Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Molly Katherine Steele Date

Anticipating Challenges to Norovirus Vaccines: Modeling Analyses to Understand How Norovirus Transmission and Genetic Diversity Could Affect Vaccine Development and Implementation

By

Molly Katherine Steele
Doctor of Philosophy

Environmental Health Sciences

_____

Benjamin A. Lopman, Ph.D.
Advisor

_____

Katia Koelle, Ph.D.
Committee Member

_____

Karen Levy, Ph.D.
Committee Member

_____

Lance Waller, Ph.D.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

Anticipating Challenges to Norovirus Vaccines: Modeling Analyses to Understand How Norovirus Transmission and Genetic Diversity Could Affect Vaccine Development and Implementation

By

Molly Katherine Steele

B.S., Montana State University, 2010

M.Sc., The Pennsylvania State University, 2013

M.P.H., Emory University, 2015

Advisor: Ben A. Lopman, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in

Environmental Health Sciences

2019

Abstract

Anticipating Challenges to Norovirus Vaccines: Modeling Analyses to Understand How Norovirus Transmission and Genetic Diversity Could Affect Vaccine Development and Implementation

By Molly Katherine Steele

Noroviruses are among the leading causes of diarrheal disease worldwide. Norovirus vaccines are currently in development; however there are several characteristics of norovirus that are not well understood and could hamper vaccine development and implementation.

In **Aim 1,** we estimated how norovirus transmission varies between age groups and how this variability could affect vaccine implementation. Using a deterministic, age-structured model of transmission and vaccination, we found that children under 5 years old contributed the most to transmission (age-specific basic reproduction number (R_0) of 4.3). Thus pediatric vaccination was predicted to avert 18-21 times more cases and twice as many deaths per vaccinee compared to elderly vaccination.

The size and severity of norovirus outbreaks varies across different settings, times of year and for different genotypes, suggesting that norovirus transmission is variable at the scale of outbreaks. In **Aim 2**, we estimated the basic (R_0) and effective (R_e) reproduction numbers for norovirus outbreaks in the US and used regression models to assess whether factors such as setting and season were associated with transmissibility. We found that norovirus outbreaks in the US have modest values of R_0 and R_e (2.75 [IQR: 2.38, 3.65] and 1.29 [IQR: 1.12, 1.74], respectively) and that outbreaks had higher transmission within long-term care/assisted living facilities, during winter, and when norovirus was the confirmed etiology.

GII.4 noroviruses evolve rapidly, and the extent of cross-immunity between strains is poorly understood. In **Aim 3**, we quantified the level of cross-immunity between multiple GII.4 strains. We developed a set of coupled single-strain models to estimate changes in population-level susceptibility and calculate the level of cross-immunity between strains. We estimated that the level of cross-immunity between Farmington Hill to Hunter and Hunter to Den Haag was high (0.84 - 0.91 and 0.91 - 0.94, respectively) and low between New Orleans to Sydney (0.34 – 0.73).

The variability of transmission both at the population-level and at the level of outbreaks provides key insights into which populations to target for vaccination. Certain GII.4 strains are associated with high cross-immunity, indicating vaccine formulations containing these strains may provide cross-protection against future strains of GII.4 norovirus.

Anticipating Challenges to Norovirus Vaccines: Modeling Analyses to Understand How Norovirus Transmission and Genetic Diversity Could Affect Vaccine Development and Implementation

By

Molly Katherine Steele

B.S., Montana State University, 2010

M.Sc., The Pennsylvania State University, 2013

M.P.H., Emory University, 2015

Advisor: Ben A. Lopman, Ph.D.

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in

Environmental Health Sciences

2019

Acknowledgements

First and foremost, I would to thank my dissertation committee, without whom this work would not have been possible. I would like to extend my deepest gratitude and appreciation to Ben Lopman, who encouraged me to embark on this doctoral journey, though I had my reservations. His unwavering patience, guidance and support over the last six years have shaped me into a better public health researcher and have been invaluable to completing this dissertation. Katia Koelle has provided education and support vital for the mathematical and statistical approaches used in this dissertation, and made deriving equations enjoyable (which I thought could never be possible). I have greatly appreciated Karen Levy's thoughtful feedback and encouragement to think of the bigger picture and how my research fits into public health. I am grateful for Lance Waller's helpful advice throughout my dissertation. Thank you to everyone who made this work possible, including Justin Remais, Andreas Handel, Lilly Pang, Karen Ellis, Lindsay Beck-Johnson, the members of the NoVaMod research group, collaborators at the Centers for Disease Control and Prevention Division of Viral Diseases, and collaborators at Public Health England.

I would like to thank faculty at the Rollins School of Public Health for their support and guidance throughout the doctoral program. I am especially grateful to my community of friends in the doctoral program who provided laughter, fun and much needed support when times were tough. Julia Baker has been the best conference travel buddy and provided calm and steady support when I felt things beginning to spin out of control. Thank you to Joana Yu for making sure that I took breaks from analysis and writing, and letting me take "staycations" with her furry friends.

And finally, I dedicate this dissertation to my wonderful family, Ken, Deb, Lura, Kevin, Ginger and Po. I am endlessly grateful for their love, support, encouragement, purrs and belly rubs, which have been immeasurable to maintaining my sanity over the last eight years I've been in graduate school.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Global mortality and the burden of norovirus

The Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) estimated that the global number of early deaths in 2017 was 56 million, where early deaths are defined as deaths that occur prior to the estimated life expectancy. Diarrheal diseases are the fifth leading cause of excess mortality, responsible for approximately 1.57 million deaths. Most diarrheal related deaths occurred in children under the age of five (534,000 deaths) and the elderly ≥ 70 years old (624,000 deaths).[1] In addition to a high burden of mortality, acute gastroenteritis (AGE) has the second highest morbidity burden of all infectious diseases, causing an annual estimated loss of 81 million disability-adjusted life years (DALYs).[2] While diarrheal related DALYs and mortality have declined by approximately 28% and 30%, respectively, since 2007, disease and death due to diarrhea still represent a substantial health burden globally.[1–3]

Noroviruses are among the leading causes of endemic diarrheal disease worldwide, associated with approximately 18% (95% CI: 17-20%) of AGE cases.[4] Estimates from the GBD study in 2016 and two systematic reviews indicate that noroviruses cause between 140 million to 677 million AGE episodes and between 71,000 to 212,000 deaths annually.[5–7] Historically, rotavirus gastroenteritis was the leading cause of diarrheal disease globally. However, with the establishment of routine rotavirus vaccination in many countries, the rotavirus burden has declined substantially and norovirus is now becoming the predominant cause of diarrheal diseases.[8,9]

1.2 Burden of norovirus in the United States

In the United States, noroviruses are the leading cause of AGE,[10–12] responsible for an average 570–800 deaths, 56,000–71,000 hospitalizations, 400,000 emergency department admissions, 1.7–1.9 million outpatient admissions, and 19–21 million illnesses annually.[11,13,14] Norovirus can infect people of all ages, however, certain populations and age groups are more vulnerable to chronic disease or severe outcomes. Approximately 90% of norovirus-associated deaths in the US occur among the elderly (~20 per 1,000,000 persons per year).[15] Children under five years of age experience the highest incidence of AGE (~20 per 100 persons per year)[16] and have the highest rates of outpatient, emergency department, and inpatient visits (233, 38, and 9.4 per 10,000 persons per year, respectively).[13,14]

1.3 Norovirus seasonality

The incidence of norovirus disease is variable within a given year. Norovirus is colloquially known as the "winter vomiting disease" due to its documented winter seasonality in temperate climates.[17] In the US, the majority of norovirus cases and outbreaks occur during the cool months, from October through March, with peaks in cases and outbreaks occurring from November through January. While the exact cause of norovirus seasonality is not well known, several studies have indicated that weather factors (i.e., temperature and humidity) may have an impact on the seasonality of norovirus infections.[17–19] Norovirus case and outbreak reports are inversely correlated with temperature,[17,19] and norovirus surrogate virus (e.g., murine norovirus, feline calicivirus) survival declines with increasing temperatures.[20,21]

1.4 Viral characterization

Noroviruses are non-enveloped, icosahedral viruses of the family *Caliciviridae*. The viral genome contains positive-sense single-stranded RNA with three open reading frames (ORFs). ORF1 encodes for nonstructural proteins used for viral replication, ORF2 encodes for viral protein 1 (VP1), and ORF3 encodes a minor structural protein (VP2).[22] The viral capsid is predominately composed of VP1 and has two distinct domains: the shell (S) domain, which forms the core of the virus, and protruding (P) domain. The P domain is further divided into two sub-domains, P1 and P2. P2 is a hypervariable region that extends the furthest from the capsid shell and facilitates binding to host receptors.[22–24] Noroviruses are categorized into six established norovirus genogroups (GI – GVI). Within each genogroup, noroviruses are further subdivided into genotypes based on amino acid sequence similarity of the VP1 capsid protein and the RNA polymerase region of ORF1.[25,26] There are three genogroups (GI, GII and GIV) with up to 33 genotypes can infect humans.[26] A single genotype, genogroup II genotype 4 (GII.4), is estimated to be the cause of 60% to 80% of all norovirus infections.[27,28]

1.5 Evolution of Noroviruses

Norovirus evolution is driven by the accumulation of point mutations and recombination. Most norovirus genotypes do not evolve rapidly, however GII.4 noroviruses have been rapidly evolving since the mid-1990s.[29] GII.4 noroviruses undergo punctuated antigenic evolution, leading to new strains emerging in populations every two to five years.[26,30–32] These antigenic changes are thought to be the result of selection pressures from population immunity. Mutations in the VP1 gene lead

to changes in structural capsid proteins that change antigenicity, particularly the P2 domain of the capsid, which allow noroviruses to evade host immunity.[28,30,33–39] Phylogenetic analyses and immunological characterization have indicated that 14 different strains of GII.4 noroviruses have emerged since the mid-1990s, seven of which circulated globally.[31,40–44] The first GII.4 strain that circulated globally was US 95/96.[28] Following that first pandemic strain, new pandemic GII.4 strains emerged in 2002 (GII.4 Farmington Hills), 2004 (GII.4 Hunter), 2006 (GII.4 Den Haag), 2009 (GII.4 New Orleans) and 2012 (GII.4 Sydney).[26,36,40] Recently, changes in the RNA polymerase of GII.4 strains, which may increase transmission, have been suggested as a driver of GII.4 evolution as well.[45] As such, a new strain of norovirus, Sydney 2015, emerged in late 2014 and has circulated globally. This new strain shares a highly similar viral capsid with Sydney 2012, however Sydney 2015 has a GII.P16 polymerase (GII.P16-GII.4 Sydney) while the Sydney 2012 strain had a GII.P4 polymerase (GII.P4-GII.4 Sydney).[45,46]

1.6 Natural history and transmission of norovirus

Noroviruses spread via the fecal-oral or vomitus-oral route. Noroviruses are hardy viruses that can persist on surfaces for up to two weeks, are infectious from 0-60°C and resistant to many common household chemical cleaners.[20,21,47,48] These viruses are highly infectious. Two studies have estimated the infectious dose sufficient to cause infection in 50% of those exposed (ID_{50}) to the prototype strain of norovirus, Norwalk virus (genogroup I genotype1, GI.1). These estimates range from 18 to 1,320-2,800 gene equivalent copies,[49,50] Individuals will shed copious amounts of virus in their stool and vomitus. Infected individuals can shed up to 10^5-10^{11} viral copies per gram of feces[51–53] and on average will shed 8.0×10^5 and 3.9×10^4 viral copies per mL in vomitus for GI

and GII viruses, respectively.[54] The viral shedding period is highly variable between individuals and has been documented to begin as early as 18 hours after exposure and extend as long as eight weeks post exposure among immunocompetent hosts.[52,53,55] As such, transmission of norovirus can occur prior to an individual developing symptoms and well after they have recovered.

The duration of viral shedding has also been documented to vary depending on the genogroup causing infection. One recent challenge study found that among participants challenged with Norwalk (GI.1) virus, the median duration of shedding was 17 days (range: 5-27 days) with peak viral shedding occurring between 3 and 15 days post infection. Participants challenged with Snow Mountain (GII.2) virus tended to have a shorter duration of shedding, with a median of 5 days (range 2-25 days) and peak viral shedding occurring 2-6 days post infection.[52]

The low infectious dose, high viral shedding, and environmental persistence of noroviruses mean they can be easily transmitted through several routes, including through person-to-person, foodborne, waterborne, and environmental pathways.[50,56–59] Transmission of norovirus, as characterized by the basic reproduction number (R_0, the average number of secondary cases that arise from a primary case in fully susceptible population), has been estimated from several transmission modeling studies; however there is large variation in these estimates (range of estimated R_0: 1.1 to 7.2).[60] Importantly, young children may be important drivers of norovirus transmission. Observational studies have identified contact with a young child with norovirus gastroenteritis as a risk factor for diarrhea for older children and adults.[16,61]

Following exposure to noroviruses, individuals will enter a relatively short incubation period.[62,63] A recent meta-analysis of data on 1,022 norovirus outbreaks found the average incubation period to be 32.8 hours (95% CI: 30.9-34.6).[63] Symptoms of norovirus infection include a sudden onset of vomiting, watery, non-bloody diarrhea, abdominal cramps, fever and malaise.[64–67] The average duration of symptoms, also estimated by meta-analysis, is 44.2 hours (95% CI: 38.9-50.7).[63] Children less than one year old are more likely to experience diarrhea given infection with norovirus, while vomiting and diarrhea are the predominant symptoms of norovirus in individuals who are older than one year.[68,69] Persons infected with norovirus are more likely to experience vomiting, and less likely to have bloody diarrhea and long-term sequelae compared to other causes of gastroenteritis, especially bacterial pathogens such as *Salmonella* and *Campylobacter*.[68–75] Among immunocompetent individuals, noroviruses typically cause acute, self-limiting infections; however, immunocompromised individuals experience more severe disease and, sometimes, chronic norovirus infections.[70,76–79]

1.7 Host susceptibility and immunity

Susceptibility to norovirus depends, in part, on an individual's genetics.[66,80–85] Secretor positive individuals have a functional fucosyltransferase-2 (FUT2) gene, which encodes for α(1,2)fucosyltransferases that produce H type-1 antigens, which serve as precursors to A and B antigens.[86] Secretor negative individuals have non-functional FUT2 genes and do not produce H type-1, resulting in an absence of histo-blood group antigens (HBGAs) antigens in saliva and mucosa.[86] Certain norovirus genotypes are strongly associated with HBGAs; secretors (i.e., those

who express HGBAs in mucosa) are susceptible to GI.1 and GII.4 norovirus infections, however non-secretors are almost completely resistant to GI.1 and GII.4 norovirus infection.[80,82,83,87,88]

The duration of immunity to norovirus is not well understood. Challenge studies have reported that the duration of immunity is short term, typically six months to two years;[66,84,89,90] however the challenge doses administered in these studies were several thousand-fold greater than the amount of virus capable of causing infections. Immunity may be more robust to lower challenge doses. A mathematical modeling study suggests the duration of immunity to norovirus is longer term, around four to six years; however this model did not account for multiple strains of norovirus.[91] Therefore the duration of immunity to norovirus may be shorter than what was estimated from this model.

Currently, we have little understanding of whether cross-immunity exists between different genogroups or genotypes of norovirus. One challenge study conducted in 1974 explored heterotypic versus homotypic immunity to three norovirus genotypes: GI.1 (Norwalk virus), GII.1 (Hawaii virus) and GI.5 (Montgomery County virus). In this study, participants were challenged twice, either with the same virus or a virus of a different genotype. Participants challenged with the same genotype, or a genotype within the same genogroup did not become ill, but those challenged with a genotype from a different genogroup became ill.[66] This suggests there may be heterotypic immunity between genotypes within the same genogroup. Though there have been no further published challenge studies with multiple norovirus genotypes, *in vitro* studies (i.e., those employing enzyme immunoassays (EIAs) or antibody blocking assays) have indicated that immune responses are stronger between genotypes within the same genogroup than genotypes

from different genogroups.[36,43,92–97] Further, we have limited understanding of whether there is cross-immunity between different strains of GII.4 norovirus. When new GII.4 strains emerge, they may lead to pandemics (e.g., Farmington Hills 2002 and Den Haag 2006 strains) or simply replace previous strains without disturbing the endemic pattern (e.g., New Orleans 2009 and Sydney 2012 strains), which suggests differences in the level of cross-protection between strains.[28,33,40,98–100]

1.8 Norovirus in outbreak settings

Norovirus is the most common cause of acute gastroenteritis outbreaks in the US. The Centers for Disease Control and Prevention (CDC) monitors acute gastroenteritis outbreaks through a passive surveillance database known as the National Outbreak Reporting System (NORS). State, local and territorial health departments submit web-based forms for enteric disease outbreaks for all modes of transmission.[101] Between 2009 and 2017, 17,822 norovirus outbreaks were reported to NORS, the vast majority of which (78%) were spread by person-to-person transmission. Outbreaks that spread by person-to-person transmission can occur in many different settings; in the US the top settings where outbreaks occurred were long term care facilities (61%), schools (9%), hospitals (3%), and child care facilities (3%).[102] The size and severity of outbreaks can vary across different settings, times of year and for different genotypes, suggesting that the transmissibility of norovirus may be variable across different outbreak contexts.[103]

1.9 Norovirus vaccines

In response to the substantial health burden posed by noroviruses, there have been recent efforts to develop norovirus vaccines.[34,104–112] While an *in vitro* culturing system has recently developed this system is not widely available.[113] Given this limitation vaccine development has typically employed the use of virus like particles (VLPs). When recombinant norovirus VP1 capsid proteins are expressed, these capsid proteins will self-assemble VLPs, which are non-replicating, but morphologically and antigenically indistinguishable from native virus.[109,112,114,115] Initial human challenge studies have demonstrated that norovirus VLP vaccines produce immunogenic responses in humans that are comparable to what the virus itself causes.[108,109,111,112,114] As a result, one bivalent intramuscular product has undergone Phase IIb field efficacy trials.[108,109,112] This bivalent norovirus vaccine contains VLPs derived from GI.1 virus and a GII.4 consensus sequence containing epitopes from three GII.4 strains: Houston 2002, Yerseke 2006 and Den Haag 2006.[116] Data from efficacy studies and surrogate neutralization analyses show that this vaccine is well tolerated, elicits a strong immune response and may provide immunity against genotypes and strains not represented in the vaccine formula.[111,117,118] Initial data from a GII.4 challenge study showed the vaccine did not significantly reduce infection relative to placebo dosed controls; however the vaccine did reduce the severity of disease.[108]

1.10 Challenges that norovirus poses to vaccine development and implementation

Several characteristics of norovirus along with limitations in our current understanding of norovirus immunity have hampered vaccine development. First, noroviruses are error-prone single-stranded RNA viruses, resulting in great genetic and antigenic diversity and rapid viral evolution, particularly within the GII.4 genotype.[30,36] Given the rapid evolution of GII.4 norovirus,

norovirus vaccine development may face challenges similar to those seen with influenza vaccines.[119] As described above, human immunity to norovirus is complex and incompletely understood, though current knowledge suggests strain- or genotype-specific protection is of limited duration.[66,84,89,90] Given the diversity of noroviruses, a vaccine would ideally provide immunity against genogroups and genotypes not represented in the vaccine as well as new genotypes of norovirus (notably GII.4 noroviruses) that evolve over time. The extent of cross-protection provided between genogroups and genotypes, particularly GII.4 norovirus strains, has important implications for the design of vaccines. If the level of cross-protection between current and future strains of norovirus is high, then norovirus vaccines may not need to be reformulated regularly. If, however, the level of cross-protection between current and future strains is low, then norovirus vaccines may need to be reformulated frequently to keep pace with GII.4 evolution.

Second, observational studies and outbreak data suggest that norovirus transmission may be variable by age and across different settings. As such, there is an array of possible vaccine strategies ranging from untargeted mass vaccination (like the current influenza vaccine guidance) to age-targeted (e.g. young children or elderly) or targeting of groups important in transmission (e.g. food handlers). When considering age-targeted vaccine strategies, children under the age of five experience the highest incidence of disease and have the highest rates of healthcare utilization.[13,14,16] Further, data from observational studies and a mathematical modeling study suggest that young children may be important drivers of norovirus transmission.[16,61,91] Targeting young children for vaccination could be an important strategy to reduce the burden of disease in this age group, but also to provide population-level benefits. If young children are important drivers of norovirus transmission, then targeting this age group could lower overall transmission

of norovirus resulting in indirect benefits to the population (via herd protection). The elderly are another age group to consider for vaccination, as they experience the vast majority of norovirus related deaths.[15] Vaccination of this group may be considered if the public health goal is to reduce norovirus-associated mortality. Additionally, in the US the majority of norovirus outbreaks occur in long-term care and assisted living facilities.[102] A vaccination strategy targeting elderly populations in long-term care facilities could be considered to mitigate and/or prevent outbreaks in these settings.

In addition to targeting specific age groups, groups with high risk of transmission and disease can be considered for vaccination. Known sub-populations at risk for norovirus transmission and disease include: food service workers, healthcare workers, immunocompromised individuals, military personnel and travelers. Current evidence suggests that food service workers and healthcare workers may serve important roles in transmission and act as sources for foodborne disease and outbreaks in healthcare settings.[120,121] Thus these groups could be considered for vaccination to reduce the burden of norovirus outbreaks in these settings. Travelers and military personnel could also be targeted for vaccination as they experience high burdens of norovirus AGE. 3-17% of AGE episodes during or after travel to low-income regions (i.e., travelers' diarrhea) are associated norovirus[122] and norovirus outbreaks are common in the military given confined environments that can promote transmission (e.g., Navy ships and training camps).[123,124] Vaccination could also target immunocompromised individuals as they have a high risk for chronic infections and death, and experience prolonged viral shedding, which may promote transmission.[70,76–79]

With the wide variety of potential vaccine strategies, it will be challenging to determine which strategies to focus on for vaccine development and implementation. It will likely be many years still before field studies are conducted to determine the efficacy of different vaccine implementation strategies. Further, field efficacy studies will be time-consuming and expensive, which may limit which vaccine strategies can be tested. Mathematical models of norovirus offer a unique opportunity to simulate the efficacy and impact of any number of vaccination strategies to help guide vaccine development and implementation, as detailed in the section below.

