REVIEW ARTICLE
Heterogeneity in norovirus shedding duration affects community risk

M. O. MILBRATH1*, I. H. SPICKNALL2, J. L. ZELNER3,4, C. L. MOE5 AND J. N. S. EISENBERG2,4

1 Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI, USA
2 Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA
3 Departments of Sociology and Public Policy, University of Michigan, Ann Arbor, USA
4 Center for the study of Complex Systems, University of Michigan, Ann Arbor, USA
5 Center for Global Safe Water, Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA

Received 16 August 2012; Final revision 22 January 2013; Accepted 8 February 2013; first published online 18 March 2013

SUMMARY
Norovirus is a common cause of gastroenteritis in all ages. Typical infections cause viral shedding periods of days to weeks, but some individuals can shed for months or years. Most norovirus risk models do not include these long-shedding individuals, and may therefore underestimate risk. We reviewed the literature for norovirus-shedding duration data and stratified these data into two distributions: regular shedding (mean 14–16 days) and long shedding (mean 105–136 days). These distributions were used to inform a norovirus transmission model that predicts the impact of long shedders. Our transmission model predicts that this subpopulation increases the outbreak potential (measured by the reproductive number) by 50–80%, the probability of an outbreak by 33%, the severity of transmission (measured by the attack rate) by 20%, and transmission duration by 100%. Characterizing and understanding shedding duration heterogeneity can provide insights into community transmission that can be useful in mitigating norovirus risk.

Key words: Epidemiology, Norwalk agent and related viruses, risk assessment, Susceptible-Infected-Removed (SIR) model, transmission.

INTRODUCTION
Norovirus is notorious for causing highly explosive epidemics, while also creating significant disease through community endemic transmission. As the most common cause of epidemic gastroenteritis across all age groups [1], noroviruses are responsible for more than 90% of viral gastroenteritis and about 50% of all-cause gastroenteral outbreaks worldwide [2]. In the USA, norovirus infections cause an estimated 71000 hospitalizations annually, costing nearly $500 million per year [3]. The role of endemic norovirus, while not as well characterized as its role in outbreaks, is major; endemic incidence is estimated at around 5% per year for all ages and 20% per year in children aged <5 years [4, 5].

Norovirus is transmitted through aerosolized vomitus and faecal contamination, either directly from person-to-person or indirectly through the
environment. Although there is little data to inform the relative roles of the different transmission pathways, there is evidence to support both symptomatic and asymptomatic transmission [6, 7]. Symptomatic shedding such as a vomiting event has been shown to be the source of many outbreaks, but it is likely that symptomatic individuals remove themselves from the community, limiting their rate of contact. Asymptomatic individuals, on the other hand, mix readily in the public. Additionally, post-symptomatic shedding has been shown to occur in high concentrations [8–10], at levels similar to symptomatic individuals [10].

While typical norovirus infections are self-limiting, with the shedding period lasting from a few days to a few weeks, many individuals can shed noroviruses long past symptom resolution [11–13], and some can shed for months or even years. Infants are often in this long-shedding group; an immature immune system or the presence of maternal antibodies (found in about 75% of infants aged <6 months [14]) can lead to extremely lengthy shedding periods. Similarly, immunocompromised individuals, who are unable to fully clear the infection, may shed norovirus for months or even years [15, 16].

Despite these data indicating the presence of long sheddors and the resultant heterogeneity in shedding duration, current norovirus transmission models assume short and homogeneously distributed durations of viral shedding. For example, of two recent models of nosocomial norovirus transmission, one assumes an infectious shedding period of <2 days [13] and the other a distribution from 1 to 8 days [17], while a recent household transmission model estimates an average infectious period of 1·17 days [18]. These short infectious periods may be appropriate in understanding transmission in small, enclosed environments such as a house or hospital, where symptomatic individuals shed a high magnitude of virus and quickly exhausts a closely connected and limited population; however, they cannot capture the effects of long sheddors in sustaining community transmission.

To better characterize the effect of long-term shedding on norovirus risk, we reviewed the existing literature for empirical data on human norovirus shedding duration. These data are used to estimate distributions of shedding duration that include realistic heterogeneity. This information was then used to inform a norovirus transmission model. Using this model, we demonstrate how the presence of a long-term shedding group affects risk outcomes including the number of infections, duration of transmission, and probability of outbreaks.

