Modelling norovirus transmission and vaccination

K.A.M. Gaythorpe\textsuperscript{a,\*}, C.L. Trotter\textsuperscript{b}, A.J.K. Conlan\textsuperscript{b}

\textsuperscript{a}Department of Medicine, Imperial College London, United Kingdom
\textsuperscript{b}Department of Veterinary Medicine, University of Cambridge, United Kingdom

\textbf{A R T I C L E   I N F O}

Article history:
Received 26 March 2018
Received in revised form 20 July 2018
Accepted 22 July 2018
Available online 31 July 2018

Keywords:
Norovirus
Mathematical model
ODEs
Markov model
Vaccine

\textbf{A B S T R A C T}

\textit{Background:} Norovirus is thought to be responsible for a fifth of all acute gastroenteritis cases globally each year. The population level transmission dynamics of this very common virus are still poorly understood, in part because illness is under-reported. With vaccines undergoing clinical trials, there is a growing need for appropriate, empirically grounded models, to predict the likely impact of vaccination.

\textit{Methods:} We developed a dynamic age-specific mathematical model of norovirus transmission and vaccination, informed by available data, particularly age-stratified time series case notification data. We introduce the use of a self-reporting Markov model to account for variation by age and over time in the statutory reporting of norovirus in Germany. We estimated the model using a sequential Monte Carlo particle filter. We then extended and applied our estimated model to investigate the potential impact of a range of immunisation strategies. We performed sensitivity analyses on the mode of vaccine action and other vaccine-related parameters.

\textit{Results:} We find that routine immunisation could reduce the incidence of norovirus by up to 70.5\% even when those vaccines do not provide complete protection from disease. Furthermore, we find that the relative efficiency of alternative strategies targeting different age groups are dependant on the outcome we consider and are sensitive to assumptions on the mode of vaccine action. Strategies that target infants and toddler are more efficient in preventing infection but targeting older adults is preferable for preventing severe outcomes.

\textit{Conclusions:} Our model provides a robust estimate of a dynamic transmission model for norovirus at the population level. Vaccination may be an effective strategy in preventing disease but further work is required to ascertain norovirus vaccine efficacy, its mode of action and to estimate the cost-effectiveness of immunisation against norovirus.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

\textbf{1. Introduction}

Norovirus is the most common viral cause of gastroenteritis in humans, with one in five cases of acute gastroenteritis attributable to norovirus \cite{1,2}. The virus has been estimated to cause approximately 700 million illnesses and 219,000 deaths globally each year, at a cost of around $60 billion, mostly attributable to losses in productivity \cite{3}. However, estimates of the true burden of norovirus are uncertain because many individuals will not seek health care for what is usually a self-limiting illness. In the UK, only one in 218 cases of norovirus is thought to be notified \cite{4}.

Several vaccines that offer the potential to reduce the burden of illness due to norovirus are in development, with the most advanced product from Takeda Pharmaceuticals in the proof of concept stage \cite{5}. To date only one modelling study, informed by US data, has explored the potential impact of such vaccines \cite{6}. A wider range of modelling approaches and estimation techniques have been applied to norovirus at the outbreak level \cite{7}. However, by their nature outbreak settings are highly diverse and context specific. As a consequence there is disparity between model structures and key parameter estimates at the outbreak level. In contrast, at the population level most studies have utilised the Infectious Intestinal Disease survey data from the UK \cite{7}. The discrepancies between models are well illustrated by the variation in the basic reproduction number, $R_0$ between modelling studies. While population level models have tended to use similar data and approaches, with $R_0$ in the region of 1.6 \cite{8}, outbreak reproduction numbers have been found to be as high as 14 \cite{9}. These differences may be real, if the outbreak represents an exceptional situation or subset of the population. However, the low mean age at first infection implied by serological data suggests that the population level $R_0$ may be greater than 10 \cite{10–13}. Under-estimating $R_0$ has important implications for transmission, with $R_0$...
implications for control strategies; with a higher $R_0$ resulting in a higher critical vaccination threshold for elimination.