1.11 The value of dynamic models for studying infectious disease epidemiology

Infectious diseases are driven by a series of "dependent happenings," meaning that the frequency with which an event occurs depends on some other quantity, resulting in dynamical feedback.[125,126] For example, the incidence of infectious diseases depends on the prevalence of disease in a given population. Thus, in many cases, when public health interventions reduce the prevalence of disease, the incidence of disease also declines. Unlike traditional statistical models used for many classical epidemiological studies, mathematical models incorporate these dynamical feedbacks to characterize the dynamic nature of infectious disease systems and predict how these systems may respond to public health interventions.[127,128] Mathematical models have been used extensively to predict the dynamics of many infectious diseases in a range of contexts and with a diversity of analytical goals. For example, mathematical models have been used to describe the transmission dynamics of influenza viruses,[129,130] to predict the probability of international spread of emerging influenza strains,[131] and to evaluate the potential impact of different intervention strategies, such as vaccination and school closures.[132–135]

Mathematical models offer a unique opportunity to address questions about norovirus epidemiology that are challenging to address through laboratory, field and observational studies. Transmission of norovirus is difficult, if not impossible, to observe and quantify in the field. As such many studies have employed variations of dynamic models to quantify norovirus transmission and predict dynamics of disease at the population-level and across a range of settings.[91,136–138] As reviewed by Gaythorpe et al., norovirus transmission models have provided a range of estimates for R_0 (i.e., the average number of infections that arise from a primary infection in a completely susceptible population) between 1.1 to 7.2.[60] An important direction for these models is to harness this understanding of transmission to estimate the ability of different types of public health interventions, such as vaccination strategies, to interrupt transmission and maximize population-level benefits.

Mathematical models can also be used to better understand norovirus immunity and cross-protection between evolving GII.4 strains. Until very recently, attempts to culture norovirus *in vitro* using an array of cell systems have been unsuccessful.[139–141] Recently, human noroviruses have been successfully cultured in stem cell–derived, human intestinal enteroids (HIEs),[113] though this system is not yet widely available in research labs. In the absence of a widely available culturing method, much of our knowledge of norovirus immunity and cross-protection has been inferred from challenge studies, studies of VLPs and surrogate neutralization assays.[36,43,92–97] Mathematical models can be used to supplement the findings from challenge and *in vitro* studies and translate this understanding of norovirus immunity and cross-protection to the population-level. All existing norovirus transmission models assume a single strain of norovirus, or that exposure to a single strain provides immunity to all subsequent strains to which individuals are

exposed.[91,121,136,138,142] While these models have provided important insights into the epidemiology of noroviruses, these results may not be generalizable given the vast genetic and antigenic diversity of noroviruses. There is a growing body of research into modeling the dynamics of pathogens with multiple strains, particularly influenza virus.[143–146] Similar to GII.4 noroviruses, influenza A viruses undergo antigenic drift resulting in the emergence of new strains that evade population immunity.[119,147] Thus, multi-strain transmission modeling studies have been used to predict the dynamics and evolution of influenza A viruses.[145,148] One such study combined a genetic-antigenic model with a transmission model to measure how changes in cross-immunity can drive the evolution of interpandemic influenza A over time.[119] Given the patterns of evolution between GII.4 noroviruses and influenza A viruses, the multi-strain models that have been developed to describe and predict cross-immunity and evolution influenza A can be adapted and applied to GII.4 norovirus.

2 Significance

Noroviruses pose a substantial burden disease across the age range. Norovirus vaccines are currently in development; however there are several characteristics about norovirus (i.e., transmission and immunity) that are not well understood and could hamper vaccine development and implementation. While it is essential to assess the efficacy of these vaccines, to ensure the success of norovirus vaccination programs, it is equally important to understand how aspects of norovirus biology and epidemiology, such as transmission and genetic diversity, could affect vaccine efforts.

The overall goal of this dissertation is to provide insight into key aspects of norovirus epidemiology that could pose challenges to vaccine development and implementation. In this dissertation, I examined how norovirus transmission varies between age groups and how this variability could affect vaccine strategies with implications for implementation (Aim1), characterized norovirus transmission across different settings of outbreaks (Aim 2); and quantified the level of immune escape between multiple GII.4 norovirus strains (Aim 3).

2.1 Aim 1 Rationale and Overview

Current norovirus vaccine evaluations have been trialed among adults; however, as noroviruses affect all ages and are transmitted through multiple routes, an array of vaccination strategies warrants consideration. Few studies have sought to quantitatively assess the population-level impact of vaccines on norovirus illness. One study used a Markov model to assess the impact of norovirus vaccines and associated economic benefits of vaccination.[149] This model however does

not consider the dynamics of disease transmission, and thus cannot account for the potential indirect benefits of vaccination (i.e., herd immunity and reductions in transmission).[149] Therefore the estimated impacts of vaccination from this study may be a substantial underestimate.[149] Further, observational studies suggest that young children are important in norovirus transmission; contact with a young child with norovirus gastroenteritis has been identified as a significant risk factor for diarrhea for older children and adults.[16,61] If young children are indeed important in transmission, then targeting this age-group could lead to wider population benefits by reducing overall disease transmission. The elderly suffer the vast majority of deaths due to norovirus,[15] thus targeting the elderly for vaccination could reduce the incidence of the norovirus deaths substantially. A better understanding of population-level, age-specific variation in norovirus transmission is required so that we can determine which age groups to target for vaccination to maximize population-level benefits.

AIM 1: Identify how norovirus transmission varies between age groups and assess how this variability could affect vaccine impact.

Aim 1A: Understand norovirus transmission by estimating age-specific transmission parameters.

A dynamic model of norovirus transmission and immunity in the US was developed to estimate age-specific transmission parameters and determine how different age groups contribute to transmission.

Aim 1B: Characterize age-specific and population level impacts of different vaccination strategies on the epidemiology and disease burden of norovirus.

The model from Aim 1A was modified to include vaccination compartments to simulate the effects of pediatric and elderly vaccination strategies on the incidence of five clinical outcomes: outpatient (OP) admission, emergency department (ED) admission, hospitalization (IP), and death due to norovirus.

2.2 Aim 2 Rationale and Overview

In addition to understanding population-level variations in transmission, it is also important to consider the variability of transmission at the scale of outbreaks. Outbreaks generally affect sub-populations (e.g., food handlers, immunocompromised, healthcare workers) whose risk for transmission and disease may be substantially different from what is observed at a population-level. Thus public health interventions (such as vaccination strategies) required to mitigate or prevent outbreaks may differ substantially from population-level interventions.[150] Analyses of norovirus outbreak data have shown that the size and severity of outbreaks varies from different settings, times of year and for different genotypes. The duration and amount of shedding during norovirus infection is variable between individuals and with different genogroups, and contact patterns change for different settings (i.e., higher number of contacts made at home vs school or work).[151] Thus we can expect norovirus transmission is highly variable across different outbreak contexts.[35]

One method to better understand transmission during outbreaks is to calculate the basic and effective reproduction numbers (R_0 and R_e). R_0 represents the average number of secondary cases that arise from a primary case in a completely susceptible population. R_e represents the average number of secondary cases that arise from a primary case in a population that is not necessarily fully susceptible.[152,153] Importantly, when $R_0>1$ an epidemic will likely occur. The value of R_0 is often used to determine the type and level of control methods required to mitigate outbreaks (i.e., reduce R_0 to less than 1).[153,154] A more comprehensive estimation of norovirus transmission is required to better understand whether transmission varies in different times and settings.

AIM 2: Characterize how the basic (R_0) and effective (R_e) reproduction number of norovirus outbreaks in the US varies across different settings, seasons and genotypes.

The basic (R_0) and effective (R_e) reproduction numbers were estimated for norovirus outbreaks reported to NORS in the US between 2009 and 2017. Then regression models were used to assess whether norovirus transmission is associated with outbreak setting, census region, season, year, whether norovirus was suspected or confirmed, and norovirus genotype (categorized as either GII.4 or non-GII.4).

2.3 Aim 3 Rationale and Overview

A gap in our current knowledge of norovirus immunity is the extent of cross-protection provided when individuals are exposed to different strains of norovirus. As detailed above, data from *in vitro* EIAs, antibody blocking assays, and challenge studies suggest immunity tends to be strain- or genotype-specific,[66,84,89,90] although there may be limited cross-immunity between strains

within the same genogroup.[36,43,155] All existing models of norovirus transmission assume there is a single strain of norovirus, or that exposure to a single strain provides immunity to all subsequent strains to which individuals are exposed.[91,121,136,138,142] These existing models provide important insight into the epidemiology of norovirus, however fail to capture the observed inter-annual variability that can result from emerging GII.4 strains. Additionally, while these models capture certain outbreak patterns, they may not accurately reflect the underlying epidemiological processes driving these patterns. Thus, these models could make incorrect predictions of the impact of different types of interventions, such as vaccination.

AIM 3: Estimate changes in population level susceptibility to GII.4 noroviruses and the extent of immune escape and cross-protection between GII.4 strains.

To address this aim, a set of coupled single-strain models approach was used to estimate the change in population-level immunity during GII.4 strain transitions. This estimated change in susceptibility was used to quantify the level of immune escape and infer the level of cross-protection between GII.4 strains.

3 Manuscript for Aim 1

[Steele, M. K., Remais, J. V., Gambhir, M., Glasser, J. W., Handel, A., Parashar, U. D., & Lopman, B. A. (2016). Targeting pediatric versus elderly populations for norovirus vaccines: a model-based analysis of mass vaccination options. *Epidemics*, *17*, 42-49.]

Targeting pediatric versus elderly populations for norovirus vaccines: a model-based analysis of mass vaccination options

Molly K. Steele, Justin V. Remais, Manoj Gambhir, John W. Glasser, Andreas Handel, Umesh D. Parashar, Benjamin A. Lopman

Abstract

Background: Noroviruses are the leading cause of acute gastroenteritis and foodborne diarrheal disease in the United States. Norovirus vaccine development has progressed in recent years, but critical questions remain regarding which age groups should be vaccinated to maximize population impact.

Methods: We developed a deterministic, age-structured compartmental model of norovirus transmission and immunity in the US population. The model was fit to age-specific monthly US hospitalizations between 1996 and 2007. We simulated mass immunization of both pediatric and elderly populations assuming realistic coverages of 90% and 65%, respectively. We considered

two mechanism of vaccine action, resulting in lower vaccine efficacy (lVE) between 22% and 43% and higher VE (hVE) of 50%.

Results: Pediatric vaccination was predicted to avert 33% (95% CI: 27%, 40%) and 60% (95% CI: 49%, 71%) of norovirus episodes among children under five years for lVE and hVE, respectively. Vaccinating the elderly averted 17% (95% CI: 12%, 20%) and 38% (95% CI: 34%, 42%) of cases in 65+ year olds for lVE and hVE, respectively. At a population level, pediatric vaccination was predicted to avert 18-21 times more cases and twice as many deaths per vaccinee compared to elderly vaccination.

Conclusions: The potential benefits are likely greater for a pediatric program, both via direct protection of vaccinated children and indirect protection of unvaccinated individuals, including adults and the elderly. These findings argue for a clinical development plan that will deliver a vaccine with a safety and efficacy profile suitable for use in children.

INTRODUCTION

Noroviruses are the leading cause of acute gastroenteritis in the United States,[10–12] responsible for an average 570–800 deaths, 56,000–71,000 hospitalizations, 400,000 emergency department admissions, 1.7–1.9 million outpatient admissions, and 19–21 million illnesses annually.[11] Severe norovirus outcomes occur among pediatric and elderly populations, with 90% of norovirus-associated deaths in the US occurring among the elderly.[15] Children under five years of age experience the highest incidence (five times the general population)[16] and have the highest rates of outpatient, emergency department, and inpatient visits (233, 38, and 9.4 per 10,000 persons per year, respectively).[13,14] Given this substantial burden and limited options for prevention and treatment,[156] vaccines are considered an important means of providing protection from norovirus illness.[12]

Safety, immunogenicity, and efficacy studies on norovirus vaccines have been encouraging, with at least one bivalent intramuscular product likely to progress to Phase III field efficacy trials.[12] Current vaccine evaluations have been conducted among adults. However, as noroviruses affect all ages and are transmitted through multiple routes, an array of vaccination strategies warrants consideration. At the current stage of vaccine development, the time is ideal for examining the population impact that various norovirus vaccination programs could have on the dynamics of disease to guide vaccine development and inform policymakers on potential impacts.

Here, we present an age-structured dynamic transmission model to project the effects of different vaccination strategies on the epidemiology and disease burden of norovirus in the US, including the incidence of five clinical outcomes (cases, outpatient visits, emergency department visits,

inpatient visits, and deaths) for each of four age classes (0-4, 5-17, 18-64, and 65+ years). The model was used to compare vaccination strategies targeting pediatric versus elderly populations, both in terms of impact on disease burden and relative efficiency under various assumptions about vaccine efficacy.

METHODS

We adapted a previously-published, deterministic, age-structured compartmental model that simulates norovirus transmission and estimates disease incidence in the US.[91] The model follows a Susceptible-Exposed-Infected-Recovered (SEIR-like) framework (Figure 3-1, Text S1). We consider four age classes: 0-4, 5-17, 18-64, and 65+ years old, and applied realistic, age-specific population sizes, aging and death rates, and a heterogeneous contact structure (Table 3-1, Text S1, Table 3-S1). Lacking detailed mixing data specific to the US, we used average contact patterns from representative samples of eight European countries in the POLYMOD study.[151]

We estimated age-specific susceptibilities (q_i) to allow the four age classes (i; 0-4, 5-17, 18-64, and 65+ year olds) to exhibit heterogeneous probabilities of infection given exposure to an infectious contact. We also considered models with different numbers of estimated age-specific susceptibilities (q_i) and where transmission was dependent on susceptible or infectious individuals (Text S1, Table 3-S2); the results of this paper focuses on the best-fit model, where the probabilities of infection on contact for 5-17 and 18-64 year olds were equal ($q_2 = q_3$).

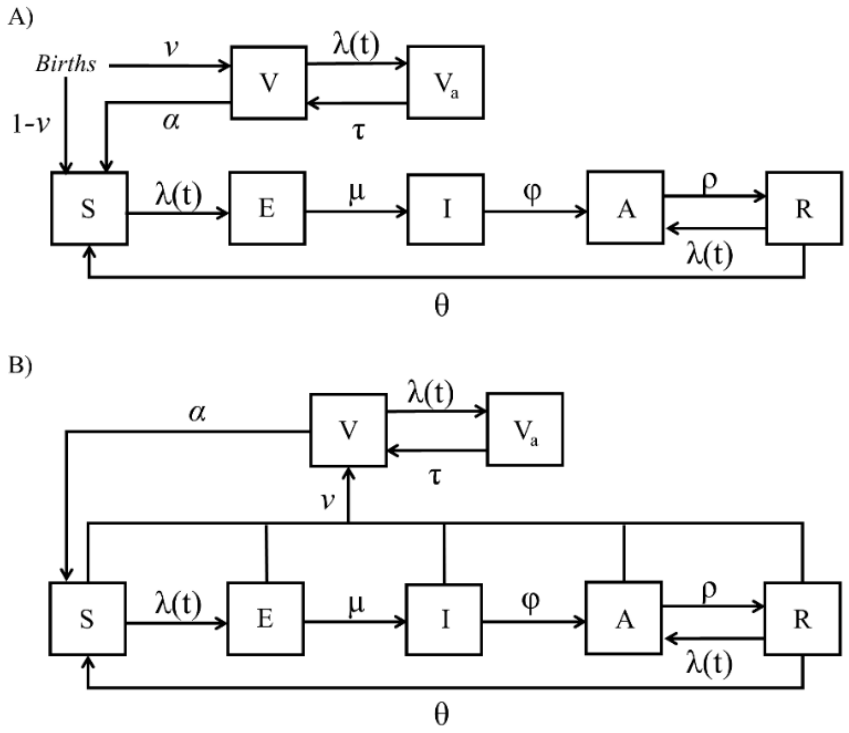
We assume maternal immunity is short-lived and negligible.[157] Therefore, absent vaccination, children are born into the susceptible class (S). Susceptible individuals are subjected to a force of infection ($\lambda_i(t)$), and progress through pre-symptomatic (E), symptomatic (I) and post-symptomatic (A) stages at rates inversely proportional to the duration of incubation (μ), symptomatic illness (φ), and asymptomatic shedding (ρ), respectively, before entering the recovered compartment (R). In this framework, individuals acquire natural immunity that protects against disease, but not against infection, until immunity wanes.[16,83] From the recovered compartment, persons can become asymptomatically infected (A) or susceptible to disease as natural immunity wanes (θ). To simulate seasonality, we applied a seasonal forcing parameter (β_1) that governs the peak-to-mean amplitude in transmissibility. To estimate clinical outcomes, we multiplied the projected disease incidence by age-specific probabilities (given norovirus illness) of outpatient (OP) admission, emergency department (ED) admission, hospitalization (IP), and death due to norovirus. These probabilities were determined from US population estimates and described in more detail in previous work.[149]

Model simulation, fitting, and analysis were conducted in R version 3.1.1.[158] Specific R packages used for these analyses are detailed in the supplement. We fit the model to age-specific monthly counts of norovirus-associated hospitalizations by maximum likelihood to estimate the susceptibility ($q_{1\ldots4}$) and seasonality (β_1,ω) parameters.[14] We assumed the monthly numbers of hospitalizations in each age group were Poisson distributed with mean equal to the model-predicted age-specific incidence multiplied by the probability of hospitalization.[159] Age-specific R_0 values were calculated following procedures detailed in the supplement and Simmons et. al., 2013 (Text S1).[91]

Vaccine Scenarios

We assumed vaccine response was "take-type:" either protection against disease was complete or vaccinated individuals remained fully susceptible.[160] We assumed vaccines confer protection in the same manner as we conceptualize natural immunity: providing protection against disease, but not infection. Thus, vaccinated individuals can become asymptomatically infected or susceptible to disease as vaccine-induced immunity wanes (τ) (Figure 3-1).

Figure 3-1. Model schematic of the movement between six states of norovirus infection. In the absence of vaccination, persons are born directly into the susceptible pool (S), become exposed at the force of infection ($\lambda(t)$), and then progress through the exposed (E), symptomatic (I) and asymptomatic (A) stages at rates inversely proportional to the duration of these states (μ, φ, ρ) before entering the recovered compartment (R). From the recovered compartment, persons can become asymptomatically infected at the force of infection or can become susceptible to disease through the waning of natural immunity (θ). In the presence of a pediatric vaccination (panel A), a proportion of births entering the system will receive protection from vaccines (v) and enter the vaccinated compartment (V). In the presence of elderly vaccination (panel B), a proportion of the elderly will receive protection from vaccines (v) and enter the vaccinated compartment (V). Only children under five and the elderly can flow into vaccinated compartments. Under both pediatric and elderly vaccine scenarios, vaccinated individuals can become asymptomatically infected at the force of infection or can become susceptible to disease through the waning of vaccine immunity (α).



After model fitting, we simulated routine, age-targeted vaccination of infants around the time of birth with vaccine coverage of 90% (i.e. Pediatric immunization) and individuals turning age 65 and every five years thereafter with vaccine coverage of 65% (i.e. Elderly immunization). Vaccine coverage for these scenarios was based on recent age-specific uptake of measles and influenza vaccines, respectively.[161,162] No vaccine efficacy (VE) estimates from field trails exist for norovirus vaccines, so we considered two different values, based on different interpretations of vaccine-challenge studies. These studies suggest monovalent or bivalent norovirus vaccination followed by a homotypic challenge reduces disease by approximately 50% among vaccinated individuals.[108,109] In low-efficacy vaccine scenarios (lVE Pediatric and Elderly) we assume only those immunologically susceptible to norovirus at the time of vaccine administration gain additional protection from disease. About 44% and 55% of 0-4 year olds are susceptible and

recovered, respectively, prior to vaccination, resulting in a VE around 25% for 0-4 year olds. In the elderly population, 85% are susceptible and 14% are recovered prior to vaccination, thus VE is approximately 43% for the elderly. Under a more optimistic high VE (hVE) scenario, we assume vaccination confers a 50% reduction in disease incidence over one year among vaccinated individuals.

For these four vaccine scenarios—and for a scenario without vaccination—we estimated age-specific incidence of disease and clinical outcomes. Analyses of long-term impacts of vaccination were conducted after the system had reached equilibrium, approximately 40 years after vaccine introduction. We simulated a well-established vaccination program with coverage similar to other vaccines. More detailed analysis are required to model the scale-up of vaccine coverage in the first few years after implementation and the associated epidemiological impacts. We calculated population direct and indirect effects of vaccination by comparing vaccine to no-vaccine simulations. We also assessed the efficiency of vaccine simulations, defined as the number of clinical outcomes averted per vaccinee (Text S1).

Parameters and Simulations

Parameter values (and ranges) for natural history, and clinical outcome probabilities were set to values identified in observational/challenge studies, and previous modeling studies (Table 3-1).

Table 3-1. Parameter input values, ranges tested in uncertainty analyses, and sources.

