
Transmission routes of rare seasonal diseases: the case of norovirus infections

Rushton, Stephen P.¹, Sanderson, Roy A.^{1*}, Reid, William D. K.², Shirley, Mark D.F.¹, Harris, John P.^{3,4}, Hunter, Paul R.^{4,5}
and O'Brien, Sarah J.^{1,2,4}

¹*Modelling, Evidence & Policy Research Group, School of Natural and Environmental Science, Newcastle University, Newcastle upon Tyne NE1 7RU, UK*

²*Ecology Research Group, School of Natural and Environmental Science, Newcastle University, Newcastle upon Tyne NE1 7RU, UK*

³*Public Health and Policy, University of Liverpool, L69 3GL, UK*

⁴*National Institute for Health Research, Health Protection Research Unit in Gastrointestinal Infections, Liverpool, L69 3GL, UK*

⁵*Norwich Medical School, University of East Anglia, Norwich 33 NR4 7TJ, UK*

Keywords: coupled dynamic social networks, community infection, norovirus, stochastic, fomites, food

Abstract

Norovirus (NoV) is the most commonly recognised cause of acute gastroenteritis, with over a million cases globally per year. Whilst usually self-limiting, NoV poses a substantial economic burden because it is highly contagious and there are multiple transmission routes. Infection occurs through inhalation of vomitus; faecal-oral spread; food, water and environmental contamination. Whilst the incidence of disease is predictably seasonal, much less is known about the relative contribution of the various exposure pathways in causing disease. Additionally asymptomatic excretion and viral shedding make forecasting disease burden difficult. We develop a novel stochastic dynamic network model to investigate the contributions of different transmission pathways in multiple coupled social networks representing schools, hospitals, care-homes, and family households in a community setting. We analyse how the networks impact on transmission. We used ward-level demographic data from Northumberland, UK to create a simulation cohort. We compared the results with extant data on NoV cases from the IID2 study. Connectivity across the simulated cohort was high. Cases of NoV showed marked seasonality, peaking in early winter and declining through the summer. For the first time we show that fomites and food appear to be the most important exposure routes in determining the population burden of disease.

*Author for correspondence (roy.sanderson@newcastle.ac.uk). ORCID ID 0000-0002-9580-4751

Introduction

Norovirus (NoV), (formally known as Norwalk virus, Norwalk-like virus or small round-structured virus), is a comparatively recently-identified calcivirus originally discovered following an outbreak of acute gastroenteritis in Norwalk, Ohio[1]. Norovirus is the most common cause of acute gastroenteritis worldwide[2]. The disease is usually self-limiting in the immunocompetent but poses particular issues for semi-closed social settings where there is population mixing and potential for close contact between infectious individuals[3], contaminated environments and susceptible individuals. The virus may contaminate food[4] or water[5] causing infection through the faecal-oral route. Symptomatic individuals generate millions of viral particles which are aerosolised in vomitus and faeces thus contaminating the environment[6]; the dose required for infection has been estimated to be <18 particles[7]; the virus may persist on surfaces in the environment for up to 50 days after deposition[8] and individuals may shed virus for several weeks after recovery [9]. Furthermore, up to 16% of the population may be asymptomatic excretors[10]. These features pose particular problems for human environments where social mixing takes place. Outbreaks often lead to closures of health care, care-home and educational establishments, posing further burdens for health and social care. The annual costs of norovirus infection in the UK are estimated to be £80M[11].

Whilst the pattern of disease is strongly seasonal in temperate countries[12], predicting the burden and incidence of disease in any particular place is difficult for several reasons. Firstly, the incidence of disease is comparatively rare (47 community cases per 1000 person-years)[13]. Secondly, as the disease is self-limiting it is under-recorded[14] as many cases do not seek medical care and it has been estimated that only around 1 in 300 cases are recorded in official statistics[13]. The problem of predicting disease is also further complicated by hyper-mutation leading to rapid strain evolution[15].