METHODS

Literature review search strategy
We reviewed the literature for individual-level human viral shedding duration data through the electronic database Scopus by a key word search using the terms ‘norovirus’ paired with ‘shedding’ and/or ‘excretion’. These articles were supplemented with sources identified from bibliographies of the resultant studies and with unpublished data from known norovirus researchers.

We restricted the results to English-language human studies containing individual-level data that were acquired through the use of polymerase chain reaction (PCR). Studies that compare PCR with earlier techniques (electron microscopy and ELISA) demonstrate substantial improvements in detection capabilities [8, 19], and thus more accurate estimates of viral shedding duration. PCR does not distinguish between viable and non-viable viruses, and because noroviruses are non-culturable, there is no way to determine the infectivity of those detected through this method [20]. Even with this constraint, PCR data can be used to estimate the duration of infectiousness. It has been shown that passage through a human host does not diminish infectivity [21], meaning that shed viruses, although not culturable, remain viable. Second, norovirus has an extremely small minimum infectious dose [22]—challenge studies have demonstrated infection with 4·8 RT–PCR units [8], and the estimated average probability of infection for a single norovirus particle is about 0·5 [21]. This low infectious dose, along with the high number of particles that are excreted, support the hypothesis that detection indicates a sufficient dose, even if some viral particles have mutated to have non-infectious capsids while remaining sufficiently intact to protect the easily degradable RNA.

Estimating the distribution of shedding durations
In the challenge studies, shedding duration was calculated from the first positive stool sample to the last positive stool sample. For other types of studies (e.g. outbreak investigations), the date of the first positive stool sample was not known, so we assumed that
viral shedding starts with the onset of symptoms and continues until the last positive sample. If shedding lengths were reported in days after inoculation [8], we subtracted the number of days to symptom onset to account for the incubation period, and for consistency among the datasets. Although pre-symptomatic shedding has been reported in <30% of the population [23], it is generally for <1 day [8, 24, 25] and there is no indication that it would affect the individual shedding groups differently.

We used two methods to identify the long-shedder population (Table 1). First, we examined the literature to identify individual characteristics potentially related to long-term shedding. If an individual was identified in the original study as having one or more of these characteristics s/he was categorized into the long-shedding group (operational long shedders). Because not all individuals in these categories were long shedders, and because others who do not have these characteristics may shed for long times, we developed a comparative method. With the second definition, long shedders were those whose shedding durations exceeded a set maximum shedding length, regardless of individual characteristics (functional long shedders). We chose the value of 34 days as this cut-off, as it was the maximum duration in controlled experiments using only healthy individuals [C. L. Moe, previously unpublished data (available in the Supplementary material)] [24, 26]. Because of the uncertainty in these estimates, we also performed a sensitivity analysis on this parameter.

We fitted the data to multiple distributions, and using a likelihood ratio test we found that both the lognormal and gamma distributions fit the data well. We chose gamma distributions based on biological plausibility [27], precedence [28], and because this distribution can be included in a deterministic model using multi-compartmental waiting times when the shape parameter is >1 (see [18] for an explanation). The stratified data were fitted to gamma distributions using the dgamma function in R [29], which estimates parameters using maximum likelihood.

<table>
<thead>
<tr>
<th>Criteria used for data stratification</th>
<th>Regular shedders</th>
<th>Long shedders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational</td>
<td>&gt;1-year-old and competent</td>
<td>and/or &lt;1-year-old</td>
</tr>
<tr>
<td>Functional</td>
<td>≤34 days</td>
<td>&gt;34 days</td>
</tr>
</tbody>
</table>

**Table 1. Criteria used for data stratification**

---

![Model schematic. S. Susceptible; E, exposed; I, infected; R, recovered; subscript (R), regular shedding; subscript (L), long shedding. Parameters: β, transmission probability; ε, incubation period; γR, 1/regular-shedder infectious period; γL, 1/long-shedder infectious period; ρ, fraction that is long-shedding.](image)

**Fig. 1.** Model schematic. S, Susceptible; E, exposed; I, infected; R, recovered; subscript (R), regular shedding; subscript (L), long shedding. Parameters: β, transmission probability; ε, incubation period; γR, 1/regular-shedder infectious period; γL, 1/long-shedder infectious period; ρ, fraction that is long-shedding.