Parameter	Symbol	Input value	Range (+/- standard deviation)	Distribution	Source
Duration of incubation period	μ	32.8 hours	(30.9 – 34.6)	Uniform	Devasia et al 2014 [63]
Duration of symptomatic infectiousness	φ	48 hours	(38.9 – 50.7)	Uniform	Devasia et al 2014 [63]
Duration of asymptomatic infectiousness	ρ	10 days	(1-20)	Uniform	Atmar et. al 2008 [53]
Duration of natural immunity	θ	5.1 years	(4.0–6.7)	Uniform	Simmons et. al 2013 [91]
Relative infectiousness during incubation and asymptomatic period	ε	0.05	(0.045, 0.055)	Uniform	Simmons et. al 2013 [91]
Duration of vaccine asymptomatic infectiousness	α	10 days	(1-20)	Uniform	Assumed equal to duration of natural infection
Duration of vaccine immunity	τ	5.1 years	(4.0–6.7)	Uniform	Assumed equal to duration of natural immunity
OP admission probability					
0-4 years		0.168	(0.100–0.235)	Uniform	Bartsch et al 2012 [149]
5-17 years		0.168	(0.111–0.226)	Uniform	Bartsch et al 2012 [149]
18-64 years		0.06	(0.019–0.106)	Uniform	Bartsch et al 2012 [149]
65+ years		0.103	(0.063–0.143)	Uniform	Bartsch et al 2012 [149]
IP admission probability					
0-4 years		0.00428	+/- 0.000178	Normal	Bartsch et al 2012 [149]
5-17 years		0.00182	+/- 0.000074	Normal	Bartsch et al 2012 [149]
18-64 years		0.00228	+/- 0.000092	Normal	Bartsch et al 2012 [149]
65+ years		0.01733	+/- 0.000709	Normal	Bartsch et al 2012 [149]
Death probability					
0-4 years		0.00000625	+/- 2.57×10^{-7}	Normal	Bartsch et al 2012 [149]
5-17 years		0.00000466	+/- 1.81×10^{-7}	Normal	Bartsch et al 2012 [149]
18-64 years		0.00000466	+/- 1.81×10^{-7}	Normal	Bartsch et al 2012 [149]
65+ years		0.000435	+/- 0.000018	Normal	Bartsch et al 2012 [149]
ED visit probability					
0-4 years		0.0179	(0.0112-0.0246)	Uniform	Bartsch et al 2012 [149]
5-17 years		0.0199	(0.0114-0.0280)	Uniform	Bartsch et al 2012 [149]
18-64 years		0.026	(0.0153-0.0368)	Uniform	Bartsch et al 2012 [149]
65+ years		0.0325	(0.0199-0.0452)	Uniform	Bartsch et al 2012 [149]
Fitted Parameters					
Susceptibility of 0-4 year olds	q_1	0.208	(0.141, 0.402)[a]		Estimated

Susceptibility of 5-17 and 18-64 year olds	$q_{2,3}$	0.032	(0.023, 0.057) [a]		Estimated
Susceptibility of 65+ year olds	q_4	0.020	(0.014, 0.035) [a]		Estimated
Seasonal amplitude	β_1	0.034	(0.008, 0.089) [a]	-	Estimated
Seasonal offset	ω	2.147	(1.961, 2.266) [a]	-	Estimated

a. Range in the fitted value based on 1,000 random samples of the fixed parameter

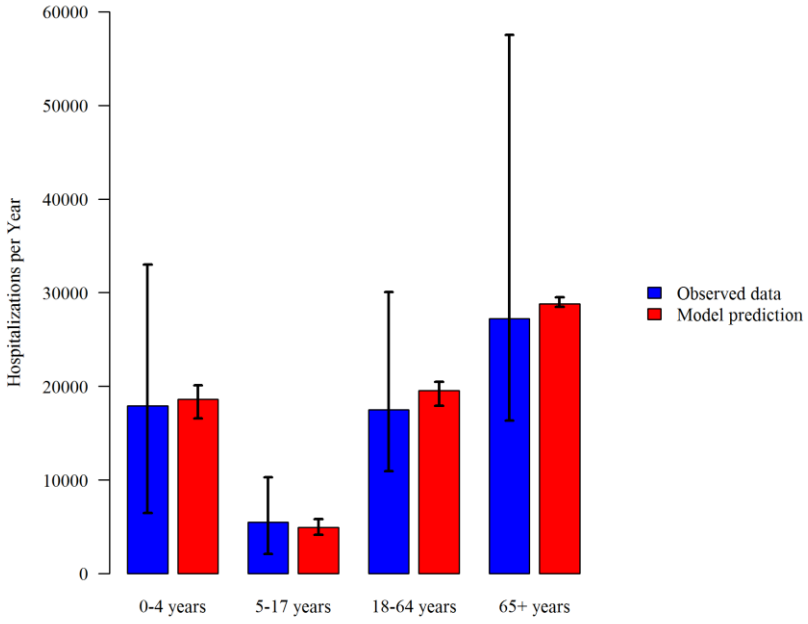
We used Latin hypercube sampling to generate 1,100 random samples of parameter sets and then re-fit the transmission probabilities and seasonality parameters, in the absence of vaccination, for each parameter set. We report the median and ranges of values for the fitted parameters based on the randomly sampled parameter sets (Table 3-1). We then ranked the 1,100 sampled and fitted parameter sets by their negative log likelihood (NLL) value. The 100 parameter sets with the highest NLL values were discarded, and we ran each vaccine scenario with the remaining 1,000 parameter sets. For summary statistics, we report medians and 2.5/97.5 percentiles of the annual clinical outcomes averted. These annual data are four year averages. In order to quantify the sensitivity of model projections to uncertainty in each parameter's value, we calculated partial rank correlation coefficients (PRCC) for natural history and vaccine parameters. PRCC values were calculated between model parameters and the percentage of cases averted in the total population and the age group targeted for vaccination in a given vaccination scenario.

RESULTS

Model Fitting

The best fit model based on the minimum Akaike information criterion (Table 3-S2) included three age-specific probabilities of infection on contact. The observed and predicted average annual norovirus hospitalizations were 71,461 and 71,906, respectively (Figure 3-2). A seasonal forcing of 3.4% (95% CI: 1.1%, 8.1%) of peak-to-mean amplitude provided the best fit to observed seasonal variation in monthly hospitalizations (Table 3-1, Figure 3-S2, Table 3-S2). In the best-fit model, 0-4 year olds contributed the most to transmission, with an age-specific basic reproduction number (R_0) of 4.3 compared to 1.4, 1.2, and 0.4 from 5-17, 18-64, and 65+ years, respectively (Text S1).

Figure 3-2. **Age-specific observed and predicted in best fitting model of hospitalizations per year in the United States.** The error bars in the observed data represent the range in annual hospitalizations over the 11 year data set. The error bars on the model data represent the range in annual hospitalizations based on the range in estimated and natural history parameter values identified in Table 3-1.

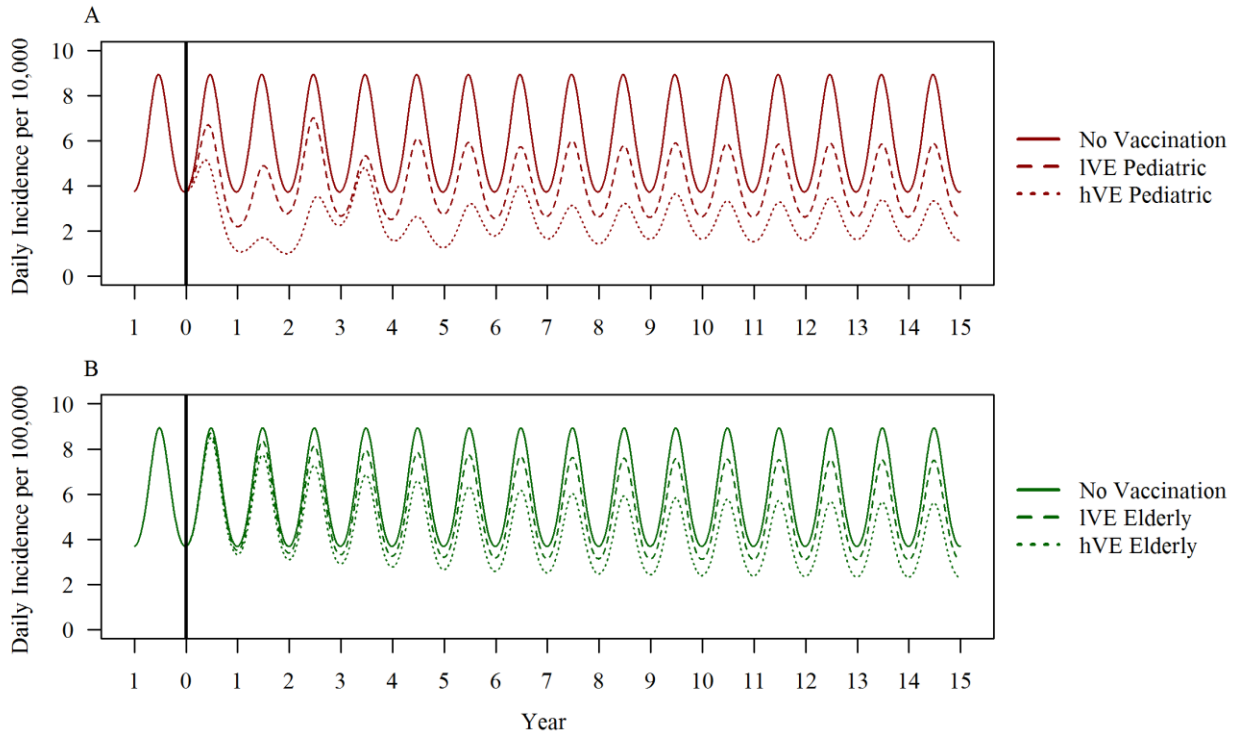


Vaccine Impact

The Pediatric programs rapidly reduced disease incidence in 0-4 year olds. Disease incidence exhibited inter-annual variability for the first several years before reaching lower equilibria (Figure 3-3A). In the first five years of the lVE Pediatric program, incidence among 0-4 year olds was reduced by 24%, 41%, 22%, 37%, and 30%. In the first five years of the hVE Pediatric program, incidence among 0-4 year olds was reduced by 42%, 78%, 59%, 46%, and 68%.

Elderly vaccination led to gradual reductions in disease incidence, achieving a new equilibrium of lower incidence in approximately 15 years (Figure 3-3B). In the first five years of the lVE Elderly program, incidence among the elderly was reduced by 3%, 6%, 9%, 11%, and 12%. In the first five years of the hVE Elderly program, incidence among the elderly was reduced by 5%, 13%, 19%, 23%, and 26%.

Figure 3-3. Predicted incidence of disease within the age-group targeted for vaccination over time. (A) Impact of the Pediatric vaccine programs on incidence of disease in 0-4 year olds. (B) Impact of the Elderly vaccine programs on incidence of disease in 65 year olds and older.



The lVE Pediatric program at equilibrium was predicted to avert 33% (95% CI: 27%, 40%) of all clinical outcomes in 0-4 year olds annually (Table 3-2, Figure 3-4A, Figure 3-5A). Approximately 71% of the averted outcomes were achieved through direct effects and 29% through indirect effects (Figure 3-4A). In older age classes, 14-16% of cases were averted primarily through indirect effects (Figure 3-4A, Figure 3-5A, Table 3-2).

Table 3-2. Outcomes averted (95% Confidence Interval) annually with a pediatric vaccine program with vaccine coverage of 90% and vaccine efficacy of 22% (lVE Pediatric).

Age Group	Cases Averted	Outpatients Averted	ED Visits Averted	Hospitalizations Averted	Deaths Averted
0-4 years	1,430,000 (1,185,000, 1,724,000)	237,000 (141,000, 356,000)	25,700 (15,700, 37,500)	6,200 (5,100, 7,400)	9 (7, 11)
5-17 years	419,000 (219,000, 705,000)	69,000 (32,000, 132,000)	8,100 (3,500, 16,400)	800 (400, 1,300)	2 (1, 3)
18-64 years	1,157,000 (701,000, 1,876,000)	70,000 (22,000, 160,000)	30,000 (14,900, 58,400)	2,600 (1,600, 4,300)	5 (3, 9)
65+ years	266,000 (173,000, 410,000)	27,000 (14,000, 49,000)	8,500 (4,600, 15,500)	4,600 (3,000, 7,000)	115 (75, 180)
Total (#)	3,282,000 (2,295,000, 4,720,000)	407,000 (259,000, 614,000)	72,400 (45,900, 115,700)	14,200 (10,100, 20,100)	132 (87, 202)
Total (%)	19% (13%, 27%)	22% (15%, 30%)	18% (12%, 26%)	20% (14%, 28%)	16% (11%, 25%)

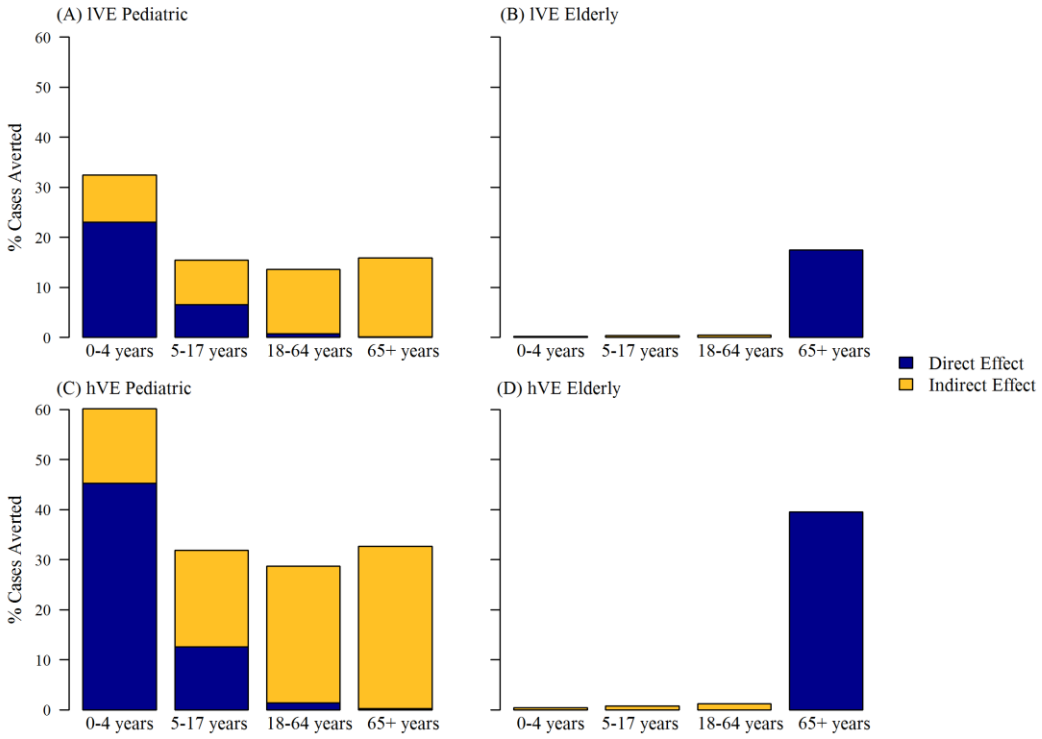
The hVE Pediatric program at equilibrium was predicted to avert 60% (95% CI: 49%, 71%) of all clinical outcomes in the 0-4 year olds annually (75% through direct and 25% indirect protection; Table 3-S3, Figure 3-4C, Figure 3-5C). In older age classes, 29-33% of cases were averted primarily through indirect protection (Table 3-S3, Figure 3-4C, Figure 3-5C).

A lVE Elderly program at equilibrium would avert approximately 17% (95% CI: 12%, 20%) of all clinical outcomes almost exclusively through direct effects in the elderly (Table 3-3, Figure 3-4B, Figure 3-5B). Minimal impacts were conferred on other age groups as less than 1% of outcomes in 0-64 year olds were averted through indirect effects (Table 3-3, Figure 3-4B, Figure 3-5B).

Table 3-3. Outcomes averted (95% CI) with routine elderly immunization with vaccine coverage of 65% and vaccine efficacy of 43% (lVE Elderly).

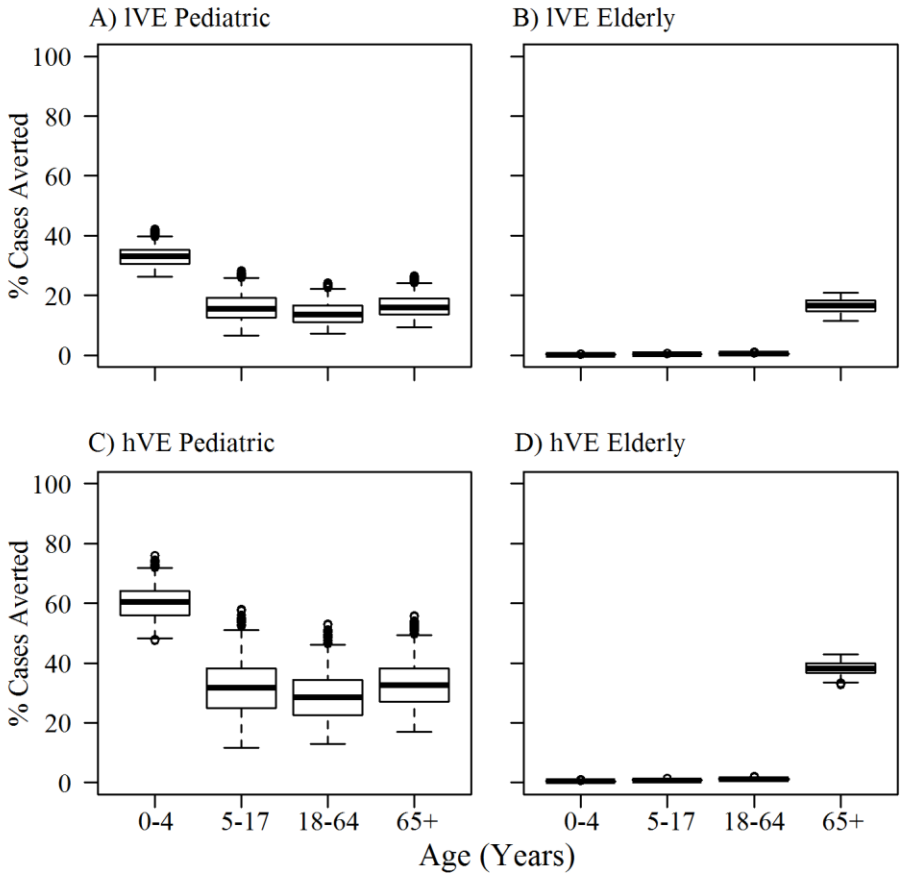
Age Group	Cases Averted	Outpatients Averted	ED Visits Averted	Hospitalizations Averted	Deaths Averted
0-4 years	8,500 (4,700, 14,800)	1,400 (630, 2,800)	150 (70, 290)	36 (19, 65)	0.05 (0.03, 0.09)
5-17 years	9,200 (5,400, 15,400)	1,500 (760, 2,900)	180 (80, 370)	17 (10, 28)	0.04 (0.03, 0.07)
18-64 years	42,300 (25,700, 65,900)	2,500 (810, 5,600)	1,100 (530, 2,100)	100 (60, 150)	0.20 (0.12, 0.31)
65+ years	276,900 (204,700, 344,400)	28,100 (16,100, 43,400)	8,800 (5,100, 13,700)	4,800 (3,600, 5,800)	120 (88, 151)
Total (#)	336,900 (240,300, 436,500)	33,800 (20,100, 50,900)	10,200 (6,100, 15,600)	4,900 (3,700, 6,000)	120 (88, 151)
Total (%)	1.9% (1.4%, 2.5%)	1.8% (1.1%, 2.9%)	2.5% (1.6%, 3.7%)	6.9% (5.1%, 8.5%)	15% (11%, 18%)

Figure 3-4. Direct (blue) and indirect effects (yellow) of each vaccine scenario. (A) Low vaccine efficacy (lVE) Pediatric program (B) Low vaccine efficacy (lVE) Elderly program (C) High vaccine efficacy (hVE) Pediatric program (D) High vaccine efficacy (hVE) Elderly program.



The hVE Elderly program at equilibrium was predicted to avert 38% (95% CI: 34%, 42%) of all clinical outcomes almost exclusively through direct effects (Table 3-S4, Figure 3-4D, Figure 3-5D). Minimal impacts were conferred on younger age groups, with approximately 1% or less of outcomes averted through indirect effects (Table 3-S4, Figure 3-4D, Figure 3-5D).

Figure 3-5. Boxplots representing the range of uncertainty in the percent of cases averted over a one year time period, given uncertainty in parameter input values. (A) Low vaccine efficacy (lVE) Pediatric program (B) Low vaccine efficacy (lVE) Elderly program (C) High vaccine efficacy (hVE) Pediatric program (D) High vaccine efficacy (hVE) Elderly program.



Pediatric programs were more efficient than Elderly programs. Per 100,000 vaccinees assuming lVE, the Pediatric program averted 21 times more cases; 26 times more OP visits; 15 times mores ED visits; 6 times more IP admissions; and twice as many deaths (Table 3-4) as Elderly programs. For hVE, the Pediatric programs averted 18 times more cases; 21 times more OP visits; 13 times more ED visits; 5 times more IP admissions; and twice as many deaths (Table 3-4) as Elderly

programs. For every one case averted through direct effects, 3 and 5 cases in the total population were averted through indirect effects with lVE and hVE Pediatric programs, respectively. For every one case averted through direct effects, 0.5 and 1 cases in the total population were averted through indirect effects with lVE and hVE Elderly programs, respectively.

Table 3-4. Clinical outcomes averted per 100,000 vaccinees (95% CI) over 1 year.