Forecasting norovirus disease is difficult unless due attention is paid to the different modes of transmission. Whilst faecal-oral, droplet-oral and vomitus-oral, and transmission via fomites are known pathways for infection it is not clear which of these mechanisms is responsible for spread and persistence in the wider community[16]. Outbreaks are recorded in UK public health systems, but these are usually associated with institutional settings like schools, hospitals and care-homes for the elderly[17] where case ascertainment is straightforward. Whilst outbreaks in these institutions can be expensive and have high local impact these systems usually have comparatively small populations at risk. It is essential to quantify the relative importance of transmission pathways and sources of infection at the wider community-level so that risks of outbreaks at the institutional level can be minimised at a higher level[18]. Conventional deterministic SEIR models are not ideal to investigate disease spread in these scenarios as transmission events are unlikely to be described adequately using these models. Here we develop and validate a microsimulation model of norovirus infections which investigates transmission pathways in a community in the context of both the social networks within it, and also the underlying host-pathogen interaction represented by the immune response, recovery, asymptomatic shedding and environmental contamination. Networks associated with being in a family, attending education, attending hospital and care-homes are linked dynamically through time in relation to the working week and community demographics. The model identifies the relative significance of different mechanisms of exposure and transmission in the social setting of a digital representation of the population of a voting ward in NE England. We link exposure, transmission in relation to dose-response and changes in immunity in individuals in the population following exposure. Model results are compared to observations of community-level norovirus infections from a prospective, population-based cohort study (IID2, January 2009- September 2011)[13] and also with norovirus incidence from public health records for NE England.

Methods

The model consisted of a combined microsimulation and individual-based model for predicting the temporal dynamics of norovirus spread amongst individuals in a spatially-defined community on the basis of different contact mechanisms.

The model has two components (Figure 1):

1. Community simulator to create modelled cohort
2. Disease dynamics simulator within-cohort disease spread

1. Community simulator

The community simulator was a microsimulation model that was used to define a population and the network of interactions between individuals in terms of 4 types of social network: family, schools, hospitals and care-homes. An individual belonged to one or more networks depending on their age (primary/secondary-school, care-home) or health (hospital). Disease spread occurred through daily contacts between individuals as they came into contact through the

relevant social network. The simulator creates a population with defined socio-economic structure at the level of the Census Ward. Since census wards typically have populations of 4000-6000 a model cohort of 5000 people was implemented. Data describing population age, gender, socio-economic status, household composition and family size were used as parameters to define a simulated-cohort. A cohort was created to reflect a representative ward, with individual age, gender, family, school, hospital and care-home residency allocated stochastically. Input data for model were obtained from Northumberland County Council which characterise the socio-demographics of the population in each ward in Northumberland[19]. The simulator assumed that individuals lived in households/families that exist as single-person households, single parents with children, couples, couples with children. The proportion of the modelled population in each category was determined on the basis of the known age distribution amongst known household structure in the ward. Children above the age of 5 were assumed to attend a primary-school; those > 11 a secondary-school and all children attended school until age 18, for 5 days a week. Individuals in a ward were assumed to have access to one or more hospitals and care-homes. Both the number of schools and hospitals were defined as model inputs, and we assumed that there were 4 primary- and one secondary-schools. Hospital attendance was a stochastic variable determined on the basis of admission rates to hospitals. Patients were assumed to remain in hospital for 7 days. Care-home residency was modelled on the basis of the known proportion of the over 65 population known to occupy care-homes. The network structure of the model cohort was then analysed in terms of the connectivity of individuals within the population as determined by the social networks in which they participated (specifically schools and families). The population ages over time in the disease dynamics model, with individuals dying after 80 years; babies are born during the simulation to replace deaths.. Data were analysed using the igraph and network packages[20].

2. Disease dynamics

Disease dynamics was modelled at the individual level. Disease transmission was assumed to occur when susceptible individuals consumed contaminated food and/or came into contact with infected individuals, or environments in which infected individuals had been ill, in four social settings. These were in: the home (within-family exposure); primary-school (within-school exposure); secondary-school (within school-exposure); hospitals (within-hospital exposure) and care-homes (within-care-home exposure). The disease model operated on a daily time-step by simulating weeks with 5-days school attendance and 2-day weekends for each individual in each network as appropriate (Figure 1). Each social setting was effectively a sub-section of the total population of the cohort which changed with the day. Individuals attending any of the settings came into contact with other individuals in that setting (family, primary-schools, secondary-schools, hospitals and care-homes) who may or may not have been infectious/carriers of norovirus (Figure 1). In the case of human:human transmission, the probability of a susceptible individual becoming infected on any day was then modelled as a Poisson process similar to that of the Kermack-McKendrick SEIR model, where individual infection is calculated as the product of a transmission coefficient (representing the number of individuals that become infected by contact with one infected individual) and the number of infected individuals currently in that group in that setting at that time point. Individuals may also come into contact with residual viral particles left in the environment of the premises following illness having occurred there. The viability of these viral particles following deposition from an infected individual was assumed to decline following a Weibull distribution[21, 22]. In both cases the dose of viral particles that a susceptible individual received on contact with an infectious individual or contaminated environment was modelled stochastically. Individuals were also exposed to infection through consumption of contaminated oysters. Whilst this may technically be considered an environmental source of contamination, in practice the contamination was spatially external to the community. Presence of an infectious individual in a social setting was assumed to take precedence over a contaminated environment in causing disease. The probability that an individual became ill after exposure was then determined on the basis of the received dose of viral particles[23] and the estimated immune status of the individual. Immune status was modelled as an exponential decline in immunity from the time of last exposure. The proportional decline in immunity in each time step was estimated from the value of the exponent needed to reduce the estimated immunity from complete (1) to a low level (less than 10^{-20}) by the maximum duration of immunity. The duration of the infectious state and the rate of decline of immune status were variable as model inputs.