**Model design**

The data from the review were used to parameterize two modified susceptible (S), exposed (E), infectious (I), recovered (R) transmission models (SEIR models) based on the schematic in Figure 1 and the equations below.

\[
\begin{align*}
\frac{dS_R}{dt} &= -\beta S_R (I_L + I_R), \\
\frac{dS_L}{dt} &= -\beta S_L (I_L + I_R), \\
\frac{dE_R}{dt} &= \beta S_R (I_L + I_R) - \varepsilon E_R, \\
\frac{dE_L}{dt} &= \beta S_L (I_L + I_R) - \varepsilon E_L, \\
\frac{dI_R}{dt} &= \varepsilon E_R - \gamma_R I_R, \\
\frac{dI_L}{dt} &= \varepsilon E_L - \gamma_L I_L, \\
\frac{dR_R}{dt} &= \gamma_R I_R, \\
\frac{dR_L}{dt} &= \gamma_L I_L.
\end{align*}
\]

We first examined a standard deterministic transmission model that includes both a regular-shedding group (represented by a subscript R) a long-shedding group (represented by a subscript L). Second, we examined a stochastic, discrete-time version of the deterministic model. In both of these models, individuals may be in any of the following states, reflecting the within-host stages of norovirus infection: susceptible (S_R or S_L), exposed/incubating (E_R or E_L), infectious/shedding (I_R or I_L), or recovered (R_R or R_L).

For the deterministic model, vital dynamics such as host births and deaths are excluded because of the relatively short time scale of the simulations (2500 days). We also assumed permanent immunity after recovery. We assumed the absence of competing strains and co-infection, and that host mixing between the regular- and long-shedding groups is proportional.
Infectious states (IR and IL) are assumed to exert an equal force of infection due to (1) low infectious dose [21], (2) high levels of asymptomatic excretion [7, 30], and (3) lack of correlation between virus titre in the first positive stool with shedding length [9]. We use a singular infectious period that does not account for symptoms. We examined the more complete process that includes symptomatic transmission and found that our conclusions were robust; even when 50% of infections resulted in a 1-day symptomatic period that was 1000 times more contagious, the importance of long shedding remained (see Supplementary material).

The population is split into two groups: a fixed proportion (\(\rho\)) is in the long-shedding group, with the remainder (1 - \(\rho\)) in the regular-shedding group. When time (\(t\)) = 0, an infectious seed of one individual is in the regular exposed state (ER), and the remainder are in the susceptible states (SR, SL). Susceptible individuals become exposed (ER or EL) at an average per-infected rate of \(\beta\), and transition into the corresponding infectious/shedding state (IR or IL) at an average rate of \(\alpha\), which is assumed to be the same for both shedding groups. Individuals enter recovered states (RR and RL) at rate \(\gamma_R\) for regular shedders, and \(\gamma_L\) for long shedders. Estimated parameters used in these simulations are provided in Table 2.

We estimate the percentage of the U.S. population in the operational long-shedding group, \(\rho\) (infants and/or immunocompromised), to be about 5%. The number of infants born in the U.S. in 2007 was 4317119 [31] in a population of 304059724 [32], about 1.4% of the population. The number of immunocompromised individuals in the USA has been previously estimated to be about 10 million individuals (3.6% of the U.S. population), a number that includes organ transplant recipients, HIV-infected individuals, and cancer patients [33].

We assessed risk in the deterministic model by examining changes in the basic reproduction number (\(R_0\)), which is a measure of the potential of a disease to spread in a population. \(R_0\) is typically defined as the expected number of secondary infections produced by a single index case in a completely susceptible population [34]. It can be expressed as the product of the expected duration of the infectious period and the rate at which secondary infections occur; in a standard SEIR model with a mean infectious period of \(1/\gamma\) and a transmission rate of \(\beta\), \(R_0\) is given by equation (1).

\[
R_0 = \frac{\beta}{\gamma}.
\]  

In our model, where multiple infectious phases are possible, \(R_0\) is the sum of the expected number of secondary cases generated by an individual in each state, weighted by the probability that the index case will occupy that state. For the model presented here, \(R_0\) is a product of the transmission probability \(\beta\) and the expected time an individual will spend in each of the infectious states IR and IL (1/\(\gamma_R\) and 1/\(\gamma_L\)) and the probability \(\rho\) of being in the long-shedding group [equation (2)].

\[
R_0 = \beta \left( \frac{1 - \rho}{\gamma_R} + \frac{\rho}{\gamma_L} \right).
\]  