Vaccine strategy	Cases averted per 100,000 doses	Outpatient visits averted per 100,000 doses	ED visits averted per 100,000 doses	Hospitalizations averted per 100,000 doses	Deaths averted per 100,000 doses
lVE Pediatric					
0-4 years	39,500 (32,700, 47,600)	6,600 (3,900, 9,800)	710 (430, 1,040)	170 (140, 200)	0 (0, 0)
Total	90,600 (63,400, 130,300)	11,200 (7,100, 16,900)	2,000 (1,270, 3,200)	390 (280, 560)	4 (2, 6)
lVE Elderly					
65+ years	3,500 (2,600, 4,400)	360 (210, 560)	110 (66, 170)	61 (46, 74)	2 (1, 2)
Total	4,300 (3,100, 5,600)	430 (260, 650)	130 (80, 200)	63 (50, 80)	2 (1, 2)
hVE Pediatric					
0-4 years	72,400 (58,800, 87,000)	11,900 (7,100, 18,000)	1,280 (790, 1,920)	310 (250, 370)	0 (0, 1)
Total	178,000 (119,400, 256,200)	21,400 (13,000, 33,400)	3,940 (2,450, 6,340)	760 (530, 1,070)	7 (5, 11)
hVE Elderly					
65+ years	8,100 (7,100, 9,400)	840 (510, 1,200)	270 (160, 380)	141 (120, 160)	4 (3, 4)
Total	9,900 (8,400, 12,000)	1,010 (600, 1,400)	310 (200, 430)	145 (130, 160)	4 (3, 4)

Sensitivity Analysis

For the Pediatric vaccine programs, the duration of natural immunity (θ), duration of vaccine induced immunity (α), and the probability of infection on contact for 5-64 year olds ($q_{2,3}$) had the most influence on the percent of cases averted in the total population (Table 3-5). The most influential parameters on the percent of cases averted in the total population for the Elderly vaccine programs were the duration of natural immunity (θ), duration of vaccine induced immunity (α), and 65+ year olds (q_4) (Table 3-5).

When other parameters were fixed, the percentage of cases averted in the total population ranged from 15% (95% CI: 12%, 20%) to 23% (95% CI: 18%, 30%) and 28% (95% CI: 22%, 39%) to 45% (95% CI: 34%, 60%) across the tested range in duration of vaccine immunity (Table 3-1) for lVE and hVE Pediatric programs, respectively. The percentage of cases averted in the total population ranged from 1% (95% CI: 1%, 2%) to 2% (95% CI: 2%, 3%) and 4% (95% CI: 3%, 5%) to 5% (95% CI: 4%, 6%) across the tested range in duration of vaccine immunity for lVE and hVE Elderly programs, respectively.

Table 3-5. Partial rank correlation coefficients (PRCC) between selected model parameters and the percent of cases averted in the total population for each of four vaccination strategies.

	Symbol	lVE Pediatric	lVE Elderly	hVE Pediatric	hVE Elderly
Natural History Parameters					
Duration of Incubation	μ	0.04	0.06	0.03	0.07
Duration of Symptomatic Infection	φ	-0.19	0.11	-0.15	0.21

Duration of Asymptomatic Infectiousness	ρ	0.19	0.11	0.19	0.14
Duration of Natural Immunity	θ	-0.84 (2) [a]	-0.66 (3) [a]	-0.81 (2) [a]	-0.57 (3) [a]
Pre/Post Symptomatic Infectiousness	ε	-0.09	-0.02	-0.06	0.03
Transmissibility of 0-4 year olds	q_1	0.04	-0.35	0.10	-0.37
Transmissibility of 5-17 and 18-64 year olds	$q_{2,3}$	-0.41 (3) [a]	-0.41	-0.37 (3) [a]	-0.44
Transmissibility of 65+ year olds	q_4	0.35	0.82 (2) [a]	0.31	0.87 (2) [a]
Vaccine Parameters					
Duration of Vaccine Immunity	α	0.94 (1) [a]	0.97 (1) [a]	0.94 (1) [a]	0.90 (1) [a]
Duration of Vaccine Asymptomatic Infectiousness	τ	-0.27	-0.11	-0.25	-0.11

[a] The top 3 most influential parameters for each vaccine program are indicated by ranks in parentheses

DISCUSSION

Results from this transmission modeling study suggest the overall population impact of norovirus vaccination can vary substantially depending on the age group targeted. Pediatric programs offered the greatest reductions in all clinical outcomes, with 33% to 60% decreases among 0-4 year olds, and 14% to 33% reductions in older age groups achieved primarily through indirect protection. Pediatric programs were 18-21 times more efficient at preventing cases and 5-6 and two times more efficient at preventing IP admissions, and deaths, respectively, when compared to Elderly programs. Elderly programs averted between 17% and 38% of cases in the elderly, and provided

protection almost exclusively through direct effects. This is a result of the minimal contribution that the elderly make to disease transmission. In fact, Pediatric programs were predicted to confer similar benefits to the elderly as Elderly programs. Taken together, these results indicate targeting pediatric populations for vaccination leads to greater direct and indirect benefits for the total population than vaccine programs that target the elderly. Children under five have higher disease, OP, and ED admission rates; thus vaccines can directly prevent these outcomes. The indirect benefits of Pediatric programs are a result of reductions in disease transmission, owing to the importance of young children in transmission. This finding is consistent with observational studies that identified contact with a young child with norovirus gastroenteritis as a risk factor for diarrhea for older children and adults.[16,61] Large indirect benefits have been observed with the introduction of pediatric rotavirus and pneumococcal vaccines in the US, with unvaccinated populations protected through reductions in the overall force of infection.[163,164]

A second important finding was the identification of key parameters that influence the impact of vaccine strategies. For both Pediatric and Elderly vaccination programs, the duration of vaccine-induced immunity, and age-specific transmission parameters ($q_{2,3}$ and q_4 for Pediatric and Elderly programs, respectively) strongly determined the outcome of the analysis. These are parameters for which we have limited empirical data because transmission is largely unobservable,[121] and no vaccine studies have included long-term follow-up for clinical outcomes.[111] In order to better predict the impacts of norovirus vaccines in future work, this analysis highlights the high value of collecting information on transmission from observational studies and conducting clinical trials that can estimate the duration of protection for both children under five and the elderly.

The predicted impacts of lVE scenarios were modest due to technical reasons related to our model construction. First, we assume vaccination only provides additional protection for those who are susceptible to disease (in the S class) at the time of immunization. Individuals who have acquired natural immunity will receive no added protection. This assumption strongly limits the impact of vaccination for adults as many will have acquired immunity, whereas we assume all children are susceptible at the time of infant immunization. A second explanation is that we assume exponential waning of both natural and vaccine immunity, so while the average duration of protection 5.1 years, most individuals have a shorter-duration immunity while a few have longer-term protection. Compartmental models can be modified to assume other distributions for waning immunity; however no data are available to inform the functional form of waning immunity to norovirus. The hVE scenarios which were based on a 50% vaccine efficacy[108,109] resulted in more optimistic impacts. The values and concepts of vaccine action that are most appropriate can be informed by future clinical trial data.

There are several limitations to this study. First, there is uncertainty in the robustness of the epidemiological data used to fit the model. We used US hospitalization data, which are model estimates, and community incidence rates were informed by a UK study. In that study, incidence in older age groups was low and may have been biased downwards.[16] Fitting to such low incidence limited the potential impact of elderly immunization in our model and limited the role of elderly people in transmission. Second, there is considerable uncertainty in model parameters due to our limited understanding of natural history of norovirus disease and transmission, particularly the relative roles of pre- and post-symptomatic transmission. Third, our model construction assumes a single strain of norovirus; thus infection from, or vaccination against, one strain of norovirus

provides protection against all other infections. This is a major simplification, as noroviruses are highly genetically diverse and natural immunity provides only limited cross-protection within genogroups.[66] However, this simplification may be partially accounted for by the duration of immunity parameter, particularly in the low vaccine efficacy scenario. In addition, novel genogroup 2 type 4 (GII.4) strains emerge every two to four years, that may evade host population immunity. Current data are insufficient to establish the degree of cross-protection to norovirus, or to parameterize a multi-strain model as has been accomplished for influenza.[165] For a more complete understanding of norovirus transmission and vaccination, these are important areas for further empirical studies and, subsequently, model development. Additionally, we did not consider a model that incorporated a class of individuals that are genetically resistant to norovirus infection. As more data become available on the effect of vaccination among genetically resistant individuals, future modeling studies should consider such a class. Another important limitation is that we assumed that VE was the same for all clinical outcomes and disease severity. This may not be the case. For rotavirus, VE is greater for severe outcomes.[164] Finally, while we developed a model to predict the impact of infant and elderly vaccination, there have been no studies of VE in pediatric populations and only one immunogenicity study in the elderly.[111] Human safety, immunogenicity and efficacy studies have all involved experimental challenge of adults (typically 18-49 years old).[108,109,112] While these results are promising, clinical trials will be pivotal in determining VE among infants and the elderly. Though our study made several simplifying assumptions—as all models do—the dynamic transmission framework presented here offers a more comprehensive understanding of total population benefits of vaccination than previous studies that included only direct effects.[149]

In summary, our results quantitatively demonstrate that the potential public health value of a norovirus vaccine is likely greatest with pediatric immunization. This finding argues for a clinical development plan for a vaccine with a safety and efficacy profile suitable for use in children. To improve models for future analyses, better data are needed on the duration of natural and vaccine immunity, the extent of cross-protection and process of norovirus infection. Future modeling studies should incorporate norovirus strain diversity to examine the implications of multiple, evolving strains for vaccination. As more data become available on the extent of cross-protection and the duration of vaccine immunity, this modeling framework can be adapted to more precisely estimate population-level impacts of norovirus vaccination. Models should also be adapted to developing world settings where the force of infection is higher [166] and disease burden is greater.[4]

Supplemental

Text S1.

Detailed Model Description: The ordinary differential equations that support the model structure as shown in Fig 1 are as follows:

Where k =1 represents 0-4 year olds

$$\frac{dS_k}{dt} = B(1-v) + \theta R_k + \alpha V_k - (\lambda_i(t) + a_k + D_k)S_k$$

$$\frac{dE_k}{dt} = \lambda_i(t)S_k - (\mu + a_k + D_k)E_k$$

$$\frac{dI_k}{dt} = \mu E_k - (\varphi + a_k + D_k)I_k$$

$$\frac{dA_k}{dt} = \varphi I_k + \lambda_i(t)R_k - (\rho + a_k + D_k)A_k$$

$$\frac{dR_k}{dt} = \rho A_k - (\lambda_i(t) + \theta + a_k + D_k)R_k$$

$$\frac{dV_k}{dt} = Bv + \tau V a_k - (\lambda_i(t) + \alpha + a_k + D_k)V_k$$

$$\frac{dV_{ak}}{dt} = \lambda_i(t)V_k - (\tau + a_k + D_k)Va_k$$

Where k = 2, 3 represent 5-17 and 18-64 year olds, respectively

$$\frac{dS_k}{dt} = a_{k-1}S_{k-1} + \theta R_k + \alpha V_k - (\lambda_i(t) + a_k + D_k)S_k$$

$$\frac{dE_k}{dt} = a_{k-1}E_{k-1} + \lambda_i(t)S_k - (\mu + a_k + D_k)E_k$$

$$\frac{dI_k}{dt} = a_{k-1}I_{k-1} + \mu E_k - (\varphi + a_k + D_k)I_k$$

$$\frac{dA_k}{dt} = a_{k-1}A_{k-1} + \varphi I_k + \lambda_i(t)R_k - (\rho + a_k + D_k)A_k$$

$$\frac{dR_k}{dt} = a_{k-1}R_{k-1} + \rho A_k - (\lambda_i(t) + \theta + a_k + D_k)R_k$$

$$\frac{dV_k}{dt} = a_{k-1}V_{k-1} + \tau Va_k - (\lambda_i(t) + \alpha + a_k + D_k)V_k$$

$$\frac{dV_{ak}}{dt} = a_{k-1}Va_{k-1} + \lambda_i(t)V_k - (\tau + a_k + D_k)Va_k$$

Where k = 4…8 represent 65-69, 70-74, 75-79, 80-84 and 85+ year olds, respectively

$$\frac{dS_k}{dt} = a_{k-1}S_{k-1}(1-v) + \theta R_k + \alpha V_k - (\lambda_i(t) + a_k + D_k)S_k$$

$$\frac{dE_k}{dt} = a_{k-1}E_{k-1}(1-v) + \lambda_i(t)S_k - (\mu + a_k + D_k)E_k$$

$$\frac{dI_k}{dt} = a_{k-1}I_{k-1}(1-v) + \mu E_k - (\varphi + a_k + D_k)I_k$$

$$\frac{dA_k}{dt} = a_{k-1}A_{k-1}(1-v) + \varphi I_k + \lambda_i(t)R_k - (\rho + a_k + D_k)A_k$$

$$\frac{dR_k}{dt} = a_{k-1}R_{k-1}(1-v) + \rho A_k - (\lambda_i(t) + a_k + \theta + D_k)R_k$$

$$\frac{dV_k}{dt} = a_{k-1}V_{k-1}(1-v) + v(a_{k-1}(S_{k-1} + E_{k-1} + I_{k-1} + A_{k-1} + R_{k-1} + V_{k-1} + Va_{k-1}) + \tau Va_k - (\lambda_i(t) + a_k + \alpha + D_k)V_k$$

$$\frac{dV_{ak}}{dt} = a_{k-1}Va_{k-1}(1-v) + \lambda_i(t)V_k - (\tau + a_k + D_k)Va_k$$

Where

S_k= Susceptible to infection

E_k = Exposed to infection

I_k = Infected with symptoms

A_k = Infected, asymptomatic

R_k = Recovered, immune to disease but not infection

V_k = Vaccinated, immune to disease but not infection

Va_k = Vaccinated, asymptomatic

B = births entering the system

a_k = rate of individuals aging out the group; where $k = 8$, $a_k = 0$

D_k = proportion deaths from age group k exiting the system

θ = rate of waning natural immunity

α = rate of waning vaccine immunity

v = proportion of individuals protected by vaccination represented as the vaccine coverage multiplied by the vaccine efficacy.

$\lambda(t)$ = the force of infection

μ = rate of moving out of exposed

φ = rate of symptomatic infection

ρ = rate of asymptomatic infectious period

τ = rate of vaccinated asymptomatic infectious period

The force of infection was modeled as:

$$\lambda_i(t) = b_i q_i \sigma(t) \sum_{j=1}^{4} c_{ij} \frac{(\varepsilon E_j + I_j + \varepsilon A_j)}{N_j}$$

where b_i is the total number of contacts an individual of age group i makes in a day, q_i represents the age-specific transmissibility of norovirus, c_{ij} is the proportion of contacts that members of age group i make with age group j, N_j is the population size of age group j, and ε represents how infectious exposed and asymptomatic individuals are relative to symptomatically infected individuals. When j=4, values for E, I, A and N are summed across the 5 elderly age groups (65-69, 70-74, 75-79, 80-84 and 85+ year olds). As norovirus exhibits a strong seasonal pattern in the United States, the model incorporates the effect of seasonality ($\sigma(t)$):

$$\sigma(t) = 1 + \beta_1 \times \cos(2\pi t + \omega)$$

where β_1 is the amplitude of the seasonal fluctuation and ω is the seasonal offset parameter.

R Packages: Model simulation and fitting were conducted in R version 3.1.1 using the *nloptr* and *deSolve* packages.[158,167,168] We used the *lhs* package in R to conduct Latin hypercube sampling to generate random samples of parameters given the ranges and distributions specified for each.[169] For our sensitivity analysis (described in more detail below) we calculated partial rank correlation coefficients (PRCC) using the *sensitivity* package in R.[170]

Alternative Model Structures: Other model structures were considered (Table 3-S1) that implement a different number and representation of the age-specific probability of infection given contact (q_i or q_j). In the model structure selected for analysis (see model 4 in Table 3-S1), q_i represents the probability of transmission and is dependent on the age group (i) of the susceptible individual. Alternatively, q_j may also be written to represent the probability of transmission dependent on the age group (j) of the infected person (see models 2 and 3 in Table 3-S1). Each alternative model structure produced different estimates for q_i and q_j. Particularly, the two alternative models that represented q_j as the probability of transmission dependent on the age group of the infected person, predicted that 0-4 year olds were highly infectious to older age groups ($q_1 \approx 0.3$); however, all older age groups were minimally infectious (see models 2 and 3 in Table 3-S1). As it is highly unlikely that only 0-4 year olds contribute to the transmission of norovirus,[91] these alternative model structures were rejected as implausible. Of the two models that represent q_i as the probability of transmission is dependent on the age group i of the susceptible individual,

the model that used four transmission probabilities (q_1, $q_{2,3}$, q_4), where the probabilities of infection on contact for 5-17 and 18-64 year olds were equal ($q_2 = q_3$), exhibited a better fit to the observed data than the model with two transmission probabilities (q_1, q_2) (see models 1 and 4 in Table 3-S1). Thus the model with three different age-specific transmission probabilities (q_1, $q_{2,3}$, q_4) (see model 4 in Table 3-S1), represented as the susceptibility of age group *i* to infections from all other age groups, was selected for further analysis.

Table 3-S1. Alternative models exploring different numbers and interpretations of age specific transmission probabilities (q_i).

Model No.	Model Description	Force of infection	No. estimated parameters	Estimated values	Negative Log(L)	AIC
1	The force of infection is dependent on age specific transmissibility parameters (q_1= 0-4 years and q_2=5+ years). These q parameters determine what proportion of infectious contacts from the community an individual is susceptible to	$\lambda_i(t) = b_i q_i \sum_{j=1}^{4} c_{ij}\, \sigma(t) \frac{(\varepsilon E_j + I_j + \varepsilon A_j)}{N_j}$	4	β_1=0.060 ω =2.169 q_1 =0.262 q_2 =0.032	289835.0	579679.9
2	The force of infection is dependent on age specific transmissibility parameters (q_1= 0-4 years and q_2=5+ years). These q parameters determine what proportion of infectious contacts from the community are successful	$\lambda_i(t) = b_i \sum_{j=1}^{4} q_j c_{ij}\, \sigma(t) \frac{(\varepsilon E_j + I_j + \varepsilon A_j)}{N_j}$	4	β_1= 0.049 ω = 2.361 q_1= 0.287 q_2= 1.71x10^{-8}	298886.9	597781.9
3	The force of infection is dependent on age specific transmissibility parameters (q_1= 0-4 years, $q_{2,3}$=5-64 years, q_4=65+ years). These q parameters determine what proportion of infectious contacts from the community are successful	$\lambda_i(t) = b_i \sum_{j=1}^{4} q_j c_{ij}\, \sigma(t) \frac{(\varepsilon E_j + I_j + \varepsilon A_j)}{N_j}$	5	β_1=0.046 ω =2.246 q_1=0.282 $q_{2,3}$=8.694 x10^{-4} q_4=2.888x10^{-9}	299138.1	598286.3
4	The force of infection is dependent on age specific transmissibility parameters (q_1= 0-4 years, $q_{2,3}$=5-64 years, q_4=65+ years). These q	$\lambda_i(t) = b_i q_i \sum_{j=1}^{4} c_{ij}\, \sigma(t) \frac{(\varepsilon E_j + I_j + \varepsilon A_j)}{N_j}$	5	β_1 =0.034 ω = 2.146 q_1= 0.208	239847.7	479705.6

parameters determine what proportion of infectious contacts from the community an individual is susceptible to	$q_{2,3}$= 0.032 q_4= 0.020

Contact Structure: Lacking detailed mixing data specific to the US, we used POLYMOD data[151] describing epidemiologically relevant contact patterns from representative samples of eight European countries. From these data we determined the average total number of all reported contacts (both physical and non-physical) made by individuals in each age group, and created a contact matrix with age-specific proportions of contacts (Table 3-S2). Several steps were taken to obtain the proportion of contacts made by age group *i* with age group *j* (c_{ij}). Raw numbers of the age-specific daily number of contacts from eight European countries (Belgium, Germany, Finland, Great Britain, Italy, Luxembourg, The Netherlands, and Poland) were collected from the POLYMOD study.[151] Next, these contact data were summed into the four defined age groups of this study (0-4 years, 5-17 years, 18-64 years, and 65+ years). To obtain rates of contact (R_{ij}), the counts of contacts (C_{ij}) were divided by the number of participants in each contact group (N_i):

$$R_{ij} = C_{ij}/N_i \quad (1)$$

As the raw data for contacts made between age groups *i* and *j* are unlikely to be symmetric due to reporting errors (meaning each contact between an individual in age group *i* and an individual in age group *j* is recorded by both the individual in age group *i* and the individual in age group *j*), the contact rates were corrected for differences in reporting by different age groups with the following equation described by Eames et. al.:[171]

$$B_{ij} = (N_i * R_{ij} + N_j * R_{ji})/2N_i \quad (2)$$

Where N_i is the number of individuals in age group i, N_j is the number of individuals in age group j, and R_{ij} is the contact rate of age group age group i with age group j. $2N_j$ assumes that contacts of individuals within the sample (N_i) made contacts with individuals outside of the sampled population. Finally, the following equation was used to obtain the proportion of contacts made by age group i with age group j (c_{ij}):

$$c_{ij} = B_{ij}/a_i \quad (3)$$

Where B_{ij} represents the corrected contact rates (see equation 2) and a_i represents the sum over all age groups (Table 3-S2). The total number of contacts each age group makes in one day are as follows:

0-4 years (a_1) = 8.12

5-17 years (a_2) = 15.52

18-64 years (a_3) = 13.94

65+ years (a_4) = 8.53

Table 3-S2. Proportion of contacts made by age group *i* (rows) with age group *j* (columns).

	0-4 years	**5-17 years**	**18-64 years**	**65+ years**
0-4 years	0.276	0.180	0.506	0.039
5-17 years	0.034	0.633	0.313	0.019
18-64 years	0.046	0.151	0.753	0.050
65 + years	0.044	0.116	0.616	0.225

Age-Specific R_0 Calculation: A detailed description of how R_0 was calculated can be found in the technical appendix from Simmons et. al., 2013.[91] Briefly we constructed a Next Generation Matrix (NGM) for our model system and calculated the eigenvalues of this matrix. R_0 is then equal to the largest eigenvalue, or spectral radius, of the matrix. The age-specific R_0 values are the row sums of the NGM, and represent the number of cases in any age group generated by an individual in age-group *i* in one generation at the beginning of the epidemic.