The model was constructed in the R programming language using custom scripts[24].

Sensitivity analysis to quantify impacts of transmission mechanism, food and environmental exposure and immunity on disease spread

We varied four groups of epidemiological parameters. Firstly, the risk of transmission under the different social settings of family, primary/secondary-school, care-home and hospital. Secondly, the risk from environmental contamination and the level of contamination arising from the presence of sick individuals in each social setting. Thirdly, the extent to which members of the cohort consumed contaminated oysters, ignoring age, and finally, the duration of infectious period in which individuals could spread disease and the duration of immunity following becoming colonised/infected.

We used Latin Hypercube Sampling (LHS)[25] to create ranges for these parameters and ran the model for 3 years for 40 runs with different input parameters. See Table 1 for details of the range of inputs used in the model. Where we had no data we used plausible ranges. LHS provides a robust method for sensitivity analysis where, as in our study, there was

considerable uncertainty as to the true parameter values[see section 3.1 of Shirley et al.26]. The total number of norovirus cases arising from each transmission pathway each day was collated from the outputs. Since the predicted number of cases was highly periodic we detrended the case data in relation to time using harmonic regression. We then used the intercept from the harmonic regressions as an indicator of baseline disease burden after adjusting for seasonality under each LHS scenario, We used each the intercept as the dependent variable and the epidemiological parameters as independent variables in partial least squares regression to calculate the relative contribution of the covariate to the baseline disease burden. We calculated Variable Influence on Projection (VIP) factors to compare the relative contributions of the epidemiological parameters to total burden of disease and burden arising under each transmission setting (family school etc). VIP values greater than 1.0 have an above-average weight in explaining the dependent variable [27], whilst accounting for any collinearity between predictions.

Comparison of microsimulation model predictions with observed data

The model output daily counts of cases of norovirus disease by mode of contact, specifically: oysters, contact in family, schools, hospitals and care-homes and via environmental contamination in each setting. Here we summarised outputs in a number of ways to produce data that could be compared with the observed patterns of disease. Observed data consisted of monthly occurrence of norovirus outbreaks cases collated from the cohort study in the IID2 study[13] and from public health records of outbreaks in the UK over the period 2000-2010. Neither of these data sets provide a perfect cohort against which to compare model outputs. The model was run for the period May 2010 to June 2011 for comparison with data derived from the sampling period of the IID2 study and for three years to compare with the disease burden in NE England. Individuals were assigned age, gender and socio-economic class based on the socio-economic structure in NE England. We used the outputs for the runs of the sensitivity analysis as to predict cases for the whole community and compared these with the observed monthly number of cases recorded in the IID2 study and regional data using correlation.

Results

Disease network characteristics derived from community simulator

The social community model produced an initial population of 4969 individuals in 2638 families (households). The connectivity between members of the population was very high. When considering membership of schools and families alone, the mean distance on a weekday between any two members of the population was 3.3 links indicating high connectivity in this community. Whilst 1245 individuals (particularly single member households) were effectively isolated; 1300 individuals had only one link (suggesting childless couples) and 6 individuals had 481 connections to other members of the population. The mean number of connections (the degree) for individuals in the community was 55.7, but the degree distribution was highly bimodal, with one peak at the average size of a primary school (100) and another at the size of the single local secondary school (477). The four schools (3 primary, 1 secondary) are 'hubs' of dense connections, with every child at a school linked to every other child at that school.

Sensitivity analysis

We first assessed the extent to which the total burden of disease in the cohort was dependent on the disease parameters before investigating the extent to which burden arising from contact in the family, school, hospital, care-home and environmental settings was related to these parameters (Fig 2h). Environmental contamination risk, infrequent consumption of oysters and duration of illness had VIP greater than 1, indicating a significant contribution to baselines burden of disease for the community as a whole. At the level of individual transmission setting (Figure 2a-g) the transmission coefficient (for that setting was a major predictor of baseline disease for that setting (Figure 2a to Figure 2e, VIP values >1). In all these settings illness duration was also a particularly important factor. In care-homes there appeared to be particular risks from hospital and families, as well as within the care-home itself (Figure 2e). The major food risk was via infrequent consumption of oysters (Figure 2f).

