Original Contribution

Community Incidence of Norovirus-associated Infectious Intestinal Disease in England: Improved Estimates Using Viral Load for Norovirus Diagnosis

Gemma Phillips*, Clarence C. Tam, Stefano Conti, Laura C. Rodrigues, David Brown, Miren Iturriza-Gomara, Jim Gray, and Ben Lopman

* Correspondence to Gemma Phillips, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom (e-mail: gemma.phillips@lshtm.ac.uk).

Initially submitted October 27, 2009; accepted for publication January 13, 2010.

Existing estimates of the incidence of infectious intestinal disease (IID) caused by norovirus are based on electron microscopy or reverse transcription-polymerase chain reaction (RT-PCR). Neither method accurately represents norovirus disease burden: Electron microscopy has poor diagnostic sensitivity, and RT-PCR has poor diagnostic specificity. In this study, viral load measurements were used to identify cases of norovirus-associated IID and to produce new incidence estimates for England. IID cases were ascertained in the Study of Infectious Intestinal Disease in England (1993–1996), and stool specimens were tested by semiquantitative real-time RT-PCR for norovirus. The age-adjusted community incidence of norovirus-associated IID was 4.5/100 person-years (95% credibility interval: 3.8, 5.2), equating to 2 million episodes/year. Among children aged less than 5 years, the community incidence was 21.4/100 person-years (95% credibility interval: 15.9, 27.7), and the incidence of consultations to general practitioners for norovirus-associated IID was 3.2/100 person-years (95% credibility interval: 2.6, 3.8), with 100,000 children visiting their general practitioner for norovirus-associated IID each year. Norovirus is the most common cause of IID in the community in England and is responsible for a similar number of pediatric primary care consultations as rotavirus.

Norovirus is the most common cause of infectious intestinal disease (IID) in the community in high-income countries (1–4), and a substantial prevalence of norovirus infection has been reported among IID cases seeking medical care (5). Existing estimates of norovirus-associated IID incidence in the community and among individuals presenting to their general practitioner in England are based on electron microscopy, which has poor diagnostic sensitivity for identifying norovirus-associated IID (6–8); it is very likely that these estimates underrepresent the burden of norovirus disease.

Reverse transcription-polymerase chain reaction (RT-PCR) is now the preferred diagnostic method for norovirus. However, semiquantitative real-time RT-PCR testing has demonstrated a wide range of viral loads in norovirus-infected IID cases (8); many IID cases shed norovirus at the same concentration as healthy individuals, with no recent history of IID (8, 9). It is therefore unlikely that all IID cases with norovirus infection detected by RT-PCR have disease caused by norovirus; another pathogen is probably causing illness in IID cases shedding norovirus at very low concentrations. Only individuals with IID caused by norovirus should be included in estimates of norovirus disease burden.

We demonstrated in previous work that viral load measurements can be used to identify IID cases with disease caused by norovirus and to exclude IID cases with “asymptomatic” norovirus infection concurrent with disease caused...
by another pathogen (8). In this study, we used viral load measurements from IID cases in the Study of Infectious Intestinal Disease in England to improve estimates of the incidence of norovirus-associated IID in the community and leading to general practice consultations. Accurate estimates of norovirus-associated IID incidence at the community level are essential for understanding the introduction of norovirus into health-care settings, where outbreaks cause substantial economic burden and service disruption (10), and for informing potential vaccination programs (11, 12).

MATERIALS AND METHODS

Recruitment and all-cause IID incidence

Data are taken from the Study of Infectious Intestinal Disease in England (“the IID Study”), conducted between 1993 and 1996 (13). The incidence of IID in the community, caused by any pathogen, was estimated in a prospective cohort, which was demographically representative of the population of England. Cohort members were actively followed up, with weekly null reporting, to ensure that all IID episodes were recorded (14).

The incidence of general practitioner consultations for IID, caused by any pathogen, was estimated by recruiting individuals with IID presenting to one of the 70 participating general practices (14). Incidence numerators were adjusted for underascertainment of IID cases, and denominators were adjusted for registered patients no longer using the practices (4, 14).