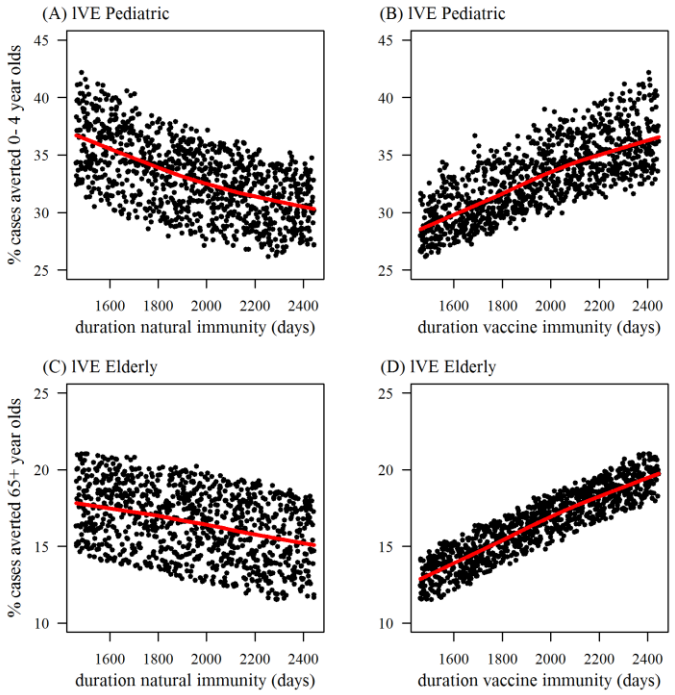
Population Direct and Indirect Effect Calculation: Total effects of vaccination were calculated by dividing the number of clinical outcomes averted under a given vaccine scenario by the number of clinical outcomes that occurred without vaccination. The population direct effects of vaccination were calculated using a formula described by Pitzer et. al.:[172]

$$DE_{i,k} = \frac{\sum_{d=1}^{365} v_{ik} x_{i,d_{pv}}}{\sum_{d=1}^{365} x_{i,d_{pv}}}$$

where $v_{i,k}$ is the average proportion of individuals with vaccine immunity (from time period d=1 to d=365) in age group i under vaccine scenario k, and $x_{i,dpv}$ is the number of norovirus cases in age group i during a pre-vaccination day d_{pv}. Our model results indicate there are population direct effects among older age groups under pediatric immunization programs. There are two explanations for this result. First, the structure of the model allows an exponential flow of aging; thus a small number of individuals with vaccine protection will flow into older age groups more quickly. Second, cycling between vaccine asymptomatic and vaccine protected states can extend the duration which individuals remain protected by vaccines. Indirect effects of vaccination were calculated by subtracting population direct effects from overall effects.

Sensitivity Analysis: To quantify the sensitivity of model projections to uncertainty in each parameter's value, we calculated partial rank correlation coefficients (PRCC) for natural history and vaccine parameters using the *sensitivity* package in R.[170] We used scatterplots to determine the relationships between the model results and parameter values were monotonic, thus confirming PRCC would provide meaningful results (Figure 3-S2). PRCC values were calculated between model parameters and the percentage of cases averted in the (a) total population and (b) the age group targeted for vaccination in a given vaccination scenario.

Figure 3-S1. Scatterplots, with LOWESS regression lines, of the correlation of percent cases averted in targeted age group and the durations of natural and vaccine immunity (θ, α). (A) lVE Pediatric program (B) lVE Pediatric program (C) lVE Elderly program (D) lVE Elderly program.



For the Pediatric vaccine programs, the duration of natural immunity (θ), duration of vaccine induced immunity (α), and probability of infection on contact for 0-4 year olds (q_1) affected the percent of cases averted in 0-4 year olds the most (Table 3-S5). For the Elderly vaccine programs, the most influential parameters on the percent of cases averted in 65 years and older the duration of natural immunity (θ), duration of vaccine induced immunity (α), and probability of infection on contact for 65+ year olds (q_4) (Table 3-S5).

Figure 3-S2. Predicted vs. observed seasonal variation in total hospitalizations per month. The grey band represents the range in the monthly number of hospitalizations from 1996-2007. The red line is the model estimated monthly number of hospitalizations. The dashed lines are hospitalization data from years (2002 and 2006) where GII.4 strains emerged, resulting in largescale epidemics. The considerable variability in the timing and magnitude of the norovirus season from year to year makes it challenging for a simple model to capture.

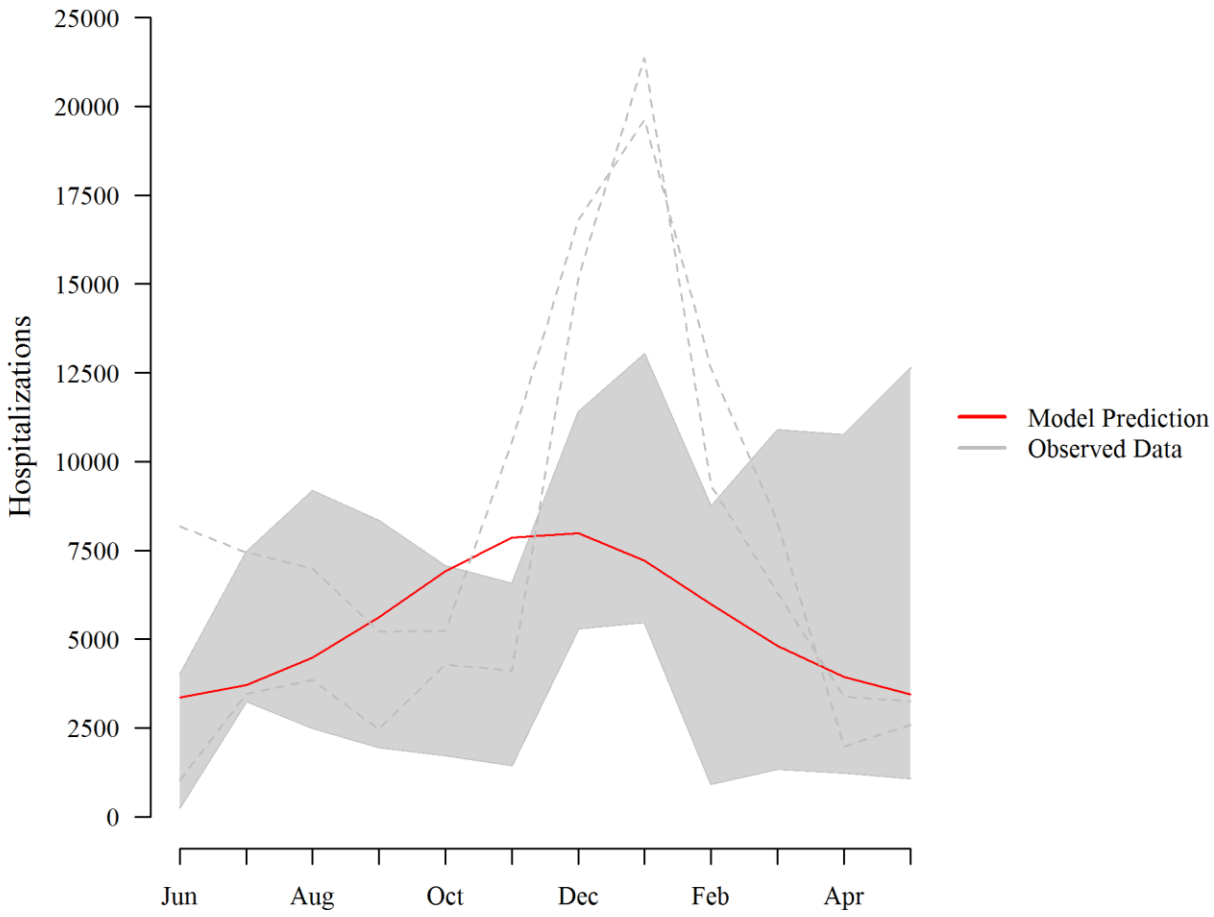


Table 3-S3. Clinical outcomes averted annually (95% CI) with a hVE Pediatric program.

Age Group	Cases Averted	Outpatients Averted	ED Visits Averted	Hospitalizations Averted	Deaths Averted
0-4 years	2,621,000 (2,128,000, 3,149,000)	429,000 (259,000, 654,000)	46,400 (28,800, 69,400)	11,200 (9,200, 13,300)	16 (13, 20)
5-17 years	850,000 (420,000, 1,414,000)	139,000 (61,000, 268,000)	15,900 (6,600, 33,300)	1,500 (800, 2,600)	4 (2, 7)
18-64 years	2,433,000 (1,398,000, 3,948,000)	144,000 (42,000, 330,000)	61,300 (29,300, 120,100)	5,600 (3,100, 9,100)	11 (6, 19)
65+ years	536,000 (331,000, 831,000)	54,000 (27,000, 100,000)	16,900 (8,700, 31,200)	9,400 (5,800, 14,200)	233 (144, 366)
Total (#)	6,444,000 (4,323,000, 9,277,000)	775,000 (471,000, 1,211,000)	142,600 (88,600, 229,500)	27,700 (19,100, 38,900)	264 (166, 410)
Total (%)	37% (25%, 53%)	42% (28%, 58%)	36% (23%, 52%)	39% (26%, 54%)	33% (21%, 50%)

Table 3-S4. Clinical outcomes averted annually (95% CI) with a hVE Elderly program.

Age Group	Cases Averted	Outpatients Averted	ED Visits Averted	Hospitalizations Averted	Deaths Averted
0-4 years	19,600 (12,200, 32,100)	3,300 (1,640, 6,200)	350 (170, 650)	85 (50, 142)	0.12 (0.08, 0.20
5-17 years	21,400 (14,200, 33,300)	3,500 (1,910, 6,400)	410 (210, 810)	39 (25, 62)	0.10 (0.07, 0.16)
18-64 years	97,300 (68,100, 141,900)	6,100 (1,890, 12,500)	2,500 (1,340, 4,600)	220 (150, 330)	0.45 (0.32, 0.68)
65+ years	636,200 (551,000, 731,500)	65,500 (39,500, 94,600)	20,700 (12,600, 29,700)	11,000 (9,800, 12,200)	276 (235, 324)
Total (#)	775,900 (652,400, 924,200)	78,800 (49,700, 112,700)	24,000 (15,600, 33,900)	11,400 (10,000, 12,700)	277 (236, 325)
Total (%)	4.5% (3.7%, 5.4%)	4.2% (2.7%, 6.4%)	5.9% (3.8%, 8.7%)	6.9% (13.9%, 17.7%)	35% (30%, 38%)

Table 3-S56. Partial rank correlation coefficients (PRCC) between selected model parameters and the percent of cases averted in the age group targeted for vaccination for each of four vaccination strategies.

	Symbol	lVE Pediatric	lVE Elderly	hVE Pediatric	hVE Elderly
Natural History Parameters					
Duration of Incubation	μ	0.02	0.03	-0.003	0.03
Duration of Symptomatic Infection	φ	-0.10	-0.05	-0.07	-0.02
Duration of Asymptomatic Infectiousness	ρ	0.14	-0.03	0.18	-0.01
Duration of Natural Immunity	θ	-0.72 (2) [a]	-0.86 (2) [a]	-0.70 (2) [a]	-0.80 (2) [a]
Pre/Post Symptomatic Infectiousness	ε	-0.13	-0.10	-0.09	-0.10
Transmissibility of 0-4 year olds	q_1	-0.55 (3) [a]	-0.11	-0.40 (3) [a]	-0.11
Transmissibility of 5-17 and 18-64 year olds	$q_{2,3}$	-0.07	-0.18	-0.10	-0.22
Transmissibility of 65+ year olds	q_4	0.33	0.24 (3) [a]	0.29	0.35 (3) [a]
Vaccine Parameters					
Duration of Vaccine Immunity	α	0.97 (1) [a]	0.99 (1) [a]	0.97 (1) [a]	0.99 (1) [a]
Duration of Vaccine Asymptomatic Infectiousness	τ	-0.24	-0.05	-0.27	-0.06

[a]The top 4 most influential parameters for each vaccine program are indicated by ranks in parentheses

4 Manuscript for Aim 2

Characterizing Norovirus Transmission from Outbreak Data in the United States

Molly K. Steele, Mary E. Wikswo, Aron J. Hall, Katia Koelle, Andreas Handel, Karen Levy, Lance Waller, Ben A. Lopman

ABSTRACT

Norovirus is the most common cause of outbreaks of acute gastroenteritis (AGE) in the US. The size and severity of outbreaks varies across different settings, times of year and for different genotypes, suggesting that the transmissibility of norovirus may be variable across different outbreak contexts. We estimated the basic (R_0) and effective (R_e) reproduction numbers for 7,094 norovirus outbreaks reported to the National Outbreak Reporting System (NORS) in the United States between 2009 and 2017 and used regression models to assess whether transmission varied by outbreak setting. We estimated the median R_0 and R_e to be 2.75 (IQR: 2.38, 3.65) and 1.29 (IQR: 1.12, 1.74), respectively. Compared to our referent (outbreaks with confirmed norovirus etiology in long-term care/assisted living facilities in the south, during winter of the July 2016 – June 2017 norovirus season; R_0=3.35; 95% confidence interval [CI]: 3.26, 3.45), outbreaks in schools and universities had a lower average predicted R_0 (2.92; 95% CI: 2.81, 3.03) and outbreaks in hospitals and other healthcare facilities had a marginally lower average predicted R_0 (3.13; 95% CI: 2.97, 3.30). Outbreaks in summer had lower average predicted R_0 (3.11; 95% CI: 2.95, 3.27)

than the referent group. These results suggest that elderly populations in long-term care/assisted living facilities could be targeted for vaccination to reduce the disease burden in these settings.

INTRODUCTION

Norovirus is the most common cause of outbreaks of acute gastroenteritis (AGE) in the US. The Centers for Disease Control and Prevention (CDC) collects data on AGE outbreaks through the National Outbreak Reporting System (NORS).[101,173] From 2009 to 2017, norovirus was reported as the suspected or confirmed etiology of 47% of AGE outbreaks reported to NORS.[102] The size and severity of outbreaks varies across different settings, times of year and for different genotypes, suggesting that the transmissibility of norovirus may be variable across different outbreak contexts.[103] Generally, the transmission potential of infectious diseases is influenced by the infectiousness of the pathogen, the duration of infectiousness and the number of susceptible contacts exposed during the infectious period.[174]

The reproduction number is a metric to quantify transmissibility of a pathogen. The basic reproduction number (R_0) is defined as the average number of secondary cases that arise from a primary case in a completely susceptible population. The effective reproduction number (R_e) quantifies the average number of secondary cases that arise from a primary case in a population that is not necessarily completely susceptible. The effective reproduction number changes through the course of an outbreak as the proportion of the population that is susceptible changes.[152,153] The basic and effective reproduction numbers of norovirus have been estimated from several transmission modeling studies; however there is large variation in these estimates (range of

estimated R_0: 1.1 to 7.2).[60] As reviewed by Gaythorpe et al., 2018, much of the variation in these estimates of R_0 across studies is due to differences in the structures, assumptions and data between transmission models.[60] Across these studies, transmission models were either deterministic, population-level models or models of norovirus outbreaks. Generally, the population-level models produced R_0 estimates of about 2, while the estimates from outbreak-level models tended to be higher and more variable. The variability of estimates from outbreak-level models is likely driven by the context specific nature of the data informing these models; outbreaks may occur in populations that are not representative of the population as a whole.[103] Gaythorpe et al., 2018 also reported that differences in assumptions regarding the contribution of asymptomatic shedding to transmission, duration of shedding and duration of immunity were drivers of the variation in estimates of R_0 across studies. These factors are challenging to parameterize as transmission, both from symptomatic and asymptomatic individuals, is largely unobservable, the duration of shedding is highly variable and there are conflicting estimates of the duration of immunity (short vs. long-term).[52,53,55,84,89,91,121]

In this study we investigated whether transmission of norovirus varies across different outbreak contexts. However, given the challenges of parameterizing key aspect of the natural history of norovirus for transmission models, which is the method that has most often been used to estimate R_0 and R_e of norovirus, we estimated R_0 and R_e of norovirus using an alternative method: the final epidemic size equation.[175,176] This method does not require information on the transmission chain (i.e., who infects whom) nor does it require assumptions about the duration of incubation, infection, shedding or immunity. To this end, we estimated R_0 and R_e using the final size method for thousands of norovirus outbreaks in the United States and then evaluated whether reproduction

numbers were associated with setting, season, year, geographic region and whether norovirus was suspected or confirmed (defined as two or more laboratory confirmed cases) as the cause of the outbreak.

METHODS

Data

We obtained data from NORS and CaliciNet on all norovirus outbreaks (defined as two or more cases of suspected or laboratory confirmed norovirus that are epidemiologically linked) that occurred between 2009 and 2017. NORS data consist of web-based reports of all foodborne, waterborne, and enteric disease outbreaks transmitted by contact with environmental sources, infected persons or animals, or unknown modes of transmission reported by state, local and territorial public health agencies. This web-based reporting system collects epidemiological information including the dates, setting (e.g., long-term care facility, child daycare, hospital, schools), and geographic location of the outbreak, as well as the estimated total number of cases and exposed population.[173] CaliciNet data consist of sequence-derived genotypes and epidemiological data from norovirus outbreaks submitted from local, state and federal public health laboratories. We obtained CaliciNet genotypes that were linked to outbreak data we acquired from NORS.

The total number of laboratory-confirmed and suspected primary cases, which excludes cases associated with secondary spread of illness (e.g., person-to-person transmission of norovirus in

households after a restaurant-based outbreak), is collected for all outbreaks reported to NORS. However, data for calculating attack rates, specifically the number of exposed persons and the subset of the exposed persons who became ill, are only collected for outbreaks with person-to-person, environmental, or unknown transmission modes. Between 2009 and 2017 there were 17,822 suspected and confirmed norovirus outbreaks reported to NORS. We excluded 10,728 outbreaks based on the following criteria (imposed hierarchically): transmission was not person-to-person (n = 3,866), the outbreak exposure occurred in multiple states (n = 8) or Puerto Rico (n = 3), the total exposed population size or major setting were not reported (n = 5,573), the total estimated primary cases and the total ill among the exposed population were not equal (n = 1,231), or the total estimated primary cases or the total ill among the exposed population were reported to be greater than the total exposed population size (n = 47) (Figure 4-S1). We therefore used 7,094 norovirus outbreaks meeting our inclusion criteria in subsequent analyses.

Estimating basic and effective reproduction numbers

We calculated the basic and effective reproductive numbers, and their associated standard errors (SE), for each outbreak using equations proposed by Becker that use the final epidemic size:[176]

$$R_0 = \frac{N-1}{C} \sum_{i=S-C+1}^{S} \frac{1}{i}$$

$$SE(R_0) = \frac{N-1}{C} \sqrt{\sum_{i=S-C+1}^{S} \left(\frac{1}{i^2} + \frac{CR_0^2}{(N-1)^2} \right)}$$

where *N* is the total population size, *C* is the total number of cases in the outbreak, and *S* is the number of susceptible individuals at the start of the outbreak. R_e is calculated by replacing *N* with *S* in the first equation above. SE(R_e) is calculated by replacing *N* with *S* and R_0 with R_e in the second equation shown above. The final size method assumes a susceptible-infected-recovered (SIR) type infection with a closed, homogenously mixing population.[176] *C* and *N* from the equations above were informed by NORS outbreak data on the estimated total number ill and exposed population, respectively. The number of susceptible individuals at the start of an outbreak (*S*) is a variable that is not observed nor can it be informed by NORS data; therefore we used norovirus challenge study data on the percent of individuals that become infected and develop AGE after challenge with virus to estimate *S*. The weighted average percent of participants that developed gastroenteritis after challenge across published studies is 47% (range: 27%-80%) (Table 4-S1).[53,90,97,108,177–179] We assumed *S* is the number of individuals susceptible to disease, as opposed to infection. To calculate *S* we multiplied 47% by *N* and rounded to the nearest integer. For some outbreaks the total number of cases (*C*) was greater than our estimated *S*; for these outbreaks we set *S* equal to *C*, corresponding to a 100% attack rate. We also calculated *S* assuming the percent susceptible was 27% and 80% of *N* to assess the sensitivity of our model results to this parameter.

Regression Analysis

After estimating R_0, R_e and associated SEs for each norovirus outbreak, we fit a generalized linear regression model to the log transformed estimated reproduction numbers (as R_0 and R_e values are not normally distributed) to assess whether outbreak setting, census region, season, year, whether norovirus was suspected or confirmed and norovirus genotype (categorized as either GII.4 or non-

GII.4) were associated with transmissibility. All variables were categorical, where the reference was assigned as the group with the most outbreaks reported. We used the iteratively reweighted least squares method to find the maximum likelihood estimates of a linear regression of log transformed reproduction numbers by outbreak setting. To determine which variables to include in our models we used a forward selection process with a linear regression of log transformed R_0 values. We considered the following variables to include in our models: outbreak setting, census region, season, year, whether norovirus was suspected or confirmed and norovirus genotype (categorized as either GII.4 or non-GII.4). Only 1,571 outbreaks (22%) for which we calculated basic and effective reproduction numbers had data on norovirus genotype. Given this small sample size, we did not include norovirus genotype in our models and performed model selection on the remaining variables. To determine which variables to include we selected the model with the lowest Akaike information criterion (AIC) value. All analyses were conducted in R version 3.4.2.[158]

Sensitivity Analysis

We tested the sensitivity of our regression model results to different modeling approaches and different assumptions of the percent susceptible at the start of an outbreak. We fit a logistic regression model of binary transmission and a negative binomial regression of the final outbreak size, using the log transformed exposed population size as an offset (i.e., a measure of the attack rate of an outbreak). Thus, we could make comparisons between the results of these models to see if the results from modeling continuous transmission were consistent with the results of modeling binary transmission and attack rates. Additionally, we re-ran all the regression models assuming

that the percent susceptible at the start of an outbreak were 27% and 80%, corresponding to the minimum and maximum values of the percent susceptible to AGE from published challenge studies (Table 4-S1).