Comparison of microsimulation model predictions with observed data

The predicted number of cases followed a cyclic pattern, with cases lowest in summer but rising to a peak in winter. Of the 40 sets of output from the LHS sensitivity analysis one set provided a close match between model predictions and observed number of cases for both data sets, with a correlation of 0.872 between predicted cases of disease and the observed number of IID2 cases. For these model scenarios transmission coefficients within primary-school, secondary-school, hospital, family and care-home networks were low (0.00140, 0.00465, 0.0036, 0.0011, 0.00329). The duration of illness (including asymptotic shedding) was 11 days environmental contamination risk was also low at 0.002, environmental particle dose was 1320 virions and the period of immunity was 253 days. Food and environmental contamination were most important exposure-pathways (Figure 2f and 2g). Predictions using the best fit parameters for IID2 data for a new data set based on estimated number of cases for the NE of England are shown as means +/- standard deviations for 10 replicates in Figure 3.

We classified predicted number of cases each month into 7 age classes: <6; 6-10; 11-15; 16-20; 21-40; 41-60 and >60. The distribution of cases amongst different age-groups was skewed. We used the total number of cases and individuals to calculate the expected number of cases in age class and the risk for each group. The ratio of predicted cases to expected cases was 1.06 for the under 6 age class; 1.75 for 6-10 year olds; 4.58 for 10-15 year olds; 2.68 for 15-20 year olds; 0.52 for 20-40 year olds; 0.49 for 40-60 year olds and 0.56 for >60 year olds; i.e. the burden of disease was predominantly amongst children. Three epidemics were identified in the output of the model; a new epidemic was assumed to have begun if a number of days had passed between cases that exceeded the generation time of the disease (11 days, estimated from the best LHS run). The first epidemic started on the first day of the model (Jan 1st) and lasted 465 days, ending in April of the following year. The R_0 of this epidemic was estimated to be 0.98 (95% CI = 0.97 - 0.99) by the exponential growth (EG) method and 1.07 (95% CI = 0.93 - 1.22) by the maximum likelihood (ML) method. The second epidemic began in September of the second year, lasted 269 days and ended in June of the third year. The estimates for R_0 were 1.01 (EG; 95% CI = 0.92 - 1.11) and 1.17 (ML; 95% CI = 0.93 - 1.45). The third epidemic began in September of the third year, lasted 89 days and ended in December of the third year. The estimates for R_0 were 1.25 (EG; 95% CI = 0.59 - 2.56) and 2.83 (ML; 95% CI = 1.77 - 4.23). All calculations of R_0 were performed using the 'R0' package [28].

Discussion

There have been many attempts to model the epidemiology of norovirus in the past. Much of this research has focussed on outbreak modelling, as outbreaks are most economically significant because they may lead to shut-down of institutions until disease is passed or controlled [29]. They provide more clearly-defined bouts of disease and form an easy basis with which to validate models. The results of outbreak modelling studies differ widely in their estimations of disease dynamics when measured in terms of R_0 : Gaythorpe et al [30] quotes model R_0 ranges of between just over 1 and 7. Variation in R_0 is known to have impacts on the utility of such models in outbreak settings of nosocomial disease [31]. This diversity probably reflects two features of the disease. First, populations are relatively small so exposure and transmission become highly stochastic rather than deterministic processes, leading to a wide variation in disease progression. Second, outbreak settings are unlikely to be environmentally or socially homogenous, with the role of fomites and social interaction varying between settings. Research on disease epidemiology in the community in contrast is logistically difficult because of the diversity of exposure and transmission pathways, and environments, in which transmission occurs. We have attempted to model at the community-level, including outbreaks in institutional settings, but capturing the whole population as a network of linked individuals as well as the institutional settings. Our results suggest that network linkages impacted on disease spread.

Theoretical models have demonstrated that epidemiological spread is highly dependent on network topology, as represented by linkages between members of the community at risk [e.g. 32, 33]. Much research has assumed that such networks remain static, or changes over timescales longer than that of the transmission and infectious period of disease. Spread in static networks is most rapid when the path length (i.e. shortest route through the network) is comparatively small as it was in our network. At a mechanistic level, spread is higher because of 'hubs' in the network, where some individuals have connections with many individuals and when contacts are more assortative, that is, where contacts are similar demographically. In our community networks, assortative features were a precondition because children were modelled as attending a restricted number of schools as well as families and hospital, whilst adults were modelled as belonging to families, hospitals and (eventually) care-homes. The number of connections to other nodes in the network for children would therefore have inevitably been higher than that of the adults. Furthermore, older people were more likely to be single without children and effectively isolated. Moosong et al [34] demonstrated that children were the most socially connected individuals in communities and concluded that this might explain the high prevalence of droplet-borne diseases (e.g. SARS) amongst children in epidemics generally.