IID cases were individuals with diarrhea (any loose stools) or significant vomiting (≥2 vomiting episodes/24 hours), lasting less than 2 weeks, without a known noninfectious cause, preceded by a symptom-free period of at least 3 weeks (14). Healthy controls, with no recent history of IID, were recruited concurrently to cases in both study components, from the community cohort or from the general practice patient registration lists (13). Informed consent was obtained from all participants at the time of recruitment.

Specimens and testing

IID cases provided a fecal specimen during acute illness, and controls provided a specimen at recruitment. Norovirus was detected by electron microscopy, and specimens were archived in frozen storage (15). All specimens, including those previously positive by electron microscopy, were later retested for norovirus using a more sensitive RT-PCR assay. All norovirus RT-PCR-positive specimens were retested by using a semiquantitative RT-PCR assay (run for 40 cycles) (16). Recruitment and stool testing in IID cases are summarized in Table 1.

The cycle threshold value from the real-time RT-PCR assay provides a proxy measure of fecal viral load; it is inversely proportional to the amount of virus present in the specimen. The distribution of norovirus cycle threshold values in IID cases and controls used in this study has been described previously (8).

### Table 1. Summary of Case Recruitment and Stool Specimen Testing in the Community Cohort and General Practice Component of the Study of Infectious Intestinal Disease, England, 1993–1996

<table>
<thead>
<tr>
<th>Table 1. Summary of Case Recruitment and Stool Specimen Testing in the Community Cohort and General Practice Component of the Study of Infectious Intestinal Disease, England, 1993–1996</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Community Cohort</strong></td>
</tr>
<tr>
<td>Base population, person-years of follow-up</td>
</tr>
<tr>
<td>Ascertained cases, no.</td>
</tr>
<tr>
<td>Stool specimens, no.</td>
</tr>
<tr>
<td>Electron microscopy positive for norovirus, no.</td>
</tr>
<tr>
<td>Stool specimen archived, no.</td>
</tr>
<tr>
<td>RT-PCR positive for norovirus, no.&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cycle threshold value determined with real-time RT-PCR, no.</td>
</tr>
</tbody>
</table>

**Abbreviation:** RT-PCR, reverse transcription-polymerase chain reaction.

<sup>a</sup> Adjusted for registered patients no longer actively using participating general practices.

<sup>b</sup> Adjusted for underascertainment.

<sup>c</sup> Stool specimens were collected from patients in only 34 of the 70 general practices recruiting cases.

<sup>d</sup> Includes those previously positive by electron microscopy.

Calculating norovirus incidence

The incidence of norovirus-associated IID (INV) was calculated as follows:

\[
INV = I \times p(NV) \times A,
\]

where \(I\) is the incidence of all-cause IID/100 person-years, \(p(NV)\) is the proportion of IID cases positive for norovirus by RT-PCR, and \(A\) is a factor used to adjust for those IID cases with norovirus infection at low viral loads who therefore do not have disease caused by norovirus.

In a previous analysis of norovirus cycle threshold values from the IID Study, we used receiver operating characteristic (ROC) analysis to select a cutoff for attributing disease to norovirus in IID cases (8). However, standard ROC analysis does not provide confidence limits around the selected cutoff. In this analysis, the cycle threshold value distributions from the reference groups in the ROC analysis were used to calculate adjustment factor \(A\), incorporating uncertainty in these distributions due to sampling error into the incidence estimate. The reference-positive group included IID cases with norovirus detected by electron microscopy, because they have viral loads representative of where norovirus infection is causing disease (17, 18). The reference-negative group included healthy controls, because they have viral loads representative of where norovirus infection is not causing any illness.