We use a discrete-time stochastic model to examine probability, duration, and severity of outbreaks, where individuals can be in the same disease states as presented in the deterministic formulation above (SR, SL, ER, EL, IR, IL, RR, RL; Fig. 1). In this model, we assumed a single population with a frequency-dependent contact rate. The model is initialized with a population of size 10000, with one infected individual at \(t=0\) and the rest of the population in the susceptible states. The total number of

---

**Table 2. Parameter definitions and values used in simulations**

<table>
<thead>
<tr>
<th>Parameter definition (units)</th>
<th>Mean value* (1/days)</th>
<th>Gamma parameters</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta)</td>
<td>Transmission rate</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td>(\varepsilon)</td>
<td>1/length of exposed period</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>(\gamma_{RO})</td>
<td>1/ regular infectious period (operational)</td>
<td>0.061 (1/16.4)</td>
<td>2.2–7.4</td>
</tr>
<tr>
<td>(\gamma_{LO})</td>
<td>1/ long-shedding infectious period (operational)</td>
<td>0.009 (1/105.6)</td>
<td>0.8–129.1</td>
</tr>
<tr>
<td>(\gamma_{RF})</td>
<td>1/ regular infectious period (functional)</td>
<td>0.069 (1/14.5)</td>
<td>2.7–5.3</td>
</tr>
<tr>
<td>(\gamma_{LF})</td>
<td>1/ long-shedding infectious period (functional)</td>
<td>0.007 (1/136.0)</td>
<td>0.7–199.4</td>
</tr>
<tr>
<td>(\rho)</td>
<td>Fraction that is long shedding</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

* Value used when parameter was not varied in simulation.
new infections $X_t$ is drawn from a binomial distribution ($S_t$, $\lambda t$), where $S_t$ is composed of both $S_R$ and $S_L$ individuals. The exposed period is assumed to follow an exponential distribution with mean duration $1/\lambda$. The regular-shedding and long-shedding infectious periods ($1/\gamma_R$ and $1/\gamma_L$) are assumed to follow gamma distributions with parameters obtained by maximum-likelihood estimation using data collected from our literature review (Table 2).

**RESULTS**

**Data extraction**

The literature search resulted in a total of 18 English-language studies with individual-level human norovirus molecular-shedding duration data (listed in Table 3), and we received two previously unpublished datasets (for unpublished data see Supplementary material).

Shedding duration ranged from 2 to 898 days across all studies ($N = 168$ individuals) with total mean and median values of 49 days and 19 days, respectively. Much of the variability in this dataset is explained by stratification of the population on two individual characteristics: compromised immunity and infancy. Sixty-two of the 168 individuals were labelled as immunocompromised and/or infants, thus considered together as operational long shedders. Mean shedding lengths between these groups differed by about 90 days (Table 4). The immunocompromised group included a child with a genetic disorder [35], transplant recipients [36–40], paediatric oncology patients [41, 42], and one individual with multiple sustained norovirus infections whose immune status was not evaluated, but was assumed to be immunocompromised [43]. Infants were those individuals whose age was reported as <1 year in the study. It is reasonable to expect differences in shedding length to vary by genotype as well, because of different levels of immune response, but we did not have sufficient data to examine this possibility.

Infants that were immunocompetent shed on average for 22-1 days, 1.3 times longer than those in the