Our models also simulated changes in the overall community network with time. This was both in the short-term through changes in participation in networks in the week and also through movement of individuals between networks as individuals aged, which resulted in them changing schools, going into hospital or entering care-homes. Empirical studies of contacts between individuals in settings such as schools [35, 36] as well as more large scale studies of social clustering across countries [37] have also shown that contact networks may be highly dynamic. When networks are dynamic the spread of disease through them also changes. Adaptive behaviour derived from knowledge of the existence of the disease could change network structure and impact on transmission and effectively stop epidemic spread [38]. Read et al [39] concluded that spread was dependent on the type of contact (close and casual) and the mode of transmission of the pathogen. They found that behavioural contacts in their network, which were more fine-scale than those used in our study, led to a network that was effectively a random-mixing model. Our results demonstrate that the burden of disease arising in the broad institutional settings of school, family, hospital are dependent on transmission coefficients specific to each setting but that duration of infectiousness and risk of environmental contamination were important drivers of predicted cases. At the community-level however, the burden of illness was dependent on environmental contamination risk, duration of illness, oyster consumption and transmission in secondary-schools. Part of the difference reflects the

serial dependency in norovirus with environmental contamination in any setting necessarily following a vomitus/faecal episode in that setting.

At a larger scale norovirus tends to be periodically pandemic (there were pandemics of norovirus in 1996, 2002, 2004, 2006, 2009 and 2012[40]) attributed to the pathogen undergoing hyper-mutation[41], leading to rapid evolution of new strains, as pandemic strains eventually lead to herd immunity. Norovirus variants may circulate in the population at low levels before acquiring other necessary mutations to facilitate their emergence as a pandemic virus[42]. Brockman and Helbing[43], in a study of global networks patterns of passenger transport, demonstrated that the arrival time for epidemic disease at a place was independent of disease parameters and more concerned with the effective distance across the contact network. It is possible that the periodic pandemic nature of norovirus disease reflects an interaction between the evolution of new strains and the nature of spread in disease via the contact networks in which it finds itself. Pandemic disease would then reflect populations linked through the presence of a small number of even more highly connected hubs than those in smaller population and institutional units with which we associate outbreaks.

There are obvious limitations of the modelling. Firstly, even though we created a simulation-cohort that represented the demography of a real population, our conceptual model of social interactions and connectivity in the population was simplified. We did not include adult workplace settings nor other individual-level interactions in inter-generational links or extended families. In addition our assessment of contact was rather crude and not based on actual measurable human behaviours, compared to empirical work on contact transmission. We also did not include all forms of food that are known to be contaminated: outbreaks in Canada, Belgium and France have been associated with consumption of salad vegetables and berry fruit[44]. However, the impacts of these simplifications are comparatively easy to assess. Workplace settings will mainly involve individual workers at establishments out-with the community. In this context their impact on network connectivity is probably minimal, except insofar as outside contacts with workplaces may act as foci for introgression of infection external to the 'residential' community. Introgression of disease into the community from these sources is likely to be highly stochastic, as it would be at the sparse but highly-linked hubs occurring at scales leading to pandemic behaviour. The absence of intergenerational links and extended families means that the connectivity in the social networks was highly conservative and as such the model will have underestimated network connectivity. We did not model variation in immune response to different norovirus strains, treating all as homologous in their impacts on development of disease and subsequent immunity. We did not model age-specific immunity, which is likely to be lower in care-homes and amongst the elderly[45], nor the effects of natural immunity that arise from the lack of histo-blood group antigens necessary for virus binding. We also ignored age-related dietary preferences, for example oysters are less likely to be consumed by children. Finally, the data used to validate the modelling were necessarily crude. Neither the regional-level nor IID2 data were at the same spatial scale and we only had data on recorded disease not the pathways that produced them. In effect our simplifications and assumptions about the social system and disease epidemiology and the mismatch between observed cases are likely to have led to a conservative and more hierarchically structured estimate of the social networks and their impacts on likely disease dynamics in this community setting.

Our research integrates different and coupled dynamic social networks with the immunology, demography, sources of contamination, at a scale larger than the outbreak setting. The models suggest that environmental contamination and food were major drivers of the number of norovirus cases in this social system. Oyster contamination, which was highly seasonal, acted as a source of repeated introgression into the community. Our results suggest that norovirus disease is multiply-scaled in time and space, and attention to different transmission routes and networks of exposure pathways is important if norovirus is to be managed other than retrospectively at the outbreak scale.