Adjustment factor \(A\) was calculated as follows:

\[
A = \sum_{i=15}^{i=39} \frac{CT_i \times RP_i}{RP_i + RN_i},
\]

where \(RP_i\) is the moving average of the proportion of the
reference-positive group at cycle threshold (Ct) value i (over \(i - 2\) to \(i + 2\)); RN_i is the moving average of the proportion of the reference-negative group at cycle threshold value i (over \(i - 2\) to \(i + 2\)); and Ct_i is the proportion of IID cases positive by real-time RT-PCR with cycle threshold value i. Adjustment factor A varies between 0 and 1. Adjustment factor A is a weighted average of the relative frequency of the reference-positive and reference-negative groups at each cycle threshold value, weighted by the proportion of all norovirus-infected IID cases at each cycle threshold value (Figure 1).

Figure 2 shows the distribution of cycle threshold values in the reference groups and the value of the subcomponent \(\frac{\text{RP}_i}{\text{RP}_i + \text{RN}_i}\), which represents the relative frequency of the reference groups. At low-cycle threshold values, where viral loads are high and there are few individuals from the reference-negative group, the subcomponent \(\frac{\text{RP}_i}{\text{RP}_i + \text{RN}_i}\) is close to 1, indicating that the majority of IID cases with norovirus infection at these concentrations have disease caused by norovirus. In contrast, at the high-cycle threshold values (low viral loads) found in the majority of the disease-free reference-negative group, the subcomponent \(\frac{\text{RP}_i}{\text{RP}_i + \text{RN}_i}\) is close to 0, indicating that very few IID cases with norovirus infection at these concentrations have disease caused by norovirus.

Adjustment factor A was calculated separately for children aged less than 5 years and for older children and adults (aged 5 years or older) in the age-stratified and age-adjusted incidence.

Incidence estimation by Monte Carlo simulation in WinBUGS

The incidence of norovirus-associated IID was calculated by using Monte Carlo simulation in WinBUGS, version 1.4, software (19). Confidence limits for norovirus-associated IID incidence are provided as Bayesian credibility intervals from the posterior sampling distribution. The all-cause IID incidence/100 person-years (I) from the IID Study was modeled by using a log-normal distribution. Proportions were modeled by using binomial distributions with noninformative uniform priors. Multinomial distributions were used to model the cycle threshold value distributions, with noninformative Dirichlet prior distributions. The simulation was run for 300,000 iterations, from 3 different sets of initial values, to check convergence.

Separate simulations were run to estimate the incidence of norovirus in the community and the incidence of general practice consultations and to calculate age- and season-stratified incidence. The numbers of IID cases with norovirus cycle threshold values limited the number of age groups in which the community incidence could be presented. Age-adjusted incidence was calculated as a weighted average of the incidence in children aged less than 5 years and in older children and adults (aged 5 years or older); weights were taken from the mid-1994 population estimate for England, obtained from the Office of National Statistics, United Kingdom. The annual numbers of cases of norovirus-associated IID were calculated from the incidence estimates and the age-stratified mid-1994 population estimate for England.
Alternative methods for estimating the proportion of IID cases with disease attributable to norovirus

We used 3 further methods to estimate the proportion of IID cases with disease attributable to norovirus, which either do not require a control group or have been used in previous studies.

**Alternative method 1.** In previous studies using only RT-PCR, not semiquantitative real-time RT-PCR, the proportion of norovirus-infected IID cases with disease attributable to norovirus has been estimated as the difference in norovirus prevalence between the control group and IID cases (5). We calculated norovirus-associated IID incidence as follows:

\[
\text{INV} = \frac{I}{C_{0}p(NV)} \div \frac{I}{C_{0}p(NV)}\]

where \( p(NV)_{\text{case}} \) represents the norovirus prevalence among IID cases, and \( p(NV)_{\text{control}} \) represents the norovirus prevalence among controls.