In this paper, we leverage population level data sources, particularly case notification data from Germany to estimate a novel population level model of norovirus transmission. We use our robustly estimated model to investigate a spectrum of vaccine strategies and effects given the uncertainty in both the application of a vaccine in the population and the mode of action of a successful vaccine candidate.

2. Materials and methods

2.1. Demographic model

For convenience we assume the population size remains constant, ignoring migration and population growth. We assume that life-expectancy follows a Gompertz model, with death rates for each age group estimated from the United Nations population data for 2015 [14].

2.2. Epidemiological model

We adapt a deterministic Susceptible, Exposed, Infectious, Recovered (SEIR) model to include individuals protected by maternal antibodies, an approximately gamma-distributed latent period and asymptomatic individuals, as shown in Fig. 1.

The model tracks annual cohorts up to the age of 70 years, with a continuous ageing model. Model output and transmission rates are aggregated into seven age groups motivated by visual inspection of the case notification data and for younger ages key groups with the German school system; 0–8, 8–18, 18–26, 26–37, 37–50, 50–70 and 70+. Individuals protected by maternal antibodies are accounted for by a Maternal (M) antibody class which individuals leave at rate $\alpha$, fixed at a value of $2.5 \times 10^{-3}$, as estimated in [10].

Individuals are infected at a rate ($\mu_i$), with the relative transmission rates between age-groups determined by a POLYMOD mixing matrix. The POLYMOD survey, conducted in eight European countries, provides social contact information stratified by age, gender, contact type and duration [15]. We use the POLYMOD data from Germany and methods employed by Eames et al., Baguelin et al. and Meyer et al. to arrive at a contact mixing matrix which is symmetrised and weighted for weekdays and weekends [16–18]. See the supplementary information for further details.

We assume that asymptomatic individuals contribute to the force of infection at a reduced level, $\nu$, derived through the inference process. As such, the force of infection for age group $i$ is as follows:

\[ \lambda_i = qZ \sum_j c_{ij} (I_j + \nu A_j + \zeta (I_j^V + \nu A_j^V)) \]

where $q$ is a transmission rate, $c_{ij}$ is the contact matrix and $Z$ is a seasonal component depending on a seasonal amplitude, $\omega_n$, and a seasonal offset term, $\omega_2: Z = 1 + \omega_1 \cos \left(\frac{2 \pi t}{364} + \omega_2\right)$. We fix the seasonal offset term, $\omega_2$, for each age class according to where the mean seasonal peak occurs. The offset values can be found in the supplementary material and illustrate the staggered nature of the seasonality of norovirus where incidence in children peaks first, followed by the older age groups. We also include the vaccinated classes, denoted with superscript $V$, which contribute to the force of infection at reduced rate scaled by $\zeta$. Once an individual is infected, they enter a gamma distributed latent period approximated by the inclusion of two ‘E’ classes. The rate individuals exit the latent classes is fixed at 1 leading to an average latent period of $2/1$ days, consistent with [19]. Once individuals exit the latent class, only a proportion $\sigma$ will be symptomatic. We also fix the value of $\sigma$ to be 0.75, as estimated from challenge studies mentioned in [20], see supplementary material. Infected individuals lose their symptoms at rate $\psi$, fixed at 0.5 thus giving an expected duration of symptoms of two days. Asymptomatic individuals continue to shed virus and are therefore considered infectious. Once individuals recover, R, they are protected from infection for a period of time; this protection wanes at rate $\delta$ (which is estimated from the inference process) and individuals become fully susceptible to infection again.