RESULTS

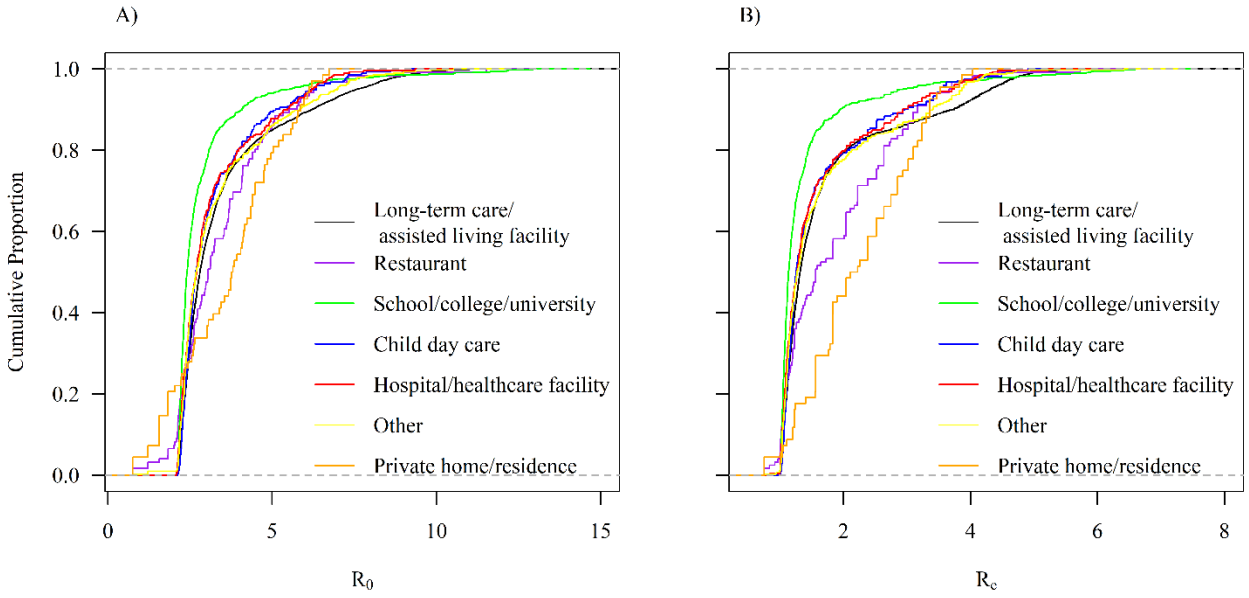
Between January 2009 to December 2017, 7,094 norovirus outbreaks meeting the inclusion criteria with data on both the number ill and the number exposed were reported to NORS. The majority of these outbreaks occurred in long-term care/assisted living facilities (n = 5,335; 75%) and occurred in winter (defined as December 1 to February 28; n = 4,016; 57%). The median outbreak size was 28 cases (IQR: 16, 47) and the median attack rate was 22% (IQR: 11%, 36%) (Table 4-1, Figure 4-1). The median R_0 was 2.75 (IQR: 2.38, 3.65), and the median R_e was 1.29 (IQR: 1.12, 1.74).

Table 4-1. Norovirus outbreaks with exposed population size reported to the National Outbreak Reporting System (NORS), 2009–2017.

	N (%)	Median Attack Rate (IQR)	Median Final Size (IQR)	Median R_0 (IQR)
All Outbreaks	7,094 (100)	22% (11%, 36%)	28 (16, 47)	2.75 (2.38, 3.65)
Major Setting				
Child day care	272 (4)	21% (13%, 36%)	18 (11, 29)	2.67 (2.39, 3.60)
Hospital/healthcare facility	271 (4)	22% (11%, 38%)	19 (11, 34)	2.70 (2.33, 3.59)
Long-term care/assisted living facility	5,335 (75)	23% (13%, 36%)	30 (17, 47)	2.81 (2.42, 3.76)
Other	350 (5)	20% (10%, 36%)	24 (15, 40)	2.66 (2.35, 3.60)
Private home/residence	42 (1)	66% (50%, 91%)	9 (6, 16)	3.80 (2.26, 4.92)
Restaurant	77 (1)	50% (27%, 64%)	10 (6, 16)	3.12 (2.53, 4.31)
School/college/university	747 (11)	12% (6%, 24%)	42 (19, 80)	2.41 (2.24, 2.92)
Season				
Winter	4,016 (57)	22% (12%, 36%)	30 (17, 51)	2.80 (2.40, 3.77)
Fall	808 (11)	21% (11%, 37%)	26 (15, 47)	2.72 (2.36, 3.63)

Spring	1,964 (28)	20% (11%, 35%)	27 (15, 44)	2.69 (2.37, 3.57)
Summer	306 (4)	17% (9%, 33%)	19 (11, 32)	2.57 (2.29, 3.33)
Outbreak Status				
Confirmed outbreak	3,114 (44)	26% (15%, 40%)	35 (20, 55)	2.99 (2.51, 4.22)
Suspected outbreak	3,980 (56)	18% (9%, 31%)	24 (14, 40)	2.59 (2.32, 3.27)
Census Region				
Region 1	1,898 (27)	17% (9%, 29%)	31 (17, 53)	2.58 (2.30, 3.23)
Region 2	2,205 (31)	25% (13%, 39%)	26 (15, 44)	2.87 (2.44, 3.98)
Region 3	2,224 (31)	23% (12%, 38%)	29 (17, 47)	2.81 (2.39, 3.93)
Region 4	767 (11)	21% (13%, 34%)	28 (16, 44)	2.75 (2.42, 3.57)
Year				
Jan 2009 - Jun 2009	243 (3)	28% (15%, 42%)	35 (20, 55)	3.09 (2.50, 4.56)
Jul 2009 - Jun 2010	275 (4)	29% (15%, 45%)	35 (19, 57)	3.17 (2.51, 4.77)
Jul 2010 - Jun 2011	592 (8)	29% (16%, 44%)	32 (19, 54)	3.12 (2.54, 4.58)
Jul 2011 - Jun 2012	679 (10)	27% (15%, 40%)	35 (19, 59)	3.01 (2.52, 4.29)
Jul 2012 - Jun 2013	967 (14)	21% (12%, 36%)	28 (16, 46)	2.73 (2.38, 3.61)
Jul 2013 - Jun 2014	913 (13)	20% (11%, 33%)	29 (18, 51)	2.68 (2.38, 3.45)
Jul 2014 - Jun 2015	941 (13)	21% (11%, 35%)	28 (16, 46)	2.74 (2.36, 3.61)
Jul 2015 - Jun 2016	1,007 (14)	17% (9%, 32%)	25 (14, 42)	2.57 (2.31, 3.29)
Jul 2016 - Jun 2017	1,070 (15)	19% (10%, 31%)	26 (14, 42)	2.63 (2.33, 3.24)
Jul 2017 - Dec 2017	407 (6)	20% (10%, 34%)	22 (14, 38)	2.66 (2.33, 3.55)

Figure 4-1. Cumulative proportion of outbreaks across major outbreak settings for (A) R_0 and (B) R_e assuming the intitial proportion susceptible is 47%. From our generalized linear regressions, outbreaks in school/college/universities (green) and hospitals/healthcare facilites (red) lower transmission than outbreaks in long-term care/assisted living facilities (black). Outbreaks in private home/residences and restaurants had higher transmission, however confidence intervals were wide due to small sample sizes.



Model Selection

The model we selected included the following variables: major setting, census region, season, year, and whether norovirus was suspected or confirmed *(AIC*=5803*)* (Table 4-S2).

Regression Analysis

The reference group for our linear regression model was outbreaks with confirmed norovirus etiology that occurred in long-term care/assisted living facilities in the south census region, during winter of the July 2016 – June 2017 norovirus season. Assuming the percent of the population susceptible to norovirus AGE at the start of an outbreak was 47%, we found that outbreaks in our referent group had an average predicted R_0 of 3.35 (95% CI: 3.26, 3.45). Compared to the reference group, outbreaks in schools/colleges/universities had a lower average predicted R_0 (2.92; 95% CI: 2.81, 3.03) and hospitals/other healthcare facilities had marginally lower average predicted R_0 (3.13; 95% CI: 2.97, 3.30). The average predicted R_0 for outbreaks in child day cares (R_0 = 3.31; 95% CI: 3.13, 3.49), private homes/residences (R_0 = 3.31; 95% CI: 2.95, 3.72), restaurants (R_0 = 3.37; 95% CI: 3.10, 3.68), and other settings (R_0 = 3.26; 95% CI: 3.11, 3.42) were not substantially different from the average predicted R_0 of outbreaks in the reference group (Table 4-2, Figure 4-1). Outbreaks in the summer, outbreaks with suspected norovirus etiology, and outbreaks in the northeast had lower average predicted R_0 (3.11, 95% CI: 2.95, 3.27; 3.02, 95% CI: 2.94, 3.11; and 3.00, 95% CI: 2.91, 3.09, respectively) than the reference group. Outbreaks that were reported from January 2009 - June 2012 all had higher average predicted R_0 (range of predicted R_0 for individual seasonal years from January 2009 - June 2012: 3.77 – 3.93) than the reference group (Table 4-2, Figure 4-1).

Table 4-2. Regression Model Results. Estimated log-linear change in R_0 (95% confidence interval) from the intercept estimate of R_0 from linear regression of log transformed R_0.

	Estimated log-linear change in R_0
Intercept	3.35 (3.26, 3.45)
Major Setting	
Long-term care/assisted living facility	Ref
Child Day Care	0.99 (0.94, 1.03)
Hospital/healthcare facility	0.93 (0.89, 0.98)
Other	0.97 (0.93, 1.01)
Private home/residence	0.99 (0.88, 1.10)
Restaurant	1.01 (0.93, 1.09)
School/college/university	0.87 (0.85, 0.90)
Season	
Winter	Ref
Fall	1.00 (0.98, 1.03)
Spring	0.98 (0.96, 1.00)
Summer	0.93 (0.89, 0.97)
Outbreak Status	
Confirmed outbreak	Ref
Suspected outbreak	0.90 (0.89, 0.92)
Census Region	
South	Ref
Northeast	0.89 (0.87, 0.91)
Midwest	1.00 (0.97, 1.02)
West	0.98 (0.95, 1.01)
Year	
Jan 2009 - Jun 2009	1.16 (1.10, 1.22)
Jul 2009 - Jun 2010	1.17 (1.12, 1.23)
Jul 2010 - Jun 2011	1.16 (1.12, 1.21)
Jul 2011 - Jun 2012	1.12 (1.08, 1.16)
Jul 2012 - Jun 2013	1.04 (1.01, 1.07)
Jul 2013 - Jun 2014	1.02 (0.99, 1.06)
Jul 2014 - Jun 2015	1.05 (1.02, 1.08)
Jul 2015 - Jun 2016	1.02 (0.99, 1.05)
Jul 2016 - Jun 2017	Ref
Jul 2017 - Dec 2017	1.04 (1.00, 1.09)

Our findings were generally robust to assumptions about the proportion susceptible at the start of the outbreak and whether we modeled the outcome of R_0, R_e, final outbreak size (Text S1, Table 4-S4, Table 4-S5, Table 4-S6, Figure 4-S2).

DISCUSSION

Using a large national outbreak dataset, we investigated transmission patterns of norovirus outbreaks. Our analysis led to several key findings. First, norovirus outbreaks in the US have modest values of R_0 and R_e (2.75 [IQR: 2.38, 3.65] and 1.29 [IQR: 1.12, 1.74], respectively). Second, we found outbreaks in long-term care/assisted living facilities had higher transmission compared to other common settings. In addition, we found higher transmission in laboratory confirmed outbreaks relative to suspected outbreaks and higher transmission for outbreaks occurring in the winter months relative to summer months.

In their recent review of norovirus modelling studies, Gaythorpe et al. found that estimates of basic reproduction numbers for norovirus ranged from 1.1 to 7.2.[60] Our estimates are similar to those reproduction numbers estimated using transmission models of norovirus at the outbreak-level;[136,180] however our estimates are higher than several studies that estimated reproduction numbers using population-level transmission models,[91,181–183] suggesting that transmission of norovirus in outbreak settings is higher than sporadic transmission in the community.

From our main analysis, we found that outbreaks in hospitals/other healthcare facilities, schools/colleges/universities and other settings had lower estimated transmission, while private homes/residences had higher estimated reproduction numbers relative to outbreaks in long-term care/assisted living facilities. Relative to outbreaks in long-term care/assisted living facilities, outbreaks that occurred in private homes/residences and restaurants had higher final sizes and schools/colleges/universities had lower estimated attack rates. Our finding that outbreaks in the winter had higher estimated transmissibility than outbreaks that occurred in summer is likely driven by the strong winter time seasonality of noroviruses in the US.[17,19] Consistent with this finding are the observations that norovirus case and outbreak reports are inversely correlated with temperature,[17,19] and norovirus surrogate virus (e.g., murine norovirus, feline calicivirus) survival declines with increasing temperatures.[20,21]

Several differences we found are likely driven by administrative or programmatic factors related to reporting, rather than differences in norovirus transmission. Suspected norovirus outbreaks without a laboratory-confirmed outbreak etiology had lower transmission than laboratory-confirmed norovirus outbreaks, perhaps because suspected norovirus outbreaks are less well investigated than confirmed outbreaks, with lower rates of case ascertainment. Outbreaks reported in the south had higher transmissibility relative to outbreaks in the northeast. This may be related to differences in the quality of reporting between these regions. There is tremendous variability in outbreak reporting between states, approximately 100-fold difference between the highest and lowest reporting states, which likely impacts the observed outbreak characteristics included herein.[120] Similarly, while NORS has been collecting outbreak reports since January 2009, in August 2012 the CDC began a concerted effort to improve norovirus outbreak reporting to NORS

and CaliciNet with the introduction of NoroSTAT.[184,185] Thus our finding that norovirus outbreaks reported prior to August 2012 were larger with higher estimated R_0 and R_e values may be due to the CDC's efforts to capture outbreaks that previously would have not been reported (i.e., smaller scale outbreaks).

There are a number of additional limitations to this study. First, our process of data selection may have introduced bias into our analyses. We excluded outbreaks that occurred in multiple states, which are likely to have higher transmissibility given the larger geographic range of these outbreaks. This exclusion could bias our estimates of transmission downwards; however, there were only 8 multi-state outbreaks, thus the bias is likely negligible. Additionally, we excluded outbreaks for which the exposed population size was not reported. Excluding these outbreaks could introduce bias if there are certain settings where the exposed population size is more likely to be reported. For example, the majority of outbreak reports in long-term care/assisted living facilities had information on the exposed population size. We only selected outbreaks where the mode of transmission was person-to-person, thus our estimates of transmissibility of norovirus are not generalizable to outbreaks where transmission occurs via other modes (e.g., foodborne, waterborne, environmental). Further, as the mode of transmission for norovirus outbreaks can be difficult to identify, there may have been outbreaks included in our analysis where the mode of transmission was misclassified as person-to-person.

A second limitation is that the final size method assumes a susceptible-infected-recovered (SIR) type infection with a homogenously mixing population.[176] This is a simplification that does not reflect true mixing patterns in many settings. Second, the exposed population size is difficult to

quantify and is not consistently reported to NORS. Thus, it is difficult to disentangle whether the differences we found in estimated attack rates across different settings arose due to true variability in the exposed population size across settings, or variability in the reliable reporting of the exposed population size. However, our analysis restricted to outbreaks in long-term care/assisted living facilities found the same trends among the variables for outbreak status, census region, season, and year as our analysis of all outbreaks, which suggests these results are reasonably robust.

Another limitation to our study is that the final size method may underestimate reproduction numbers for outbreaks in small populations with high attack rates. For example, in private homes, where attack rates may be high, if everyone in the household is infected, then no additional infections can occur. Thus the final size method cannot capture any additional transmission that could have happened if the exposed population size is larger (i.e., number of individuals in the household). Becker termed this limitation the "wasted infection potential."[186] The effects of this limitation are demonstrated in the sensitivity analysis of our assumption of the percent susceptible at the start of an outbreak. When we assumed the percent susceptible was between 47% and 80% (i.e., the effective population size for the outbreak is higher), the estimated transmissibility of norovirus in private homes/residences and restaurants was higher than transmissibility in long-term care/assisted living facilities. However, when the percent susceptible at the start of an outbreak was assumed to be 27%, the association between private homes/residences and restaurants reversed, such that these settings had lower estimated transmission relative to outbreaks in long-term care/assisted living facilities. This is because the population size that can be infected is much lower when we assume 27% susceptibility, thus more outbreaks would result in all individuals becoming infected (i.e., 100% attack rate) leading to the "wasted infection potential"

issue. For example, if a household had 15 individuals the maximum possible R_0 assuming 27% susceptibility is 7.3, which is lower than the average predicted R_0 for outbreaks in the reference group. Therefore, the results for private homes/residences and restaurants, where exposed populations sizes are low, should be interpreted with caution as these transmission values in these settings may be underestimated.

A further limitation of this analysis is that the final size method does not account for the effect of control measures. For some of the outbreaks represented in our data set, control measures were most likely implemented (e.g., isolation of ill persons, cleaning of contamination, etc.). These interventions would likely reduce the number ill such that the estimated R_0 would be lower than the true R_0 (i.e., R_0 in the absence of control measures).

Additionally, the final size method assumes that the proportion of susceptibles is known at the start of an outbreak; however, the level of susceptibility to norovirus in not well known. Certain host genetic factors are associated with the ability of norovirus to establish an infection within a human host,[80–83] leading to variable susceptibility to norovirus infection.[66,84,85] Secretor negative individuals have non-functional fucosyltransferase-2 (FUT2) genes thus certain norovirus genogroups (i.e., genogroup I and genogroup II type 4) cannot bind and therefore fail to cause infection.[80,82,83,87,88] Our estimates of the basic and effective reproduction numbers assume that the percent of the population susceptible at the start of all outbreaks in our dataset is 47%. In reality, the proportion susceptible varies from outbreak to outbreak, and potentially over time and age as the distribution of circulating norovirus genotypes changes. We also assumed that only symptomatic individuals contribute to transmission in our calculations. However, it is recognized

that individuals with asymptomatic norovirus infections may contribute to transmission, but they are likely not as infectious as individuals with symptomatic infections.[121,187]

Finally, our main analysis does not account for norovirus genotype. Due to the limited data available on norovirus genotype we were not adequately able to assess for associations with transmission. As more genotyping data become available future studies should assess whether norovirus genotype influences transmission.

We estimated reproduction numbers using the final size method for thousands of outbreaks from a national outbreak reporting system, then used these estimates to examine factors associated with norovirus transmission. Our analyses showed that norovirus transmission rates are modest. Such modest rates of effective reproduction suggest there are opportunities for effective control measures to curtail onward transmission of norovirus. However, challenges remain. Asymptomatic transmission, which we did account for in this analysis and generally goes undetected in surveillance, can limit the effectiveness of traditional control methods focused on ill individuals, even for pathogens with modest transmission.[132]

Our results provide evidence that transmission among elderly populations during outbreaks in long-term care/assisted living facilities pose an important public health burden. As such, vaccination of the elderly may be considered to reduce norovirus associated mortality and/or mitigate the burden of outbreaks in long-term care/assisted living facilities. Additionally, our finding that transmission of norovirus is higher for outbreaks that occur in winter suggests that seasonal vaccine policies could be considered to reduce the burden of disease seasonally.

SUPPLEMENT

Table 4-S1. Challenge Study Data. Data from published norovirus challenge studies on the number of participants challenged with norovirus and the number of challenged participants that subsequently developed acute gastroenteritis (AGE). We assumed that the average proportion that develop AGE across all studies, weighted by the total number of participants in each study, is the proportion that are susceptible to norovirus in our calculations of R_0 and R_e.

Study (citation)	Number Challenged	Number AGE*	Proportion with AGE
Dolin 1970 [188]	12	9	0.75
Wyatt 1974 [66]	23	16	0.70
Parrino 1977 [84]	12	6	0.50
Treanor 1988 [97]	10	8	0.80
Johnson 1990 [89]	42	25	0.60
Graham 1994 [85]	50	34	0.68
Lindesmith 2003 [83]	77	21	0.27
Lindesmith 2005 [90]	15	7	0.47
Atmar 2008 [53]	21	11	0.52
Leon 2011 [179]	15	5	0.33
Atmar 2011 [109]	41	29	0.71
Seitz 2011 [189]	13	10	0.77
Frenck 2012 [178]	40	12	0.30
Bernstein 2015 [108]	98	29	0.30
Overall	469	222	0.47†

***** Acute gastroenteritis

†Average proportion susceptible weighted by number of participants

Table 4-S2. Model Selection. Estimated log linear change in R_0 (95% CI) from the estimated R_0 for the intercept for each model in a forward selection process for a linear regression model of log transformed R_0 values. The Akaike information criterion (AIC) for each model was as follows: model 1, 6,237; model 2, 6,050; model 3, 5,935; model 4, 5,920; and model 5, 5,803.