**Alternative method 2.** We have previously defined a cutoff in norovirus genogroup II cycle threshold values for attributing disease to norovirus (8). We applied this cutoff (at cycle threshold value 30 for children aged <5 years and at cycle threshold value 33 for older children and adults) to IID cases with a cycle threshold value for either norovirus genogroup I or genogroup II. The proportion of IID cases with a norovirus cycle threshold value at or below the cycle threshold value cutoff was substituted for adjustment factor \( A \) in equation 1. To explore the effect of late specimen collection on norovirus incidence, we defined probable cases of norovirus-associated IID as those IID cases with a cycle threshold value above the cutoff, a specimen collected 5 or more days after symptom onset, and no other pathogen detected. These probable cases were added to the IID cases with a norovirus cycle threshold value at or below the cutoff.

**Alternative method 3.** We used mixture modeling to estimate the proportion of IID cases with a norovirus cycle threshold value that have disease attributable to norovirus, using only data from IID cases. This proportion was substituted for adjustment factor \( A \) in equation 1 and uncertainty represented by using a beta distribution, based on the confidence interval provided from the mixture model. Details of the mixture model are provided in the Web Appendix (http://aje.oxfordjournals.org/).

We also estimated the incidence of norovirus-associated IID based on electron microscopy testing using equation 3:

\[
\text{INV} = I \times P,
\]

where \( P \) is the proportion of cases positive by electron microscopy. The incidence of norovirus-associated IID based on classifying any norovirus RT-PCR-positive IID case as a case of norovirus-associated IID and the incidence of rotavirus-associated IID based on enzyme-linked immunosorbent assay diagnosis (20) (in children aged <5 years only) were calculated in the same way.

**RESULTS**

The crude community incidence of norovirus-associated IID was 4.1/100 person-years (Table 2); after age
adjustment, the community incidence was 4.5 episodes/100 person-years (Table 2). Incidence was highest in children aged less than 5 years, with 20% experiencing norovirus-associated IID every year. Community norovirus–associated IID incidence peaked between October and March (Table 2).

There were 0.5 general practice consultations for norovirus-associated IID/100 person-years (Table 2). The incidence of general practice consultations was highest among children aged less than 2 years, at 6.4/100 person-years. Approximately 1 of 7 children aged less than 5 years with norovirus-associated IID consulted a general practitioner, compared with 1 of 3 of those with rotavirus-associated IID in this study population (Table 2). The seasonality of general practice consultations for norovirus-associated IID was less pronounced than in the community (Table 2).

Incidence based on the cycle threshold value cutoff was slightly lower than using adjustment factor A, and the credibility intervals were narrower, as shown in Table 3 and the Web table (http://aje.oxfordjournals.org/). Subtracting the control norovirus prevalence from that in IID cases produced higher incidence estimates in young children, but lower estimates in older children and adults. Mixture modeling produced the lowest estimates.

**DISCUSSION**

This is the first study to use viral load measurements to estimate the incidence of norovirus-associated IID. A recent volunteer study showed that low norovirus viral loads, detectable by RT-PCR, are associated with asymptomatic infection (9). Consideration of viral load therefore provides the greatest diagnostic accuracy for identifying cases of norovirus-associated IID. Using such an approach, we have demonstrated that norovirus is the most common cause of IID, across all age groups, in the community in England (4), and that there is a substantial incidence of general practice consultations for norovirus-associated IID among young children (Table 4), similar to that caused by rotavirus.

Estimates of norovirus disease burden based on viral load are very likely to be more accurate than those based on electron microscopy, because electron microscopy has poor diagnostic sensitivity, or those based on RT-PCR, because it is possible to exclude IID cases who are RT-PCR positive.
but have low viral loads and are therefore unlikely to have disease caused by norovirus. We developed a method for calculating norovirus-associated IID incidence that allowed statistical uncertainty in the viral load measurements to be incorporated into the confidence limits. This was only possible with the use of Monte Carlo simulation methods to combine the multiple components of the calculation and their associated statistical uncertainty; this would have been extremely difficult using standard frequentist approaches, such as the Delta Method, because of the large number of variables in the calculation. Although the estimates presented here are based on data collected between 1993 and 1996, they provide the best available information on the burden of norovirus disease in England. Furthermore, these results are based on current diagnostic methods; as new studies are carried out, they will provide a baseline from which to assess changes in norovirus incidence over time that are not confounded by concurrent changes in the sensitivity of diagnostic methods.