<table>
<thead>
<tr>
<th>Study</th>
<th>$N$</th>
<th>Age (years)</th>
<th>Immune status</th>
<th>Genotype</th>
<th>Range (days)</th>
<th>Mean (days)</th>
<th>Median (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Florescu et al. (2008) [36]$^b$</td>
<td>1</td>
<td>1–8</td>
<td>-</td>
<td>G2</td>
<td>240</td>
<td>240</td>
<td>240</td>
</tr>
<tr>
<td>Gallimore et al. (2004) [35]$^c$</td>
<td>1</td>
<td>Child</td>
<td>-</td>
<td>rGII.3</td>
<td>156</td>
<td>156</td>
<td>156</td>
</tr>
<tr>
<td>Goller et al. (2004) [51]$^d$</td>
<td>1</td>
<td>18</td>
<td>+</td>
<td>n.r.</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Kirkwood &amp; Streitberg (2008) [43]$^f$</td>
<td>9</td>
<td>0–2</td>
<td>+ (6), – (3)</td>
<td>GII.4, GII.6</td>
<td>2–100</td>
<td>30.8</td>
<td>15</td>
</tr>
<tr>
<td>Lee et al. (2008) [37]$^g$</td>
<td>1</td>
<td>0.8</td>
<td>-</td>
<td>GII.2</td>
<td>114</td>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td>Leon et al. (2011) [24]$^h$</td>
<td>13</td>
<td>18–48</td>
<td>+</td>
<td>GII.1</td>
<td>8–35</td>
<td>21.3</td>
<td>21</td>
</tr>
<tr>
<td>Ludwig et al. (2008) [41]$^i$</td>
<td>9</td>
<td>0–16</td>
<td>+</td>
<td>GII.4, GII.3</td>
<td>22–433</td>
<td>95.7</td>
<td>46</td>
</tr>
<tr>
<td>Murata et al. (2007) [14]$^j$</td>
<td>23</td>
<td>0–3</td>
<td>+</td>
<td>n.r.</td>
<td>5–47</td>
<td>17.3</td>
<td>16</td>
</tr>
<tr>
<td>Nilsson et al. (2003) [38]$^k$</td>
<td>1</td>
<td>65</td>
<td>-</td>
<td>GII</td>
<td>670</td>
<td>670</td>
<td>670</td>
</tr>
<tr>
<td>Parashar et al. (1998) [52]$^l$</td>
<td>1</td>
<td>Adult</td>
<td>+</td>
<td>GII</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Siebenga et al. (2008) [49]$^n$</td>
<td>8</td>
<td>0–69</td>
<td>-</td>
<td>GII.4, GIIb–GII.3</td>
<td>35–182</td>
<td>89.3</td>
<td>77</td>
</tr>
<tr>
<td>Seitz et al. (2011) [26]$^o$</td>
<td>9</td>
<td>18–50</td>
<td>+</td>
<td>GII.1</td>
<td>4–34</td>
<td>17.6</td>
<td>15</td>
</tr>
<tr>
<td>Simon et al. (2006) [42]$^p$</td>
<td>12</td>
<td>0–24</td>
<td>-</td>
<td>n.r.</td>
<td>3–140</td>
<td>40.4</td>
<td>23</td>
</tr>
<tr>
<td>Tu et al. (2008) [9]$^q$</td>
<td>14</td>
<td>63–93</td>
<td>+</td>
<td>GII</td>
<td>8–44</td>
<td>25.2</td>
<td>25.5</td>
</tr>
<tr>
<td>Westhoff et al. (2009) [40]$^r$</td>
<td>2</td>
<td>73</td>
<td>-</td>
<td>GII</td>
<td>90–165</td>
<td>127.5</td>
<td>127.5</td>
</tr>
<tr>
<td>C. L. Moe, unpublished (1)$^s$</td>
<td>22</td>
<td>18–50</td>
<td>+</td>
<td>GII.1</td>
<td>8–20</td>
<td>8.3</td>
<td>7</td>
</tr>
<tr>
<td>C. L. Moe, unpublished (2)$^t$</td>
<td>9</td>
<td>18–50</td>
<td>+</td>
<td>GII.2</td>
<td>3–20</td>
<td>6.9</td>
<td>5</td>
</tr>
</tbody>
</table>

+, Immunocompetent; –, immunocompromised; n.r., not reported; r, recombinant.

$^a$ Challenge study. $^b$ Case study of infant transplant recipient. $^c$ Case study of child with genetic disorder.

$^d$ Investigation of healthcare worker. $^e$ Outbreak investigation in an elderly care facility. $^f$ Prospective hospital study.

$^g$ Retrospective study of paediatric oncology patients. $^h$ Outpatients at a paediatric clinic. $^i$ Case study. $^j$ Outbreak investigation of food handlers. $^k$ Study of transplant recipients. $^l$ 2-year hospital survey. $^m$ Outbreak investigation in paediatric oncology and haematology unit.

---

**Table 3. Description of studies and data used in empirical review**
regular-shedding group (immunocompetent and >1 year old) who shed for a mean of 16·4 days. Regardless of age, immunocompromised individuals shed on average >6 times longer than regular shedders (139·1 days for >1 year old and 106·6 days for infants). The observed effect of both compromised immunity and infancy (90·2 days) is less than the expected additive joint effect (128·4 days). We tested for additivity using linear regression in R [29]; the P value for this test was <0·01, prompting us to reject the additive effects model. This indicates that either condition can cause long shedding, but both together (being both an infant and immunocompromised) does not result in extremely long-shedding durations beyond the effect of either condition alone.