	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	3.22 (3.19, 3.25)	3.44 (3.39, 3.49)	3.57 (3.51, 3.64)	3.61 (3.54, 3.68)	3.35 (3.26, 3.45)
Child Day Care	0.95 (0.91, 0.99)	0.99 (0.94, 1.03)	0.98 (0.93, 1.02)	0.98 (0.93, 1.03)	0.99 (0.94, 1.03)
Hospital/healthcare facility	0.94 (0.90, 0.98)	0.93 (0.89, 0.97)	0.94 (0.90, 0.98)	0.94 (0.90, 0.98)	0.93 (0.89, 0.98)
Other	0.96 (0.92, 1.00)	0.96 (0.93, 1.00)	0.95 (0.92, 0.99)	0.97 (0.93, 1.01)	0.97 (0.93, 1.01)
Private home/residence	1.00 (0.89, 1.12)	0.99 (0.88, 1.11)	0.98 (0.87, 1.09)	0.99 (0.88, 1.11)	0.99 (0.88, 1.1)
Restaurant	1.02 (0.94, 1.11)	1.01 (0.93, 1.10)	1.00 (0.92, 1.08)	1.00 (0.92, 1.09)	1.01 (0.93, 1.09)
School/college/university	0.84 (0.82, 0.87)	0.88 (0.85, 0.90)	0.86 (0.84, 0.89)	0.86 (0.83, 0.89)	0.87 (0.85, 0.9)
Probable outbreak	--	0.88 (0.87, 0.90)	0.88 (0.87, 0.90)	0.89 (0.87, 0.9)	0.90 (0.89, 0.92)
Region 1	--	--	0.89 (0.87, 0.91)	0.89 (0.87, 0.91)	0.89 (0.87, 0.91)
Region 2	--	--	0.99 (0.97, 1.02)	0.99 (0.97, 1.02)	1.00 (0.97, 1.02)
Region 4	--	--	0.96 (0.93, 0.99)	0.96 (0.93, 0.99)	0.98 (0.95, 1.01)
Fall	--	--	--	1.00 (0.97, 1.03)	1.00 (0.98, 1.03)
Spring	--	--	--	0.97 (0.95, 0.99)	0.98 (0.96, 1.00)
Summer	--	--	--	0.92 (0.88, 0.96)	0.93 (0.89, 0.97)
Jan 2009 - Jun 2009	--	--	--	--	1.16 (1.10, 1.22)
Jul 2009 - Jun 2010	--	--	--	--	1.17 (1.12, 1.23)
Jul 2010 - Jun 2011	--	--	--	--	1.16 (1.12, 1.21)
Jul 2011 - Jun 2012	--	--	--	--	1.12 (1.08, 1.16)
Jul 2012 - Jun 2013	--	--	--	--	1.04 (1.01, 1.07)
Jul 2013 - Jun 2014	--	--	--	--	1.02 (0.99, 1.06)
Jul 2014 - Jun 2015	--	--	--	--	1.05 (1.02, 1.08)
Jul 2015 - Jun 2016	--	--	--	--	1.02 (0.99, 1.05)
Jul 2017 - Dec 2017	--	--	--	--	1.04 (1.00, 1.09)

Alternative Models

We assessed two alternative approaches for modeling norovirus transmission: a logistic regression to model a binary transmission outcome (i.e., high versus low transmission) and a negative binomial regression to model the final size of outbreaks, adjusting for exposed population size (i.e., modeling attack rates). For our logistic regression, we used the first and third tertiles of estimated values of R_0 and R_e, assuming the percent susceptible was 47%, to determine the cutoffs for our outcome of interest: low versus high transmission. We excluded outbreaks with transmission values within the second tertile and focus our logistic regression comparison between the lowest and highest tertiles of transmission. The third tertile of R_0 and R_e values were 3.23 and 1.52, respectively.

The trends of transmissibility across our variables of interest (outbreak setting, census region, season, year, whether norovirus was suspected or confirmed and norovirus genotype) from our main regression analysis of a continuous transmission outcome were consistent across the logistic regressions of high R_0 (R_0>3.23) and R_e (R_e>1.52) and linear regression of R_e values. (Table 4-S3) The trends of transmissibility were consistent for most of our variables of interest in the negative binomial model of final outbreak sizes; however, private homes/residences and restaurants had a much more pronounced effect on the attack rate of outbreaks, relative to long-term care/assisted living facilities (RR=2.35 (95% CI: 1.85, 3.01) and RR=1.67 (95% CI: 1.40, 2.01), respectively). As the exposed population size is difficult to quantify, and thus may not be reported reliably, we analyzed the subset of outbreaks that occurred within long-term care/assisted living facilities with our regression models. The patterns found among the variables for outbreak status, census region, season and year were consistent with what was found analyzing the full data set. (Table 4-S4)

Table 4-S3. Regression Model Results. Risk ratios of attach rates, estimated log-linear change in R_0 and R_e (95% confidence interval) relative to the intercept from linear regression of the log transformed reproduction numbers and odds ratios (95% confidence interval) of an outbreak with high transmission from logistic regression, assuming the percent susceptible at the start of an outbreak is 47%.

	RR of Attack Rates	OR of R_0>3.23	Estimated log-linear change in R_e	OR of R_e>1.52*
Intercept	0.27 (0.26, 0.29)	1.75 (1.41, 2.18)	1.63 (1.57, 1.68)	1.72 (1.39, 2.13)
Major Setting				
Long-term care/assisted living facility	Ref	Ref	Ref	Ref
Child Day Care	1.08 (0.98, 1.19)	1.03 (0.74, 1.42)	1.00 (0.95, 1.05)	1.11 (0.80, 1.53)
Hospital/healthcare facility	0.94 (0.86, 1.03)	0.62 (0.45, 0.85)	0.94 (0.89, 0.99)	0.65 (0.48, 0.89)
Other	0.98 (0.90, 1.06)	0.73 (0.55, 0.96)	0.98 (0.93, 1.02)	0.74 (0.56, 0.98)
Private home/residence	2.35 (1.85, 3.01)	1.80 (0.88, 3.87)	1.31 (1.15, 1.49)	8.47 (3.22, 29.27)
Restaurant	1.67 (1.40, 2.01)	1.41 (0.78, 2.62)	1.10 (1.00, 1.21)	1.96 (1.09, 3.67)
School/college/university	0.67 (0.63, 0.71)	0.29 (0.23, 0.36)	0.86 (0.83, 0.89)	0.30 (0.24, 0.38)
Season				
Winter	Ref	Ref	Ref	Ref
Fall	0.99 (0.94, 1.05)	1.00 (0.81, 1.23)	1.01 (0.97, 1.04)	0.97 (0.79, 1.19)
Spring	0.97 (0.93, 1.01)	0.90 (0.78, 1.04)	0.98 (0.95, 1.00)	0.89 (0.77, 1.04)
Summer	0.86 (0.79, 0.95)	0.65 (0.47, 0.88)	0.92 (0.88, 0.97)	0.62 (0.45, 0.85)
Outbreak Status				
Confirmed outbreak	Ref	Ref	Ref	Ref
Suspected outbreak	0.82 (0.79, 0.85)	0.43 (0.37, 0.49)	0.89 (0.87, 0.91)	0.42 (0.37, 0.47)
Census Region				
South	Ref	Ref	Ref	Ref
Northeast	0.77 (0.74, 0.81)	0.44 (0.37, 0.52)	0.88 (0.85, 0.9)	0.45 (0.38, 0.53)
Midwest	1.07 (1.02, 1.11)	1.02 (0.87, 1.20)	1.00 (0.98, 1.03)	1.03 (0.88, 1.20)
West	0.99 (0.93, 1.05)	0.95 (0.76, 1.19)	0.97 (0.93, 1.00)	0.94 (0.75, 1.17)
Year				
Jan 2009 - Jun 2009	1.34 (1.21, 1.48)	2.49 (1.73, 3.62)	1.20 (1.13, 1.27)	2.50 (1.73, 3.63)
Jul 2009 - Jun 2010	1.37 (1.25, 1.52)	2.59 (1.82, 3.73)	1.22 (1.16, 1.29)	2.55 (1.79, 3.65)
Jul 2010 - Jun 2011	1.34 (1.25, 1.45)	2.64 (2.01, 3.49)	1.20 (1.15, 1.25)	2.58 (1.97, 3.39)
Jul 2011 - Jun 2012	1.24 (1.15, 1.33)	2.12 (1.63, 2.77)	1.14 (1.09, 1.19)	2.16 (1.66, 2.81)
Jul 2012 - Jun 2013	1.08 (1.01, 1.15)	1.26 (0.99, 1.59)	1.05 (1.01, 1.09)	1.29 (1.02, 1.63)
Jul 2013 - Jun 2014	1.04 (0.97, 1.11)	1.19 (0.93, 1.52)	1.04 (1.00, 1.08)	1.18 (0.93, 1.51)
Jul 2014 - Jun 2015	1.08 (1.01, 1.15)	1.33 (1.05, 1.68)	1.06 (1.02, 1.1)	1.32 (1.04, 1.66)
Jul 2015 - Jun 2016	1.05 (0.98, 1.12)	1.14 (0.91, 1.44)	1.03 (1.00, 1.07)	1.15 (0.91, 1.44)
Jul 2016 - Jun 2017	Ref	Ref	Ref	Ref
Jul 2017 - Dec 2017	1.05 (0.96, 1.14)	1.40 (1.03, 1.90)	1.04 (0.99, 1.09)	1.33 (0.98, 1.80)

* logistic regression compares outbreaks with transmission in the third tertile (Re>1.52) to outbreaks in the first tertile (Re<1.17) and does not include Re values in second tertile. Linear and negative binomial regressions use full dataset

Table 4-S4. Estimated log-linear change in R_0 and R_e (95% confidence interval) relative to the intercept from linear regression, odds ratios of outbreaks with R_0>3.23 and R_e>1.52, final size adjusting for exposed population size among long-term care/assisted care facilities, assuming the percent susceptible at the start of the outbreak is 47%.

	RR of Attack Rates	Basic Reproduction Number		Effective Reproduction Number	
		Estimated log-linear change in R_0	OR of R_0>3.23*	Estimated log-linear change in R_e	OR of R_e>1.52*
Intercept	0.28 (0.26, 0.30)	3.39 (3.28, 3.51)	2.01 (1.56, 2.60)	1.63 (1.57, 1.70)	2.00 (1.55, 2.57)
Season					
Winter	Ref	Ref	Ref	Ref	Ref
Fall	0.99 (0.92, 1.06)	1.00 (0.96, 1.03)	0.92 (0.71, 1.20)	1.00 (0.96, 1.04)	0.90 (0.69, 1.16)
Spring	0.96 (0.92, 1.00)	0.97 (0.95, 1.00)	0.87 (0.73, 1.03)	0.97 (0.94, 1.00)	0.87 (0.73, 1.03)
Summer	0.83 (0.73, 0.93)	0.90 (0.85, 0.96)	0.55 (0.36, 0.85)	0.90 (0.83, 0.96)	0.54 (0.35, 0.84)
Outbreak Status					
Confirmed outbreak	Ref	Ref	Ref	Ref	Ref
Suspected outbreak	0.81 (0.78, 0.84)	0.89 (0.87, 0.91)	0.39 (0.34, 0.45)	0.89 (0.86, 0.91)	0.39 (0.34, 0.45)
Census Region					
South	Ref	Ref	Ref	Ref	Ref
Northeast	0.69 (0.66, 0.73)	0.86 (0.84, 0.88)	0.31 (0.25, 0.37)	0.84 (0.81, 0.86)	0.31 (0.26, 0.38)
Midwest	1.00 (0.95, 1.05)	0.98 (0.95, 1.00)	0.83 (0.69, 0.99)	0.99 (0.96, 1.02)	0.81 (0.67, 0.97)
West	0.96 (0.90, 1.03)	0.96 (0.93, 1.00)	0.89 (0.67, 1.18)	0.96 (0.92, 1.00)	0.89 (0.67, 1.17)
Year					
Jan 2009 - Jun 2009	1.47 (1.32, 1.63)	1.20 (1.14, 1.28)	3.39 (2.22, 5.23)	1.26 (1.18, 1.34)	3.41 (2.24, 5.25)
Jul 2009 - Jun 2010	1.46 (1.31, 1.63)	1.21 (1.14, 1.28)	3.04 (1.98, 4.74)	1.28 (1.20, 1.37)	2.92 (1.91, 4.51)
Jul 2010 - Jun 2011	1.43 (1.32, 1.55)	1.20 (1.15, 1.25)	3.27 (2.39, 4.50)	1.25 (1.19, 1.31)	3.13 (2.30, 4.29)
Jul 2011 - Jun 2012	1.28 (1.19, 1.38)	1.14 (1.10, 1.19)	2.39 (1.76, 3.25)	1.17 (1.12, 1.23)	2.35 (1.74, 3.19)
Jul 2012 - Jun 2013	1.10 (1.03, 1.18)	1.05 (1.02, 1.09)	1.35 (1.03, 1.78)	1.07 (1.02, 1.11)	1.35 (1.03, 1.78)
Jul 2013 - Jun 2014	1.08 (1.01, 1.17)	1.04 (1.00, 1.08)	1.29 (0.97, 1.74)	1.05 (1.01, 1.10)	1.28 (0.95, 1.71)
Jul 2014 - Jun 2015	1.16 (1.08, 1.24)	1.07 (1.04, 1.12)	1.59 (1.21, 2.08)	1.09 (1.05, 1.14)	1.56 (1.19, 2.04)
Jul 2015 - Jun 2016	1.08 (1.01, 1.16)	1.02 (0.99, 1.07)	1.19 (0.9, 1.58)	1.05 (1.00, 1.09)	1.19 (0.90, 1.57)
Jul 2016 - Dec 2016	Ref	Ref	Ref	Ref	Ref
Jul 2017 - Dec 2017	1.09 (0.99, 1.20)	1.06 (1.01, 1.12)	1.56 (1.08, 2.27)	1.06 (1.00, 1.13)	1.50 (1.04, 2.16)

* logistic regression compares outbreaks with transmission in the third tertile (R_0>3.23, R_e>1.52) to outbreaks in the first tertile (R_0<2.48, R_e<1.17), and does not include R_e values in second tertile. Linear and negative binomial regressions use full dataset.

Percent Susceptible

We tested the sensitivity of our results to the assumption of the percent susceptible start of an outbreak by running all the regression models assuming the percent susceptible at the start was 27% and 80%, which represent the minimum and maximum estimates of the percent susceptible to AGE from published challenge studies, respectively (Table 4-S1). By adjusting our assumption of the percent susceptible at the start of the outbreak to 27% and 80%, the median R_0 was 6.04 (IQR: 4.53, 9.38) and 1.43 (IQR: 1.33, 1.61), respectively, while the median R_e was 1.82 (IQR: 1.24, 3.83) and 1.14 (IQR: 1.07, 1.29), respectively. Assuming the proportion susceptible was 27% we found that outbreaks in long-term care/assisted living facilities were more likely to have R_0 >8.06 and R_e>3.24 relative to all other settings. (Table 4-S5, Table 4-S6, Figure 4-S1) When the percent susceptible was 80% we found that outbreaks in long-term care/assisted living facilities had increased odds of having R_0 >1.54 and R_e>1.23 compared to hospitals/other healthcare facilities, schools/colleges/universities, and other settings. Outbreaks in private homes/residences and restaurants had higher odds of having R_0 >1.54 and R_e>1.23 relative to long-term care/assisted living facilities, however the confidence intervals are wide due to small sample sizes. (Table 4-S5, Table 4-S6, Figure 4-S1) Trends in the variables for census region, season, year, and whether norovirus was suspected or confirmed for the models assuming 27% and 80% susceptibility were consistent with the models assuming 47% susceptibility. (Table 4-2, Table 4-S5, Table 4-S6)

Figure 4-S1. Cumulative proportion of outbreaks by R_0 and R_e across major setting assuming the percent of the population susceptible at the start of an outbreak is 27% (Panels A and B) and 80% (Panels C and D).

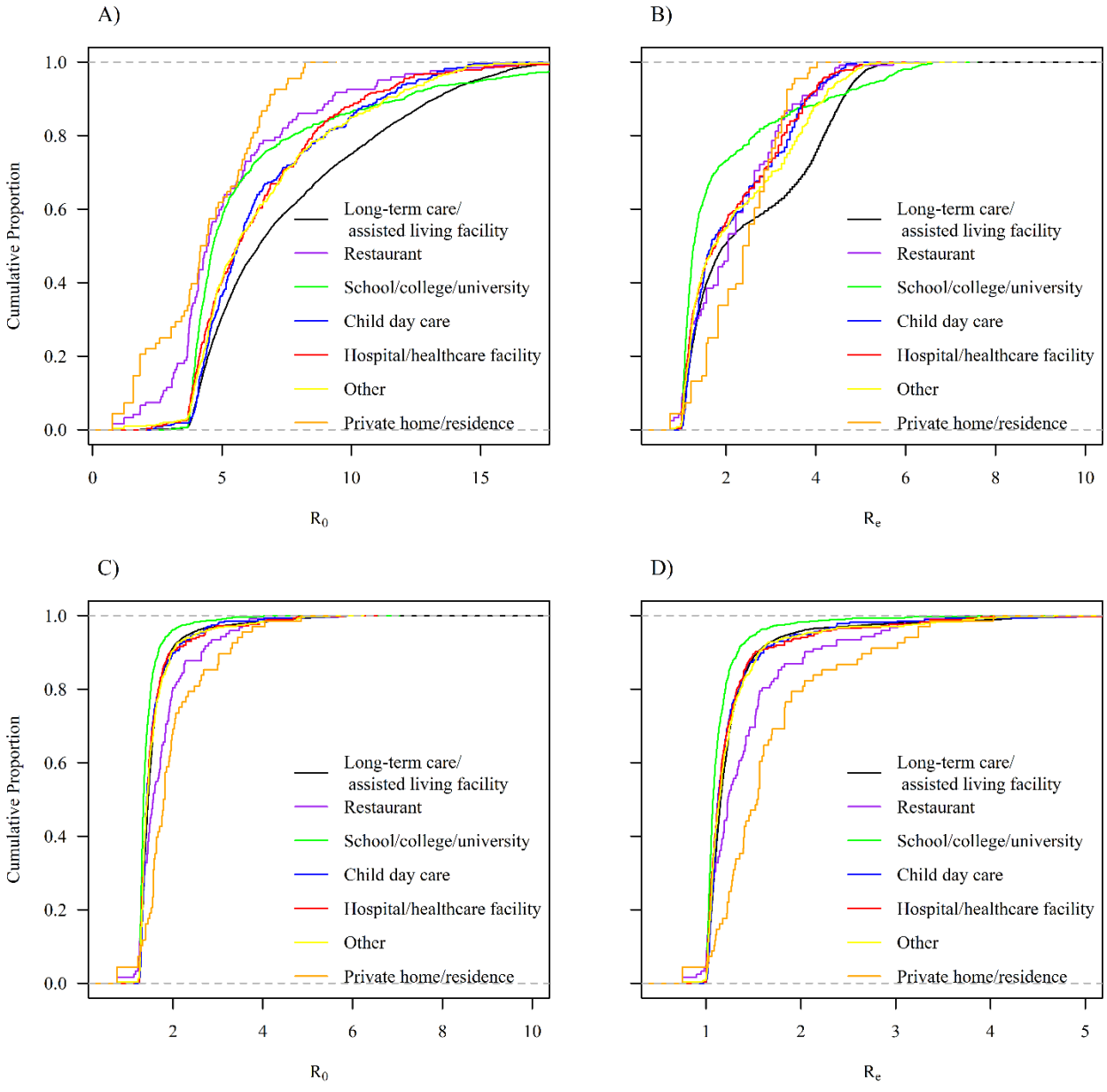


Table 4-S5. Estimated log-linear change in R_0 (95% CI) relative to the intercept from linear regression and odds ratios of outbreaks with high R_0 assuming the percent susceptible as the start of the outbreak is 27% and 80%.

	S=27%(N)		S=80%(N)	
	Estimated log-linear change in R_0	OR of R_0>8.06*	Estimated log-linear change in R_0	OR of R_0>1.54*
Intercept	7.42 (7.17, 7.68)	2.03 (1.64, 2.52)	1.57 (1.54, 1.60)	1.78 (1.43, 2.22)
Major Setting				
Long-term care/assisted living facility	Ref	Ref	Ref	Ref
Child Day Care	0.92 (0.87, 0.97)	0.64 (0.46, 0.90)	1.02 (0.99, 1.04)	1.11 (0.80, 1.55)
Hospital/healthcare facility	0.85 (0.81, 0.90)	0.46 (0.33, 0.64)	0.99 (0.96, 1.02)	0.66 (0.48, 0.91)
Other	0.91 (0.86, 0.95)	0.54 (0.41, 0.73)	0.99 (0.96, 1.01)	0.77 (0.58, 1.01)
Private home/residence	0.53 (0.47, 0.60)	0.02 (0.00, 0.12)	1.17 (1.10, 1.26)	5.17 (2.22, 14.21)
Restaurant	0.69 (0.62, 0.75)	0.17 (0.09, 0.33)	1.08 (1.03, 1.14)	1.78 (0.99, 3.31)
School/college/university	0.85 (0.82, 0.88)	0.32 (0.26, 0.39)	0.93 (0.91, 0.94)	0.30 (0.24, 0.38)
Season				
Winter	Ref	Ref	Ref	Ref
Fall	0.97 (0.94, 1.01)	0.91 (0.74, 1.12)	1.00 (0.98, 1.01)	0.94 (0.77, 1.16)
Spring	0.97 (0.94, 0.99)	0.80 (0.69, 0.93)	0.99 (0.98, 1.00)	0.88 (0.76, 1.02)
Summer	0.90 (0.86, 0.95)	0.54 (0.39, 0.75)	0.97 (0.94, 1.00)	0.62 (0.45, 0.85)
Outbreak Status				
Confirmed outbreak	Ref	Ref	Ref	Ref
Probable outbreak	0.87 (0.85, 0.89)	0.45 (0.39, 0.51)	0.95 (0.94, 0.96)	0.42 (0.37, 0.48)
Census Region				
South	Ref	Ref	Ref	Ref
Northwest	0.93 (0.91, 0.96)	0.56 (0.47, 0.65)	0.93 (0.92, 0.95)	0.44 (0.37, 0.52)
Midwest	0.99 (0.97, 1.02)	0.93 (0.80, 1.09)	1.01 (1.00, 1.03)	1.03 (0.88, 1.21)
West	1.03 (0.99, 1.06)	1.10 (0.88, 1.38)	0.98 (0.96, 1.00)	0.95 (0.76, 1.19)
Year				
Jan 2009 - Jun 2009	1.12 (1.06, 1.19)	2.07 (1.43, 3.02)	1.10 (1.06, 1.13)	2.62 (1.80, 3.83)
Jul 2009 - Jun 2010	1.11 (1.05, 1.18)	1.87 (1.31, 2.68)	1.12 (1.08, 1.15)	2.64 (1.85, 3.82)
Jul 2010 - Jun 2011	1.11 (1.07, 1.16)	1.96 (1.50, 2.58)	1.09 (1.07, 1.12)	2.59 (1.97, 3.43)
Jul 2011 - Jun 2012	1.13 (1.08, 1.18)	2.09 (1.62, 2.71)	1.06 (1.04, 1.08)	2.15 (1.65, 2.81)
Jul 2012 - Jun 2013	1.02 (0.99, 1.06)	1.13 (0.90, 1.42)	1.02 (1.00, 1.04)	1.33 (1.05, 1.69)
Jul 2013 - Jun 2014	1.01 (0.97, 1.05)	1.13 (0.89, 1.43)	1.02 (1.00, 1.04)	1.20 (0.94, 1.54)
Jul 2014 - Jun 2015	1.03 (0.99, 1.07)	1.27 (1.01, 1.60)	1.03 (1.01, 1.05)	1.37 (1.08, 1.73)
Jul 2015 - Jun 2016	1.00 (0.96, 1.03)	1.01 (0.80, 1.27)	1.03 (1.01, 1.05)	1.12 (0.88, 1.41)
Jul 2016 - Jun 2017	Ref	Ref	Ref	Ref
Jul 2017 - Dec 2017	1.04 (0.99, 1.09)	1.23 (0.91, 1.66)	1.01 (0.98, 1.04)	1.31 (0.97, 1.78)

* logistic regression compares outbreaks with transmission in the third tertile (S=27%(N): R_0>8.06; S=80%(N): R_0>1.54) to outbreaks in the first tertile (S=27%(N): R_0<4.96; S=80%(N): R_0<1.36) and does not include R_0/R_e values in second tertile. Linear and negative binomial regressions use full dataset.