There was limited resolution for estimating age-stratified incidence in the community because of the small sample size. We combined genogroup I and genogroup II norovirus infections in this analysis, rather than estimating adjustment factor A separately for each genogroup, also because of limited sample size. Similarly, in alternative method 2, we used a cycle threshold value cutoff developed for genogroup II specimens only, because no published cutoff exists for genogroup I. There is evidence that the real-time RT-PCR assay has lower efficiency for genogroup I norovirus strains (Jim Gray, Health Protection Agency Centre for Infections, personal communication, 2009), so that a given cycle threshold value may represent a higher viral load in the original stool specimen for some genogroup I strains, compared with genogroup II strains. Genogroup I noroviruses constituted less than 10% of the norovirus isolates in the study, so we believe that grouping the genogroups would result in conservative incidence estimates, rather than overestimation.

The concentration of norovirus excretion decreases substantially after symptom resolution (9). Although we made no direct adjustment for the possibility that some IID cases with high cycle threshold values may have had disease caused by norovirus, but did low viral loads at the time of specimen collection because their symptoms had already resolved, the method used to calculate adjustment factor A allows some IID cases with high norovirus cycle threshold values to be incorporated into the incidence estimate (Figure 2). It therefore indirectly allows for the possibility that some IID cases who truly had norovirus-associated IID had low viral loads at the time of testing. It is not possible to directly allow for late specimen collection using adjustment factor A, because it is calculated at the population level. Direct consideration of delay in specimen collection requires classification of norovirus disease status at the individual level, as was done when applying the cycle threshold value cutoff (alternative method 2). We recalculated the cutoff-based incidence of norovirus-associated IID, including probable cases (defined as having a high cycle threshold value, a late specimen, and no other detected pathogens) and found that the incidence was

<table>
<thead>
<tr>
<th>Alternative Methods</th>
<th>Cycle Threshold Value Cutoff</th>
<th>Incidence/100 Person-Years</th>
<th>95% Credibility Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Crude</td>
<td>4.1 (3.4, 4.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age adjusted</td>
<td>4.5 (3.8, 5.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt; 5 years</td>
<td>2.1 (1.4, 2.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 5 years</td>
<td>3.3 (2.6, 4.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subtraction Control Prevalence</td>
<td>3.9 (3.2, 4.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mixture Modeling</td>
<td>4.4 (3.8, 5.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prevalence</td>
<td>4.4 (3.8, 5.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cutoff</td>
<td>4.4 (3.8, 5.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Value</td>
<td>4.4 (3.8, 5.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>4.4 (3.8, 5.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjustment Factor A</td>
<td>4.4 (3.8, 5.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Value Cutoff</td>
<td>4.4 (3.8, 5.0)</td>
</tr>
</tbody>
</table>

Abbreviations: IID, infectious intestinal disease; RT-PCR, reverse transcription-polymerase chain reaction.
The new estimates of norovirus-associated IID incidence presented here are approximately 3 times higher in the community and 2.5 times higher at the general practitioner level than previous estimates for England based on electron microscopy (4). Accordingly, the ratio of community cases to cases presenting to general practitioners increased from 6 to 1, using electron microscopy diagnosis, to 8 to 1, using viral load measurements (4). The incidence estimates are approximately half those obtained by assuming that any IID cases with a positive RT-PCR result for norovirus has disease caused by norovirus, indicating that without consideration of viral load there is the potential for substantial overestimation of the burden of norovirus disease.