Additional Information

Data Accessibility

Analysis scripts and example data are available at DOI 10.17605/OSF.IO/T64MG

Authors' Contributions

SPR, PRH and SJO'B developed the overall project aims and objectives, and collated the relevant datasets. SPR authored the original R scripts; RAS ran model simulations and analysed outputs. MDFS contributed to the R code and analysis. All authors contributed to the final manuscript.

Competing Interests

We have no competing interests.

Funding

There is no funding to report for this project.

References

- [1] Adler, J. L. & Zickl, R. 1969 Winter Vomiting Disease. *J Infect Dis* **119**, 668-673. (DOI:10.1093/infdis/119.6.668).
- [2] Bányai, K., Estes, M. K., Martella, V. & Parashar, U. D. 2018 Viral gastroenteritis. *The Lancet* **392**, 175-186. (DOI:10.1016/s0140-6736(18)31128-0).
- [3] Becker, K. M., Moe, C. L., Southwick, K. L. & MacCormack, J. N. 2000 Transmission of Norwalk virus during a football game. *N Engl J Med* **343**, 1223-1227. (DOI:10.1056/NEJM200010263431704).
- [4] Moe, C. L. 2009 Preventing norovirus transmission: how should we handle food handlers? *Clin Infect Dis* **48**, 38-40. (DOI:10.1086/594119).
- [5] Riera-Montes, M., Brus Sjolander, K., Allestam, G., Hallin, E., Hedlund, K. O. & Lofdahl, M. 2011 Waterborne norovirus outbreak in a municipal drinking-water supply in Sweden. *Epidemiol Infect* **139**, 1928-1935. (DOI:10.1017/S0950268810003146).
- [6] Boone, S. A. & Gerba, C. P. 2007 Significance of Fomites in the Spread of Respiratory and Enteric Viral Disease. *Appl Environ Microbiol* **73**, 1687-1696. (DOI:10.1128/aem.02051-06).
- [7] Teunis, P. F., Moe, C. L., Liu, P., Miller, S. E., Lindesmith, L., Baric, R. S., Le Pendu, J. & Calderon, R. L. 2008 Norwalk virus: how infectious is it? *J Med Virol* **80**, 1468-1476. (DOI:10.1002/jmv.21237).
- [8] Boxman, I. L. A., Verhoef, L., Dijkman, R., Hagele, G., Loeke, N. A. J. M. T. & Koopmans, M. 2011 Year-Round Prevalence of Norovirus in the Environment of Catering Companies without a Recently Reported Outbreak of Gastroenteritis. *Applied and Environmental Microbiology* **77**, 2968-2974. (DOI:10.1128/Aem.02354-10).
- [9] Sabria, A., Pinto, R. M., Bosch, A., Bartolome, R., Cornejo, T., Torner, N., Martinez, A., Simon, M., Dominguez, A. & Guix, S. 2016 Norovirus shedding among food and healthcare workers exposed to the virus in outbreak settings. *J Clin Virol* **82**, 119-125. (DOI:10.1016/j.jcv.2016.07.012).
- [10] Atmar, R. L., Opekun, A. R., Gilger, M. A., Estes, M. K., Crawford, S. E., Neill, F. H. & Graham, D. Y. 2008 Norwalk virus shedding after experimental human infection. *Emerg Infect Dis* **14**, 1553-1557. (DOI:10.3201/eid1410.080117).
- [11] Tam, C. C. & O'Brien, S. J. 2016 Economic Cost of *Campylobacter*, *Norovirus* and *Rotavirus* Disease in the United Kingdom. *Plos One* **11**, e0138526. (DOI:10.1371/journal.pone.0138526).
- [12] Dowell, S. F. 2001 Seasonal Variation in Host Susceptibility and Cycles of Certain Infectious Diseases. *Emerging Infectious Diseases* **7**, 369-374. (DOI:10.3201/eid0703.017301).
- [13] Tam, C. C., Rodrigues, L. C., Viviani, L., Dodds, J. P., Evans, M. R., Hunter, P. R., Gray, J. J., Letley, L. H., Rait, G., Tompkins, D. S., et al. 2012 Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut* **61**, 69-77. (DOI:10.1136/gut.2011.238386).
- [14] Harris, J. P., Edmunds, W. J., Pebody, R., Brown, D. W. & Lopman, B. A. 2008 Deaths from norovirus among the elderly, England and Wales. *Emerg Infect Dis* **14**, 1546-1552. (DOI:10.3201/eid1410.080188).
- [15] Boon, D., Mahar, J. E., Abente, E. J., Kirkwood, C. D., Purcell, R. H., Kapikian, A. Z., Green, K. Y. & Bok, K. 2011 Comparative evolution of GII.3 and GII.4 norovirus over a 31-year period. *J Virol* **85**, 8656-8666. (DOI:10.1128/JVI.00472-11).