2.3. Norovirus notification data

The Robert Koch Institute collects statutory notifiable disease data from across Germany in the database SurvStat [21]. We estimate our model from the reported case notifications of norovirus from 2008 to 2016. This anonymised data is freely available and stratified by age, gender and region. There are known biases and systematic changes in reporting of norovirus within the German notification system that need to be appropriately modelled [22]. Under-reporting of norovirus infections is expected to be substantial with healthcare seeking behaviour for individuals with symptoms likely to vary with age [17]. Finally, there was a change in case definition during our study period. Before 2011 an epidemiological diagnosis of norovirus was sufficient for a case to be reported in SurvStat, whereas after 2011 only laboratory confirmed cases are reported. We use a novel self-correcting markov process model to account for this change in notification. Full details of the observation model can be found in supplementary information.

When estimating our models, we simulate our deterministic epidemiological model over the study period (2008–2016) and calculate the expected incidence for each weekly period as the product of the accumulated new infections ($C(t)$) multiplied by the time, age and case dependant probability of reporting. If we treat each of our M reported case notifications, per week and age class, as an independent binomial trial we can construct an approximate likelihood for our model as the product of these M trials. To estimate potential over-dispersion of the data due to model misspecification and extra-demographic stochasticity, we use a
negative-binomial likelihood with the expected value set to $C(t)\xi(t,a,C)$. Details of the statistical inference are given in the supplementary information.

2.4. Vaccination model and scenarios

We duplicate the disease transmission model for individuals who are vaccinated, but do not include the maternal protection compartment (assuming that vaccination will move those in the maternal class to recovered-but-vaccinated). We examine a variety of ways the vaccine may be deployed in the population, the effect at the population level and the modes of action at the individual level. We examine two situations upon vaccinating an individual: the first is that the vaccine does not clear infection and individuals are moved to their equivalent, vaccinated class, for example, asymptomatic individuals are moved to asymptomatic-but-vaccinated. The second situation is where the vaccine stimulates clearance of infection; in this situation all vaccinated individuals enter the vaccinated-but-removed class.

Vaccinated individuals may also experience different disease progression. All transitions in Fig. 1 highlighted in red may be adapted for vaccinated individuals. We introduce a scaling term, $\nu$, applied to the force of infection, which accounts for differing susceptibility for vaccinated individuals. We may also adapt the proportion of vaccinated individuals who become symptomatic through changing $\sigma$. Finally, we assume vaccinated individuals are less infectious by adjusting the contribution of vaccinated individuals to the force of infection by $\xi$, see Table 1 and Table A1 in the supplementary material for full definitions.

Fig. 2 shows the different components of the vaccination scenarios. We examine 6 age-dependent vaccination strategies which target respectively: infants, 1 year olds, 1 year olds and those fifty years and older, 1 year olds and the elderly over 65 years at five yearly intervals, those seventy and older at annual intervals and, finally, the elderly over 65 years at five yearly intervals. The effective coverage in each strategy is divided by age where vaccine uptake is 90% in children aged 0 or 1 and vaccine has 50% efficacy. In older individuals vaccine uptake is assumed to be 50% and vaccine has 70% efficacy. The duration of vaccine effect is fixed at an average value of five years for all ages.

Finally, the mode of vaccine action was considered. The symptoms, infectiousness and susceptibility describe the reduction in the proportion of vaccinated individuals becoming symptomatic, their contribution to the force of infection and their experience of the force of infection respectively.

We examine a representative sample of the possible permutations of vaccine strategy and mode of action by using a Latin Hypercube Sample of the vaccine scenario space. This involves taking a representative spanning set of samples to approximate the space as a whole. It should indicate the sensitivity of the campaign outcomes to each component of the vaccine scenarios.

We report four different outcomes, infections, hospitalisations, outpatient visits, deaths, and report our results in terms of outcomes averted per vaccine dose given.

### Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>Death rate per age group</td>
<td>$3.2 \times 10^{-5} \exp(0.09a)$</td>
</tr>
<tr>
<td>$q$</td>
<td>Transmissibility</td>
<td>1.17 $[1.01,1.39]$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Rate recovered individuals become fully susceptible</td>
<td>$3 \times 10^{-4}$ $[2.1 \times 10^{-4}, 4.5 \times 10^{-4}]$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Proportion of new infections that are asymptomatic</td>
<td>0.75 $[0.53,0.90]$</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Scaling of asymptomatic infectiousness</td>
<td>$5.72 \times 10^{-3}$ $[5.33 \times 10^{-3}, 5.98 \times 10^{-3}]$</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>Amplitude of seasonal forcing</td>
<td>$9.42 \times 10^{-2}$ $[8.92 \times 10^{-2}, 9.84 \times 10^{-2}]$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Recovery rate of asymptomatic individuals related to the duration of symptomatic infection</td>
<td>$7.15 \times 10^{-2}$ $[6.69 \times 10^{-2}, 7.48 \times 10^{-2}]$</td>
</tr>
<tr>
<td>$\delta(t)$</td>
<td>Reporting rate baseline for age group $a$ ($&lt;10^{-5}$)</td>
<td>0.69 $[0.65,0.72]$, 1.92 $[1.80,2.08] \times \delta(1)$, 2.03 $[1.91,2.25] \times \delta(1)$, 4.74 $[4.12,5.14] \times \delta(1)$, 0.47 $[0.44,0.53]$</td>
</tr>
<tr>
<td>$\epsilon_j$</td>
<td>Self correcting Markov process term for, $j=1$, 0–37 year olds or, $j=2$, 37+ year olds</td>
<td>$2.25 \times 10^{-2}$ $[1.94 \times 10^{-2}, 2.54 \times 10^{-2}]$, $9.57 \times 10^{-6}$ $[7.39 \times 10^{-6}, 14.40 \times 10^{-6}]$</td>
</tr>
</tbody>
</table>

3. Results

3.1. Model fit

The estimated model provides a close qualitative and quantitative fit to the data, in particular tracking the changes in the age-distribution of cases before and after the change in case notification in 2011 (Fig. 3). We find that reporting rates are lowest in young adults, consistent with published estimates [17], and that the incidence-based saturation of reporting occurs at a reduced rate for individuals under 37 years old. In children and the elderly we see both the largest numbers of cases and largest baseline reporting rates with a median of 1 in every 52.2 [14.5,75.6] cases being reported to SurvStat, higher than a similar estimate for the UK [23].

We estimate the amplitude of seasonal forcing to be approximately 9.4% [8.9%, 9.8%]. Additionally we found the duration of asymptomatic infection to be approximately 28 days (ranging between 26 and 30 days). We estimate $R_0$ in the absence of seasonality to be 4.60 [4.59,4.62] and the contribution of asymptomatic
Fig. 3. Observed and posterior predicted reported incidence. Case notification data from SurvStat are shown in solid black lines. The blue lines denote individual realisations from the estimated model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Outcomes averted per dose across all age groups given 1000 permutations of vaccine characteristics and strategy with 100 posterior samples. We plot the distribution of outcomes averted for each characteristic where the horizontal line denotes median value and thick horizontal line denotes the 25th, $q_{25}$, to 75th, $q_{75}$, quantile of the data. Outliers are shown with red crosses and the whiskers range between $q_{25} - 1.5 \times (q_{75} - q_{25})$ and $q_{75} + 1.5 \times (q_{75} - q_{25})$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
individuals to the force of infection to be quite low, approximately 0.5% of symptomatic individuals, see Table 1 for full estimates and Section 7 in the supplementary material for a discussion on the impact of seasonality on $R_0$.

### 3.2. Vaccine impact

Fig. 4 illustrates the effect of different vaccination strategies given 1000 permutations of the characteristics shown in Fig. 2 and 100 samples from the posterior distribution.

