive value of the case definitions. *Epidemiology and Infection* **137**, 662–671.
22. **Scallan E, *et al.*** (2006) Factors associated with seeking medical care and submitting a stool sample in estimating the burden of foodborne illness. *Foodborne Pathogens and Disease* **3**, 432–438.
23. **Moyo SJ, *et al.*** (2017) Comprehensive analysis of prevalence, epidemiologic characteristics, and clinical characteristics of monoinfection and coinfection in diarrheal diseases in children in Tanzania. *American Journal of Epidemiology* **186**, 1074–1083.

https://doi.org/10.1017/S0950268819001201 Published online by Cambridge University Press