This operational definition is useful for the a priori identification of high-risk individuals and for understanding the causes underlying long shedding, but was not ideal for our analysis as many cases in our dataset are missing data on age or immune status, and thus cannot be correctly classified using our operational criteria. Further, some individuals who are labelled as immunocompromised, and therefore labelled as long shedders with our operational definition, actually shed for short durations. We therefore stratified the data based on actual shedding length (functional long shedders shed for >34 days). The mean of the two groups by the functional definition differs by over 120 days (Table 4).

The fitted distribution for the unstratified data is highly variable (Fig. 2). When we stratify the data based on both of our defined shedding groups (functional and operational), much of this variability remains in the long-shedding group (coefficient of variation: 147·1 operational, 121·1 functional), while the regular-shedding groups are more homogenous (coefficient of variation: 71·3 operational, 58·4 functional) (Table 4). Using the law of total variance, the variance of the aggregated populations is 10826·2 for the operational definition and 20989·2 for the functional definition. By dividing the variance of the means by the total variance, we estimated that stratification into long-shedding and regular-shedding groups explains 35% (functional) to 37% (operational) of the total variance.

**Deterministic model results**

Changes in $R_0$ determine whether an epidemic will occur as well as the subsequent course of transmission. If $R_0 \leq 1$, transmission will be limited to sporadic secondary infections, and a major outbreak will not occur, whereas if $R_0 > 1$ an outbreak can result. Based on equations (1) and (2), and assuming $\rho = 5\%$, the inclusion of a long-shedding group [mean=105·6 days (operational) and 136·3 days (functional)], increases risk by increasing $R_0$ (1·3-fold increase for the operational definition, 1·5 for functional). Because our value of $\rho$ is estimated, we examined the effect of its increase; if $\rho$ grows to 10% of the population (e.g. increased survival rates for those with compromised systems or an increase in birth rate), the inclusion of the long-shedding group results in a 1·5-fold (operational) or 1·7-fold (functional) increase in $R_0$ over a population without long shedders. Similarly, a twofold increase in the mean shedding length of the long-shedding group ($1/\gamma_L$) results in a 1·6-fold (operational) or 1·8-fold (functional) increase in the value of $R_0$ (Table 5). The effects of changes in these long-shedding related values are independent of the transmission rate, $\beta$, meaning that resultant relative increases in $R_0$ are robust to scenarios of different transmission rates, including environmental or symptomatic transmission.

**Stochastic model results**

Results from stochastic simulations demonstrate that the presence of long shedders increases the probability that an outbreak will occur while increasing the
severity and duration of transmission when disease does not die out. We examined changes in the probability of high transmission events where high transmission was defined as >200 cases. Based on our simulation results, transmission above this level was sustained long term, and below this level would die out. In the absence of long shedders ($\rho = 0$, i.e. the entire population draws shedding durations from the functional regular-shedding distribution), there is a 23% chance that high transmission will occur in 1000 runs using our estimated parameters (Table 6). If we allow 5% of the population to draw from the functional long-shedding distribution, the probability of high transmission increases to 31%. This effect is less evident using the operational definition, where the probability remains at about 37% (Table 6). As shown in Figure 3, including a long-shedding group by either definition increases both the attack rate, and its duration of sustained transmission. With the addition of long shedders, the mean attack rate increases 145% (operational) to 225% (functional) over a population with only regular shedders (Table 6). The duration of sustained transmission also increases, approximately doubling for both definitions (operational: 633–1199 days; functional: 756–1683 days). Figure 3 demonstrates that these outcomes are dependent on our method of stratification.