The outcomes averted per dose are shown for each of our four outcomes (total infections, hospitalisations, outpatient visits, deaths) and for each vaccine strategy in Fig. 2 and Table 2. Strategies that target infants are most efficient (i.e. more cases are prevented per dose) for less severe outcomes, including total infections and outpatient visits. For more severe outcomes, including hospitalisation and death, strategies that directly target older age groups, who are at highest risk of these severe outcomes, are more efficient. Clearly for less common outcomes, the number of cases per dose is very small; put a different way, the number needed to vaccinate to prevent a death is much higher than the number needed to vaccinate to prevent an infection. The spread of results for each vaccine strategy is however rather wide with overlapping ranges.

#### Table 2

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Doses</th>
<th>Outcomes averted per million doses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Infections</td>
</tr>
<tr>
<td>A0</td>
<td>0.617</td>
<td>274173.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[187870.3]</td>
</tr>
<tr>
<td>A1</td>
<td>0.617</td>
<td>294394.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[152674.6]</td>
</tr>
<tr>
<td>A2</td>
<td>9.861</td>
<td>266594.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[117804.3]</td>
</tr>
<tr>
<td>A3</td>
<td>4.897</td>
<td>187562.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[73362.2]</td>
</tr>
<tr>
<td>A4</td>
<td>5.070</td>
<td>161394.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[50053.3]</td>
</tr>
<tr>
<td>A5</td>
<td>4.280</td>
<td>179604.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[50087.8]</td>
</tr>
</tbody>
</table>

Fig. 5. Tornado plot for each strategy displaying the relative change in infections averted per dose given a 10% change in each vaccine characteristic. The baseline scenario is 50% clearance, 50% susceptibility, 50% infectiousness and 50% symptoms for vaccinated individuals compared to un-vaccinated individuals.
In terms of the overall burden of infection, the infant strategy could reduce the total number of infections by 4.53% [3.10%, 5.29%] overall whereas the strategy targeting 70+ year olds (A4), which targets a much greater number of individuals, could reduce the total number of infections by 21.93% [6.79%, 26.05%]. The highest percentage reduction in infections, 70.5% [31.1%, 77.0%], comes as a result of strategy A2 which targets 1 and 50+ year olds, see table A2 in the supplementary material for further values.

We also examined the sensitivity of our results with respect to infections averted per dose, these are shown in Fig. 5.

We find that the most influential vaccine characteristics are clearance of infection and susceptibility. Reductions in infectiousness and symptoms are also influential but to a lesser extent suggesting that the primary benefits of vaccination would be to the individual rather than through herd effects. To examine whether a change in vaccine action between age groups could affect the relative benefits of the age-based strategies, we compared the cases averted per dose for strategy A0, targeting babies, with those of A4, targeting individuals aged over 70 years with and without clearance of infection, see Fig. 6. These represent the two most differing strategies with the most extreme change in vaccine action and we can see that the gap between the A0 and A4 strategies is narrowed by this difference in vaccine action. However, it is unlikely that there will be this degree in variation in vaccine action between age groups.

4. Discussion

We developed a population level model of norovirus transmission dynamics and vaccination calibrated using multiple sources of data that allowed us to deal with uncertainties in both disease dynamics and vaccine action. We examined the potential impact of six different age-based vaccination strategies at population level, each focusing on children, the elderly or both. Our results suggest that strategies targeting infants for vaccination are likely to be more effective at averting infections per dose than targeting the elderly, given the mode of action of a vaccine is consistent across age groups. For example, strategy A5, that targets older adults, only averts 57% of the cases that A0 (infant immunisation) averts (Table 2). As shown with other infectious diseases, such as influenza [16], the greater degree of contact among children means that vaccines targeting children can generate more profound herd effects than strategies targeting the elderly (Fig. A9). Yet, for more severe outcomes (hospitalization or death), the elderly become a more effective target, due to their higher morbidity and mortality [3]. Depending on the aim of the vaccination programme, it may be more beneficial to target those most at risk from severe disease than those most at risk from infection. A cost-effectiveness analysis would be a useful extension of this model; however it will be important to review the models once further data has been realised from clinical trials, particularly concerning vaccine characteristics and any adverse effects.