- [16] Lopman, B., Gastanaduy, P., Park, G. W., Hall, A. J., Parashar, U. D. & Vinje, J. 2012 Environmental transmission of norovirus gastroenteritis. *Curr Opin Virol* **2**, 96-102. (DOI:10.1016/j.coviro.2011.11.005).
- [17] Phillips, G., Tam, C. C., Conti, S., Rodrigues, L. C., Brown, D., Iturriza-Gomara, M., Gray, J. & Lopman, B. 2010 Community incidence of norovirus-associated infectious intestinal disease in England: improved estimates using viral load for norovirus diagnosis. *Am J Epidemiol* **171**, 1014-1022. (DOI:10.1093/aje/kwq021).
- [18] Lee, R. M., Lessler, J., Lee, R. A., Rudolph, K. E., Reich, N. G., Perl, T. M. & Cummings, D. A. 2013 Incubation periods of viral gastroenteritis: a systematic review. *BMC Infect Dis* **13**, 446. (DOI:10.1186/1471-2334-13-446).
- [19] Northumberland County Council. 2018 Northumberland Knowledge. (
- [20] Csardi, G. & Nepusz, T. 2006 The igraph software package for complex network research. *Interjournal - Complex Systems*, 1695.
- [21] Weibull, W. 1951 A statistical distribution function of wide applicability. *Journal of Applied Mechanics*.
- [22] Kim, A.-N., Park, S. Y., Bae, S.-C., Oh, M.-H. & Ha, S.-D. 2014 Survival of Norovirus Surrogate on Various Food-Contact Surfaces. *Food and Environmental Virology* **6**, 182-188. (DOI:10.1007/s12560-014-9154-4).
- [23] Atmar, R. L., Opekun, A. R., Gilger, M. A., Estes, M. K., Crawford, S. E., Neill, F. H., Ramani, S., Hill, H., Ferreira, J. & Graham, D. Y. 2014 Determination of the 50% human infectious dose for Norwalk virus. *J Infect Dis* **209**, 1016-1022. (DOI:10.1093/infdis/jit620).
- [24] R Core Team. 2016 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (
- [25] Vose, D. 2008 *Risk analysis: a quantitative guide*. John Wiley & Sons.
- [26] Shirley, M. D. F., Rushton, S. P., Smith, G. C., South, A. B. & Lurz, P. W. W. 2003 Investigating the spatial dynamics of bovine tuberculosis in badger populations: evaluating an individual-based simulation model. *Ecological Modelling* **167**, 139-157. (DOI:10.1016/s0304-3800(03)00167-4).
- [27] Wold, S., Johansson, E. & Cocchi, M. 1993 PLS-partial least squares projections to latent structures in 3D-QSAR in drug design; theory methods and applications. *PLS-Partial Least Squares Projections to Latent Structures*, 523-550.
- [28] Boelle, P.-Y. B. & Obadia, T. O. 2015 *R0: Estimation of R0 and Real-Time Reproduction Number from Epidemics*. *R package version 1.2-6*.
- [29] Sandmann, F. G., Jit, M., Robotham, J. V. & Deeny, S. R. 2017 Burden, duration and costs of hospital bed closures due to acute gastroenteritis in England per winter, 2010/11-2015/16. *J Hosp Infect* **97**, 79-85. (DOI:10.1016/j.jhin.2017.05.015).
- [30] Gaythorpe, K. A. M., Trotter, C. L., Lopman, B., Steele, M. & Conlan, A. J. K. 2018 Norovirus transmission dynamics: a modelling review. *Epidemiol Infect* **146**, 147-158. (DOI:10.1017/S0950268817002692).
- [31] Barnes, S. L., Morgan, D. J., Pineles, L. & Harris, A. D. 2018 Significance of multi-site calibration for agent-based transmission models. *IISE Transactions on Healthcare Systems Engineering* **8**, 131-143. (DOI:10.1080/24725579.2018.1431739).
- [32] Andersson, H. 1997 Epidemics in a population with social structures. *Math Biosci* **140**, 79-84. (DOI:10.1016/s0025-5564(96)00129-0).
- [33] Shirley, M. D. F. & Rushton, S. P. 2005 The impacts of network topology on disease spread. *Ecological Complexity* **2**, 287-299. (DOI:10.1016/j.ecocom.2005.04.005).
- [34] Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., et al. 2008 Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* **5**, e74. (DOI:10.1371/journal.pmed.0050074).
- [35] Kucharski, A. J., Wenham, C., Brownlee, P., Racon, L., Widmer, N., Eames, K. T. D. & Conlan, A. J. K. 2018 Structure and consistency of self-reported social contact networks in British secondary schools. *PLoS One* **13**, e0200090. (DOI:10.1371/journal.pone.0200090).