Sensitivity analysis

Our estimates are based on an incomplete dataset of all true shedding times ($1/\gamma_L$) due to publication bias and the difficulty of obtaining data on asymptomatic community infection. We therefore examined the sensitivity of our outcomes over a wide range of values

---

**Table 5. Deterministic model results: estimated fold increase in $R_0$ with addition of long shedders**

<table>
<thead>
<tr>
<th></th>
<th>$\rho = 0.05$, estimated $\gamma_L$</th>
<th>$\rho = 0.05$, estimated $\gamma_L^*2$</th>
<th>$\rho = 0.10$, estimated $\gamma_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional definition</td>
<td>1.3</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>Operational definition</td>
<td>1.5</td>
<td>1.8</td>
<td>1.7</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Shedding lengths and fitted gamma distributions of individual shedding duration data from empirical review. Data are truncated at 200 days; the figure does not include seven individuals who shed for >200 days (up to 898 days). Grey bars indicate operational long shedders (infants and immunocompromised individuals) and white bars indicate operational regular shedders (immunocompetent non-infants). The dotted vertical line at 34 days delineates the cut-off point for the functional definition. Dashed lines represent fitted gamma distributions for the operational definition (regular-shedder parameters = 2.2, 7.4; long-shedder parameters = 0.8, 129.1) and solid lines are gamma distributions fitted to the functionally defined populations, where regular shedders < 34 days, long-shedders ≥ 34 days (regular-shedder parameters = 2.7, 5.3; long-shedder parameters = 0.7, 199.4).
for this parameter. The probability of high transmission increases by about 10% as the mean long-shedding duration ($1/\gamma_L$) increases from 20 to 250, levelling off at a maximum probability of 42% for the operational definition and 34% for the functional definition when $\gamma_L > 100$ days (Fig. 4). The higher probability in the operational definition is again attributed to the wider range and higher mean shedding duration of the regular-shedding group, which contains the bulk (95%) of the population.

The total number of infected individuals also increases with $1/\gamma_L$, as shown in Figure 5. In simulations in a population of 10000 individuals, the maximum number of cases is about 8000 for both definitions (80% of the population), and remains well above the average number of cases in the absence of long shedders (46% operational, 29% functional). Transmission duration, measured as the last time at which there is at least one infected individual in the population, increases as $1/\gamma_L$ increases (Fig. 6).

Table 6. Results of stochastic simulations using parameters from the review

<table>
<thead>
<tr>
<th></th>
<th>Probability of outbreak (%)</th>
<th>Mean attack rate (%)</th>
<th>Fold increase in outbreak duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational – no long shedders</td>
<td>36</td>
<td>45.8</td>
<td>—*</td>
</tr>
<tr>
<td>Operational – with long shedders</td>
<td>38</td>
<td>67.1</td>
<td>1.9</td>
</tr>
<tr>
<td>Functional – no long shedders</td>
<td>23</td>
<td>29.4</td>
<td>1.2</td>
</tr>
<tr>
<td>Functional – with long shedders</td>
<td>31</td>
<td>66.3</td>
<td>2.7</td>
</tr>
</tbody>
</table>

* Set as baseline, column indicates fold increase over this value.

Fig. 3. Outbreak duration and severity using shedding length parameter values from our empirical review (Table 2). Only model runs resulting in high transmission events (>200 cases) are shown. Runs with both long-shedding and regular-shedding groups are represented by squares, runs with only regular shedders are represented by circles. Filled shapes indicate use of the operational (a priori) definition and open shapes indicate use of the functional definition (>34 days) for long shedders. Black lines show the mean and standard distribution for each scenario.

Long-term norovirus shedding
Fig. 4. Sensitivity of outbreak probability to variable long-shedding duration (1/γ_L). Outbreaks are defined by having >200 cases. Dashed lines show the probability for each definition in the absence of long shedders, and dotted lines represent the shedding duration of the regular-shedding groups (γ = 16.4 days, γ = 14.5 days). Open circles represent use of the operational definition, and filled grey circles represent use of the functional definition for stratification.

Fig. 5. Sensitivity of outbreak severity to variable long-shedding duration (1/γ_L). Outbreaks are defined by having >200 cases. Dashed line shows average and standard definition for 1000 runs with the functional definition (solid line) shows average and standard deviation for operational definition. Dotted lines show average number of cases in the absence of long shedders.
The maximum expected duration of about 2000 days (about 5.5 years) is reached when $1/\gamma L = 200$ days for both definitions. Again, this is much longer than the average duration in the absence of long shedders (633 days operational, 756 days functional).

**DISCUSSION**

Heterogeneity in viral shedding duration has clear implications for norovirus transmission dynamics, and should be included when estimating disease risks. In the case of norovirus, estimates of short average shedding duration do not capture the potential effects of long-term shedding individuals in sustaining community transmission. Our empirically derived characterization of norovirus shedding duration allows us to model its impact on population-level norovirus risk. Our analysis indicates that inclusion of long shedders results in a 50–80% increase in the basic reproduction number $R_0$, about a 20% increase in the number of cases, and about a 100% increase in the duration that transmission is sustained in a population.