We examined the sensitivity of our results to vaccine characteristics. A vaccine with potential to clear infection is likely to be more effective than a vaccine that does not clear infection as this reduces the effective reproduction number of norovirus. Similarly, a vaccine that directly reduces the susceptibility of vaccinated individuals to new infections performs better in almost all cases due to the reduction in the overall effective reproduction number. Less influential are reductions in the infectiousness and symptoms of vaccinated individuals, which would protect individuals in contact with a vaccinated individual rather than the individual themselves. This indicates that, given the partial protection attained from vaccination, the greater benefits are to be found through direct rather indirect effects (Fig. A9). Although potentially biologically unlikely, we also show that differences in mode of action of the vaccine by age group may influence the relative effectiveness of the age-based vaccination strategies.

A study based in the US previously modelled the potential impact of norovirus vaccination [6]. Steele et al. also found infants an
effective target for vaccination in the population of the USA; however with a greater magnitude of effect. The similarities between our model and that of Steele et al., result from similar assumptions on population structure; with no comparable study of social contacts in the USA, Steele et al. also used the POLYMOD study data. Therefore the relative probability of infection between age groups will be similar. The difference in magnitude of vaccine effect is driven by the assumptions on vaccine action. We have examined an extremely broad range of vaccine action, from small reductions in susceptibility to complete symptom loss, see Fig. 2, whereas, Steele et al. assumed that vaccination removed symptoms only.

The fit of our model (Fig. 3) suggests that we are reasonably reflecting the temporal and age-dependant dynamics of the reporting structure and underlying epidemiological processes. Our estimates for the transmissibility and duration of immunity arose from inference based on serological data and suggest a duration of immunity in the region of eight years, slightly higher than that of Simmons et al. [8]. However, we have assumed that the serological status of the population in Germany will be similar to that of Italy. We estimate $R_0$ for the mean value of the transmission parameter to be around 4.6, higher than that commonly estimated for norovirus using population level case notification data [7], yet lower than that exclusively estimated from serological data [10]. This suggests there are outstanding inconsistencies between the data and modelling assumptions that need to be resolved. We found that the duration of immunity was poorly identifiable from case notification data alone as it correlated with other key parameters. In our model, we assumed that individuals attain complete immunity upon recovery from infection, but it has also been suggested that recovered individuals may be susceptible to asymptomatic reinfection [8]. This uncertainty surrounding the nature and duration of immunity to norovirus has ramifications for the effect of vaccine strategies, and reducing this uncertainty through further research would be beneficial [20]. We have not considered adverse outcomes of the vaccine as none have currently been highlighted; however as vaccine trials progress and health economic modelling may be conducted, adverse effects may need to be incorporated. Further research into contact patterns of the elderly (particularly those groups, such as care home residents who are vulnerable to norovirus) is also warranted, given that relatively few elderly individuals participated in POLYMOD.

This work builds upon previous transmission models of norovirus by integrating different data sources and generating rigorous estimates. We highlight the potential impact of population-level norovirus vaccine and compare the efficiency of different vaccine strategies by age. We also provide insights into the sensitivity of our results to model parameters, including the mode of vaccine action, which allows us to highlight areas where further research is required. Targeting children under 2 years is more effective than vaccinating the elderly in terms of reducing the number of infec-

tions across the population, however, directly targeting the elderly will prevent more severe outcomes.

Financial support

This work was supported by Takeda Pharmaceuticals to the University of Cambridge (Grant No. RG 77788).

Conflict of interest

KAMG, CLT and AC received salary support through the grant awarded from Takeda to the University of Cambridge.

Acknowledgements

This study was funded by a grant from Takeda Pharmaceuticals to the University of Cambridge. We thank John Weil, Robert Bargetze, Gerhart Knerer and Elaine Gallagher from Takeda for stimulating and helpful discussions through the project.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.vaccine.2018.07.053.