- [36] Fournet, J. & Barrat, A. 2014 Contact patterns among high school students. *PLoS One* **9**, e107878. (DOI:10.1371/journal.pone.0107878).
- [37] Xiao, X., van Hoek, A. J., Kenward, M. G., Melegaro, A. & Jit, M. 2016 Clustering of contacts relevant to the spread of infectious disease. *Epidemics* **17**, 1-9. (DOI:10.1016/j.epidem.2016.08.001).
- [38] Kotnis, B. & Kuri, J. 2013 Stochastic analysis of epidemics on adaptive time varying networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **87**, 062810. (DOI:10.1103/PhysRevE.87.062810).
- [39] Read, J. M., Eames, K. T. & Edmunds, W. J. 2008 Dynamic social networks and the implications for the spread of infectious disease. *J R Soc Interface* **5**, 1001-1007. (DOI:10.1098/rsif.2008.0013).
- [40] Vinje, J. 2015 Advances in laboratory methods for detection and typing of norovirus. *J Clin Microbiol* **53**, 373-381. (DOI:10.1128/JCM.01535-14).
- [41] Cuevas, J. M., Combe, M., Torres-Puente, M., Garijo, R., Guix, S., Buesa, J., Rodriguez-Diaz, J. & Sanjuan, R. 2016 Human norovirus hyper-mutation revealed by ultra-deep sequencing. *Infect Genet Evol* **41**, 233-239. (DOI:10.1016/j.meegid.2016.04.017).
- [42] Eden, J. S., Hewitt, J., Lim, K. L., Boni, M. F., Merif, J., Greening, G., Ratcliff, R. M., Holmes, E. C., Tanaka, M. M., Rawlinson, W. D., et al. 2014 The emergence and evolution of the novel epidemic norovirus GII.4 variant Sydney 2012. *Virology* **450-451**, 106-113. (DOI:10.1016/j.virol.2013.12.005).
- [43] Brockmann, D. & Helbing, D. 2013 The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337-1342. (DOI:10.1126/science.1245200).
- [44] Baert, L., Mattison, K., Loisy-Hamon, F., Harlow, J., Martyres, A., Lebeau, B., Stals, A., Van Coillie, E., Herman, L. & Uyttendaele, M. 2011 Review: norovirus prevalence in Belgian, Canadian and French fresh produce: a threat to human health? *Int J Food Microbiol* **151**, 261-269. (DOI:10.1016/j.ijfoodmicro.2011.09.013).
- [45] Pettrignani, M., van Beek, J., Borsboom, G., Richardus, J. H. & Koopmans, M. 2015 Norovirus introduction routes into nursing homes and risk factors for spread: a systematic review and meta-analysis of observational studies. *J Hosp Infect* **89**, 163-178. (DOI:10.1016/j.jhin.2014.11.015).
- [46] Food Standards Agency. 2017 The Food and You Survey Wave 4. (eds. A. Sondaal, B. Bates & S. N. Yeoh). Crown Copyright.
- [47] Parrino, T. A., Schreiber, D. S., Trier, J. S., Kapikian, A. Z. & Blacklow, N. R. 1977 Clinical Immunity in Acute Gastroenteritis Caused by Norwalk Agent. *New England Journal of Medicine* **297**, 86-89. (DOI:10.1056/nejm197707142970204).

Tables

Table 1) Parameter ranges used in LHS sensitivity analysis

Parameter	Primary-school ^a (probability)	Secondary-school ^a (probability)	Hospital ^a (probability)	Family ^a (probability)	Care-home ^a (probability)	Infectious duration ^b (days)	Probability of eating oysters ^c	Frequency of eating oysters ^c	Environmental contamination risk ^a	Environmental viral particle dose ^a	Maximum immune time ^d (days)
Minimum-maximum	0.0001-0.005	0.0001-0.005	0.0001-0.005	0.0001-0.005	0.0001-0.005	2 -14	0.0001-0.005	0.0001-0.0060	0.0001 - 0.0050	10-1800	10-720

a) Reliable data not available; plausible ranges used.

b) Atmar et al [10]

c) Food Standards Agency [46]

d) Parrino et al [47]

Figure and table captions

Table 1. Parameter ranges used in LHS sensitivity analysis

Figure 1. Schematic overview of modelling process

Figure 2. Variance inflation projections for different settings (a-e) or transmission pathways (f-g). a) primary school; b) secondary school; c) hospital; d) family; e) care-home; f) food; g) environment; h) total

Figure 3. Observed (red line) and predicted (\pm SE) cases of norovirus in North East England