The importance of an increase in $R_0$ depends on its original value, as the number of excess cases that result from an increase in $R_0$ is highly dependent on the transmission rate, $\beta$. For example, if $R_0$ is already quite high (such as within a home), then an outbreak would already be likely to spread through a population and a twofold increase would not have much effect. If $R_0 < 1$ (probably true for community transmission), then this same increase can mean the difference of an outbreak dying out or of it taking off in a population (see [44, fig. 1] for further explanation). Long shedders are likely, therefore, to be more important for community transmission where they can potentially impact the probability of outbreak occurrence.

Since long shedders tend to be asymptomatic, estimating the transmission rate due to asymptomatic shedders is important. If asymptomatic individuals do not transmit norovirus then long shedders cannot play a role in norovirus transmission. However, evidence and documentation that asymptomatic individuals can be the index case of outbreaks suggest that asymptomatic transmission does occur and is probably enhanced by long shedders [6, 7].

Any effect of long shedders would become more profound with the addition of more complicated and realistic structure, including social structure and effect of varied levels of infectiousness. In a structured metapopulation comprised of loosely connected groups,
a pathogen must persist within its initial group long enough to allow for migration into another one. This type of long-shedder-driven persistence may provide an explanation for the explosive but episodic character of norovirus outbreaks: long shedders may allow the virus to circulate at low levels in the population until it either reaches an individual with a high contact rate (e.g. a food handler or symptomatic individual in a closed scenario), or a pocket of susceptible individuals (e.g. an unexposed school), resulting in a large new outbreak. A socially structured model could also account for other transmission processes, including clustering of long shedders in healthcare facilities and other high-risk zones.

The effect of long shedders may also become more profound if we relax the assumption of homotypic immunity. Immunity has been estimated to last 14 weeks [45], but this is a highly debated topic [46]. When the duration of shedding (1/γ_L) is greater than the duration of immunity, the long-shedding group can act as a reservoir, allowing the infection to persist in a population and potentially re-seed periodic epidemics, in a manner similar to herpes viruses and other pathogens with long-term carriage, such as Salmonella typhi.

Because host immune pressure can impact calicivirus RNA evolution [38], long shedders may also impact the population ecology of noroviruses [47]. Since the amino acids that mutate most frequently are those involved in immune evasion [48, 49], escape mutants may emerge over the life of a single long-shedding infection. In fact, the number of amino-acid mutations arising over the course of a year within-host is similar to the number distinguishing outbreak variants from each other [49]. Consequently, long shedders represent a potential mechanism for introducing novel variants into populations that have achieved herd immunity against the most recent circulating strains.

Our parameter estimates are sensitive to differences in technique and study design across studies (Table 1), including differences in faecal viral concentration, specimen storage, RNA extraction efficiency, presence of faecal reverse transcriptase inhibitors, and primer usage [50]. Many studies terminated sample collection while patients were still shedding [8, 38, 51], resulting in right-censored data, and systematic underestimates of actual shedding durations. Publication bias exists in the literature, so that published studies may not reflect the true distribution in the population. Future studies, with increasingly sensitive techniques, will provide more accurate estimates of actual post-infection viral shedding duration.

CONCLUSIONS
We have demonstrated the importance of including an empirically validated representation of between-host heterogeneity in norovirus natural history when assessing population-level transmission risks. These results should be incorporated in models that include other potentially important sources of transmission heterogeneity such heterogeneous contact networks, protective immunity, and strain competition. This will facilitate the development of policies and interventions that can target the individuals who are both the most susceptible and the most likely to transmit disease. As we deepen our understanding of the different types, degrees, and interactions of heterogeneity in disease transmission systems, we can make more informed policy decisions and recommendations and more effectively protect human health.

SUPPLEMENTARY MATERIAL
For supplementary material accompanying this paper visit http://dx.doi.org/10.1093/iod/iod009.

ACKNOWLEDGEMENTS
This work was funded under the U.S. Environmental Protection Agency Science to Achieve Results (STAR) grant no. RD83172701 and by the U.S. Department of Homeland Security University Programs grant no. R83236201.

DECLARATION OF INTEREST
None.

REFERENCES