The community incidence estimates are comparable to those from a study in the Netherlands, which used RT-PCR testing to identify cases of norovirus-associated IID but had a narrower case definition for IID (3 or more loose stools, or 2 or more episodes of vomiting in 24 hours), which may not have been sensitive enough to ascertain all episodes of norovirus-associated IID at the community level (1). Similarly, the incidence of general practitioner consultations for norovirus was only slightly lower than that from a recent study in Germany, which used RT-PCR diagnosis for norovirus, but again this study had a narrower case definition for IID (2 or more loose stools, or 2 or more vomiting episodes in 24 hours) (21). The incidence of norovirus-associated IID may also have been higher than normal during our study because a new variant of norovirus emerged during 1995 and 1996 (22–24); emergence of norovirus variants has been associated with increased disease incidence (25–28).

The incidence of norovirus-associated IID in the community showed a slight peak in the winter and autumn months, while general practice consultations were reasonably constant throughout the year. Outbreaks of norovirus-associated IID in community settings in the United Kingdom show very little seasonality, in strong contrast to outbreaks in health-care settings, which show marked winter-time seasonality (29). A number of factors may contribute to these differing patterns of seasonality between community disease and outbreaks in different settings. First, community norovirus outbreaks are more commonly reported from catering settings, with transmission occurring through food contamination; while the prevalence of norovirus infection among food handlers is likely determined by the incidence of community disease, the driving factor in these outbreaks is breakdown in food hygiene practices, which is not a seasonal phenomenon. Second, it has been suggested that the
marked winter-time increase in hospital admissions for respiratory infections may drive the strong seasonality of norovirus outbreaks in this setting, and that there are distinct norovirus strains circulating in hospital populations and in the community that may have different transmission characteristics (29); therefore, the incidence of community disease or general practitioner consultations would not necessarily show the marked seasonality seen in health care-associated outbreaks. However, detailed characterization of the molecular epidemiology of norovirus infections in the community is needed, for comparison with the extensive data that already exist for hospital-acquired infections (30, 31), to understand better the factors driving the different seasonality of health-care outbreaks and community disease. Finally, it is also possible that there was more out-of-season norovirus transmission during this study because of the emergence of a new norovirus variant, as described above (32).

We have demonstrated, for the first time, how viral load measurements can be used to make improved estimates of norovirus disease burden. This approach is preferable to including all IID cases who are RT-PCR positive, regardless of their viral load, because many may be shedding norovirus at low concentrations, with disease caused by another pathogen. With the widespread use of RT-PCR for norovirus diagnosis in community-based studies, we recommend using a real-time platform to allow consideration of viral load when calculating norovirus incidence; we have shown that additional real-time testing in a subset of norovirus-infected IID cases would be sufficient to use this approach, providing the subset is of a reasonable size and is representative. Further work is needed to validate the use of a cycle threshold value cutoff for use in studies without a control group. Asymptomatic norovirus infection is very common (1, 16, 21, 33–35). Therefore, this quantitative approach provides the most rigorous estimate of norovirus disease burden.

ACKNOWLEDGMENTS

Author affiliations: Infectious Disease Epidemiology Unit, London School of Hygiene and Tropical Medicine, London, United Kingdom (Gemma Phillips, Clarence C. Tam, Laura C. Rodrigues, Ben Lopman); Department of Gastrointestinal, Emerging, and Zoonotic Infections, Health Protection Agency Centre for Infections, London, United Kingdom (Gemma Phillips, Ben Lopman); Statistics Unit, Health Protection Agency Centre for Infections, London, United Kingdom (Stefano Conti); and Virus Reference Department, Health Protection Agency Centre for Infections, London, United Kingdom (David Brown, Miren Iturriza-Gomara, Jim Gray).

The authors would like to acknowledge the help of the following people: Corrine Amar, Fenella Halstead, Dalila Choudhury, and Mihaela Cirdei for completing the laboratory work; Nick Andrews and Ben Cooper for advice on the Monte Carlo simulation; George Kafatos for providing advice; and STATA code for the mixture modeling.

This work was presented at the Fifth International Conference on Vaccines for Enteric Diseases, September 9–11, 2009, Malaga, Spain.

Conflict of interest: none declared.

REFERENCES