Table 4-S6. Estimated log-liner change in R_e (95% CI) relative to the intercept from linear regression and odds ratios of outbreaks with high R_e assuming the percent susceptible at the start of the outbreak is 27% and 80%.

	S=27%(N)		S=80%(N)	
	Estimated log linear change in R_e	OR of R_e>3.24*	Estimated log-linear change in R_e	OR of R_e>1.23*
Intercept	2.41 (2.31, 2.52)	1.88 (1.51, 2.33)	1.25 (1.23, 1.28)	1.67 (1.34, 2.07)
Major Setting				
Long-term care/assisted living facility	Ref	Ref	Ref	Ref
Child Day Care	0.95 (0.89, 1.02)	0.79 (0.56, 1.11)	1.01 (0.98, 1.04)	1.05 (0.75, 1.45)
Hospital/healthcare facility	0.85 (0.80, 0.91)	0.41 (0.29, 0.58)	0.99 (0.96, 1.02)	0.63 (0.46, 0.86)
Other	0.91 (0.86, 0.97)	0.65 (0.48, 0.86)	0.99 (0.96, 1.01)	0.76 (0.58, 1.00)
Private home/residence	0.98 (0.84, 1.15)	0.86 (0.24, 3.04)	1.24 (1.15, 1.33)	8.47 (3.22, 29.28)
Restaurant	0.90 (0.80, 1.01)	0.35 (0.16, 0.74)	1.08 (1.03, 1.14)	1.53 (0.86, 2.78)
School/college/university	0.79 (0.76, 0.82)	0.27 (0.21, 0.34)	0.93 (0.91, 0.94)	0.31 (0.25, 0.39)
Season				
Winter	Ref	Ref	Ref	Ref
Fall	0.98 (0.94, 1.02)	0.89 (0.72, 1.1)	1.00 (0.98, 1.01)	0.97 (0.79, 1.19)
Spring	0.96 (0.93, 0.99)	0.82 (0.71, 0.95)	0.99 (0.97, 1.00)	0.90 (0.77, 1.04)
Summer	0.87 (0.82, 0.93)	0.53 (0.37, 0.73)	0.97 (0.94, 1.00)	0.63 (0.46, 0.86)
Outbreak Status				
Confirmed outbreak	Ref	Ref	Ref	Ref
Suspected outbreak	0.82 (0.8, 0.84)	0.41 (0.36, 0.47)	0.95 (0.94, 0.96)	0.42 (0.37, 0.48)
Census Region				
South	Ref	Ref	Ref	Ref
Northwest	0.86 (0.83, 0.89)	0.47 (0.40, 0.56)	0.93 (0.92, 0.95)	0.45 (0.38, 0.53)
Midwest	1.01 (0.98, 1.04)	0.95 (0.81, 1.12)	1.02 (1.00, 1.03)	1.01 (0.86, 1.18)
West	0.99 (0.95, 1.04)	0.98 (0.78, 1.23)	0.98 (0.96, 1.00)	0.92 (0.74, 1.15)
Year				
Jan 2009 - Jun 2009	1.25 (1.16, 1.35)	2.64 (1.83, 3.85)	1.10 (1.07, 1.14)	2.66 (1.84, 3.88)
Jul 2009 - Jun 2010	1.27 (1.18, 1.36)	2.57 (1.8, 3.69)	1.12 (1.08, 1.15)	2.64 (1.86, 3.78)
Jul 2010 - Jun 2011	1.24 (1.17, 1.31)	2.61 (1.99, 3.43)	1.09 (1.07, 1.12)	2.56 (1.95, 3.36)
Jul 2011 - Jun 2012	1.20 (1.14, 1.26)	2.21 (1.7, 2.87)	1.06 (1.04, 1.09)	2.22 (1.71, 2.89)
Jul 2012 - Jun 2013	1.05 (1.01, 1.11)	1.31 (1.03, 1.65)	1.02 (1.00, 1.04)	1.30 (1.03, 1.65)
Jul 2013 - Jun 2014	1.03 (0.99, 1.08)	1.29 (1.01, 1.65)	1.02 (1.00, 1.04)	1.25 (0.98, 1.59)
Jul 2014 - Jun 2015	1.07 (1.02, 1.12)	1.33 (1.06, 1.69)	1.03 (1.01, 1.05)	1.34 (1.06, 1.69)
Jul 2015 - Jun 2016	1.02 (0.98, 1.07)	1.12 (0.88, 1.41)	1.03 (1.01, 1.05)	1.15 (0.91, 1.45)
Jul 2016 - Jun 2016	Ref	Ref	Ref	Ref
Jul 2017 - Dec 2017	1.05 (0.99, 1.12)	1.41 (1.04, 1.91)	1.01 (0.98, 1.04)	1.37 (1.01, 1.86)

* logistic regression compares outbreaks with transmission in the third tertile (S=27%(N): R_e>3.24, S=80%(N): R_e>1.23) to outbreaks in the first tertile (S=27%(N): R_0<1.37, S=80%(N): R_e<1.09) and does not include R_0/R_e values in second tertile. Linear and negative binomial regressions use full dataset.

5 Manuscript for Aim 3

Estimating the levels of immune-escape and cross-protection for rapidly evolving norovirus strains

Molly Steele, Katia Koelle, Karen Ellis, Xiao-Li Pang, Lance Waller, Karen Levy, Ben Lopman

ABSTRACT

Noroviruses are a leading cause of acute gastroenteritis worldwide, and approximately 60% to 80% of norovirus infections are caused by a single genotype, genogroup II genotype 4 (GII.4). GII.4 noroviruses evolve rapidly, with new strains emerging every two to five years. Currently, we have little understanding of how population-level susceptibility to GII.4 norovirus changes over time and whether there is cross-immunity between different strains. We developed a model of within-cluster GII.4 norovirus transmission. We fit this model to monthly counts of GII.4 outbreaks by strain in Alberta, Canada from 2002 - 2018 to estimate changes in population-level susceptibility during strain transitions and calculate the level of cross-immunity between strains. Our model estimated that the percent susceptible to GII.4 norovirus strains ranged between 17.5% - 20.0%. We estimated that the range of cross-immunity between Farmington Hill and Hunter was 0.84 - 0.91, Hunter and Den Haag was 0.91 - 0.94, and New Orleans and Sydney was 0.34 – 0.73. These data reveal that small changes in population susceptibility are sufficient to allow new GII.4

strains to emerge, and that the level of cross-immunity is high for most strain transitions. These results suggest that norovirus vaccines containing certain GII.4 strains may provide some level of cross-protection to future strains.

INTRODUCTION

Noroviruses are highly genetically diverse, single stranded RNA viruses categorized into five genogroups (GI-GV); 33 genotypes across three genogroups (GI, GII and GIV) can infect humans.[26] Though there are many genotypes capable of infecting humans, approximately 60% to 80% of infections are caused by a single genotype, genogroup II genotype 4 (GII.4).[27,28] GII.4 noroviruses undergo punctuated antigenic change, leading to new strains emerging in populations every two to five years.[26,30–32] These antigenic changes are thought to be the result of selection pressures from population immunity; mutations lead to changes in key antigenic sites in the capsid, particularly the outermost protruding (P) domain of the capsid, which allow noroviruses to evade host immunity.[28,30,33–39] Recently, changes in the RNA polymerase, which could lead to increased transmission, of GII.4 noroviruses have been suggested as a mechanism of GII.4 evolution as well.[45,46] When new GII.4 strains emerge, they may lead to pandemics (e.g., Farmington Hills 2002 and Den Haag 2006 strains) or simply replace previous strains without disturbing the endemic pattern (e.g., New Orleans 2009 and Sydney 2012 strains), which suggests differences in the level of immune escape between strains.[28,33,40,98–100]

Currently, we have little understanding of how population-level susceptibility to GII.4 norovirus changes over time. Norovirus challenge studies have provided some insight into population level

susceptibility. Across norovirus challenge studies, the percent of participants that developed acute gastroenteritis (AGE) after challenge ranged between 27% and 80%.[53,66,83–85,89,90,97,108,109,178,179,188,189] Given that not all individuals who were challenged developed AGE, this indicates that a proportion of those challenged either were genetically resistant to infection or had acquired immunity. Secretor negative individuals are genetically resistant to infection from the certain genotypes (GI.1 and GII.4). These individuals have non-functional fucosyltransferase-2 (FUT2) genes and lack ABH antigens in saliva and mucosa,[86] thus the virus cannot bind and subsequently fails to cause infection.[80,82,83,87,88] Across challenge studies that distinguished between participants with secretor and non-secretor phenotypes, the percent of secretors who developed AGE after challenge ranged between 33% and 77%, indicating that some level of immunity to norovirus existed in these challenge populations. The majority of these studies challenged participants with GI.1 norovirus (Norwalk virus) which, unlike GII.4 norovirus, does not evolve rapidly over time.[29] Further, we have little understanding of the degree to which one GII.4 strain confers immune protection to future strains (i.e., cross-immunity). Data from *in vitro* surrogate neutralizing/blocking assays suggest immunity tends to be strain-specific, however there may be limited cross-immunity between GII.4 strains.[190–192]

Mathematical transmission models have been used to address a range of questions about norovirus such as the duration of immunity to norovirus[91] and the potential impact of age-targeted vaccination strategies.[181,193] However, all existing models of norovirus transmission assume a single strain of norovirus, or that exposure to a single strain provides immunity to all subsequent strains to which individuals are exposed.[91,121,136,138,142] These models provide important insight into the epidemiology of norovirus, however fail to capture the observed inter-annual variability that

can result from emerging GII.4 strains. Further, the estimates from these models, such as the duration of immunity and impact of vaccination, may be inaccurate in light of genetic diversity and evolution. For example, estimates of the impacts of vaccination from models that do not consider rapid evolution and emergence of new strains may be overestimated.

In this study, we sought to quantify the degree of immune escape during strain transitions and infer the level of cross-protection between five strains of GII.4 norovirus using a multi-strain modeling approach. We briefly discuss three existing GII.4 norovirus time series datasets and three multi-strain modeling approaches to explore multi-strain dynamics. Ultimately, we selected the set of coupled single-strain models to estimate population-level susceptibility to, and estimate the dynamics of five strains GII.4 norovirus in Alberta, Canada over time.

METHODS

Data

Long-term surveillance data are required to estimate changes in population-level susceptibility to and cross-immunity between GII.4 noroviruses. The ideal dataset to examine population-level changes in susceptibility would contain counts of norovirus cases or outbreaks by strain from a national surveillance system, with minimal changes in reporting over time, and cover the time from when pandemic GII.4 strains first emerged (mid-1990s) through contemporary GII.4 strains (2018/2019). Norovirus is not a notifiable disease in many countries; due to under-reporting and particularly the inconsistency of norovirus reporting over time, there are few datasets of long time

series of norovirus, and fewer still long time series of GII.4 strains of norovirus. Here, we review currently available datasets of time series of GII.4 strains.

The Centers for Disease Control and Prevention's CaliciNet GII.4 norovirus outbreak data

The Centers for Disease Control and Prevention's (CDC) CaliciNet was established in 2009 and is a national surveillance network the collects laboratory data, including genetic sequences of norovirus strains, and epidemiology data from norovirus outbreaks in the United States.[194] We obtained monthly counts of norovirus outbreaks by GII.4 strain from CDC's CaliciNet database. These data span 2009 to 2018 and include a total of 3,353 outbreaks from three GII.4 norovirus strains: New Orleans 2009, Sydney 2012, and Sydney 2015. This data set includes both strain transitions that occurred via changes in the viral capsid (i.e., New Orleans 2009 to Sydney 2012) and changes in the polymerase (i.e., Sydney 2012 to Sydney 2015). (Table 5-1, Figure 5-S1)

Public Health England GII.4 norovirus laboratory reports

Public Health England (PHE) collects laboratory reports of sporadic cases or outbreaks of norovirus; a subset lab report samples are genotyped to track viral diversity over time. [195] We obtained monthly counts of total GII.4 norovirus lab reports and monthly counts of reports where samples were sequenced by PHE. These reports come from England and Wales between January 2005 and November 2018. A total of 5,761 lab reports across five strains of GII.4 norovirus are represented in this dataset: Hunter 2004, Yerseke 2006, Den Haag 2006, New Orleans 2009, and Sydney 2012. The number of typed samples are quite sparse prior to 2011. (Figure 5-S2A) In order to fit a model to these data, we calculated the proportion of typed samples for each strain and

multiplied these proportions by total number of GII.4 reports over time and used a moving average to interpolate over missing data. (Figure 5-S2B)

Alberta, Canada GII.4 norovirus outbreak data

We also obtained GII.4 norovirus outbreak data from a literature review of norovirus genotypes. From these literature review data, we identified three studies from Alberta, Canada with monthly counts of GII.4 norovirus outbreaks.[196–198] These data span from October 2002 to December 2015 and include a total of 1,175 outbreaks from five GII.4 strains: Farmington Hills 2002, Hunter 2004, Den Haag 2006, New Orleans 2009 and Sydney 2012. Over this time series, the epidemiological pattern of norovirus shifted from biennial prior to 2011 to annual approximately one year after the emergence of GII.4 New Orleans (i.e., 2010/2011 norovirus season). (Figure 5-1A) Biennial patterns of norovirus are uncommon; however have been reported in Sweden.[199]

Models

There are several modeling approaches that can be used to simulate the dynamics of multi-strain pathogens. Here we discuss three multi-strain modeling approaches and identify which models best capture GII.4 strain dynamics.

History-based transmission models

History-based models track the individual infection histories within a population over time, with model compartments for each possible set of strains that individuals have been exposed to.[143] The

level of partial cross-immunity acquired by individuals is dependent on their infection history. There are two assumptions for how cross-immunity acts: exposure to a previous strain reduces an individual's susceptibility to infection from a subsequent strain (reduced susceptibility) and exposure to a previous strain reduces an individual's ability transmit a subsequent strain given infection (reduced transmission). Using notation described in Kucharski et al., 2016, a two strain history-based strain model has the following equations:[200]

$$\frac{dS_0}{dt} = \mu N - \beta_1(I_1 + \sigma J_1)S_0 - \beta_2(I_2 + \sigma J_2)S_0 - \mu S_0$$

$$\frac{dI_1}{dt} = \beta_1(I_1 + \sigma J_1)S_0 - (v + \mu)I_1$$

$$\frac{dI_2}{dt} = \beta_2(I_2 + \sigma J_2)S_0 - (v + \mu)I_2$$

$$\frac{dS_2}{dt} = vI_2 - \beta_2(I_1 + \sigma J_1)\tau S_2 - \mu S_2$$

$$\frac{dS_1}{dt} = vI_1 - \beta_1(I_2 + \sigma J_2)\tau S_1 - \mu S_1$$

$$\frac{dJ_1}{dt} = \beta_2(I_1 + \sigma J_1)\tau S_2 - (v + \mu)J_1$$

$$\frac{dJ_2}{dt} = \beta_2(I_2 + \sigma J_2)\tau S_1 - (v + \mu)J_2$$

$$\frac{dS_{1,2}}{dt} = v(J_1 + J_2) - \mu S_{12}$$

Where S_0 represents the number of individuals who have no infection history, I_i is the number whose first infection is with strain i (where i=1,2), S_i is the number recovered from strain i and susceptible to the next strain, J_i is the number of individuals whose second infection is with strain i, $S_{1,2}$ is the number recovered from both strains, N is the total population size, μ is the birth and

death rate, β_i is the effective contact rate for strain i, ν is the rate at which infected individuals recover from infection, and σ and τ represent the relative infectiousness and susceptibility to a second infection, respectively.

History-based models have predominantly been used to explore the dynamics of few strains as these models soon become intractable with the number of strains considered.[143] Additionally, results from a multi-strain model comparison study showed that a history-based model could not capture sequential strain replacement dynamics well. Rather, sequential strain replacement could only be replicated under a very limited set of model structure and assumptions.[148]

Status-based transmission models

Status-based transmission models track the immune status of individuals, rather than their infection history.[144] If it is assumed that cross-immunity acts to reduce transmission and immunity is polarizing (i.e., upon infection individuals will become completely immune or remain fully susceptible), status-based models can be reduced, such that the number of equations grows linearly with each strain considered. This framework can therefore be used to simulate many more strains than history-based models. The system of equations for a status-based model assuming reduced transmission is as follows:[145]

$$\frac{dS_i}{dt} = \mu N - \sum_j \beta_j S_i \sigma_{ij} I_j - \mu S_i$$

$$\frac{dI_i}{dt} = \beta_i S_i I_i - \nu I_i - \mu I_i$$

Where N is the total population size, μ is the birth and death rate, β_i is the effective contact rate for strain i, v is the rate at which infected individuals recover from infection, σ_{ij} is the level of cross immunity between strains i and j, and S_i and I_i represent the number susceptible and infected with strain i, respectively. Status-based models are capable of producing strain replacement dynamics that resemble the observed dynamics of rapidly evolving pathogens such as influenza.[145,146,148]

While the assumption that cross-immunity acts to reduces transmissibility produces a more tractable model system, this assumption may not be biologically plausible. For this assumption, cross-immunity does not prevent infection but prevents infected individuals from transmitting the pathogen to others. Additionally, individuals who have cross-immunity can gain additional immunity with continued exposure to infection. Ballesteros et al., 2009 demonstrated that this feature of the reduced transmission assumption may lead to overestimates of population-level immunity.[148] An additional complexity of status-based, as well as history-based approaches, is how to model the accumulation of cross-immunity to a strain i given previous exposure to multiple strains that provide cross-immunity to strain i. As reviewed by Wikramaratna et al., 2015, two common assumptions about the accumulation of cross-immunity are that it accumulates geometrically such that all strain exposures contribute to immunity (i.e., "product" cross-immunity), or only the strain that provides the strongest immunity contributes (i.e., "minimum" cross-immunity).[201] It is currently unclear which assumption is more appropriate as we have little understand of how immunity accumulates for noroviruses.

Quantifying the level of immune escape via within-cluster transmission models

In an attempt to better capture the observed variability of GII.4 strain dynamics we developed a set of coupled single-strain models to estimate the level of susceptibility to GII.4 strains over time. This set of coupled single-strain models is a simple Susceptible, Infected, Recovered (SIR) model that tracks the proportion susceptible to a given GII.4 strain over time (Figure 5-1A). We assume maternal immunity is negligible, therefore births enter directly into the susceptible class. Susceptible individuals (*S*) become exposed at a rate given by the force of infection (λ_i) and progress to the infected state (*I*). Infected individuals recover (*R*) from infection at a rate inversely proportional to the duration of illness (v). We assume that strain specific immunity does not wane. We model the force of infection as:

$$\lambda_i = \theta_t \beta I_i$$

Where β is the effective contact rate for all strains and I_i is the number infected with strain i. We incorporated a seasonal forcing parameter (θ_t) as follows:

$$\theta_t = 1 + \alpha\,(cos * t + \omega)$$

where α represents seasonal amplitude, ω is the seasonal offset and t represents time in years. This method allows for the consideration of many strains without being computationally intensive and assumes that all dynamics prior to the emerging strain only affect the proportion of the host population that is susceptible to the emerging strain, thus eliminating complications associated with assumptions of cross-immunity. This method however can only be applied when there is minimal or no co-circulation of strains. We therefore selected the Alberta dataset for this analysis as there is little to no co-circulation of GII.4 strains.

To estimate the number of GII.4 outbreaks by strain, we multiplied the model estimated incidence by a reporting rate that increases linearly to account for improvements in diagnostics and reporting over time:

$$\rho = mt_x + b$$

Where b is the baseline reporting rate in the first month of the dataset and m is the increase the reporting rate with each month (t_x). To fit this model, we separated the full time series of monthly counts of GII.4 norovirus outbreaks into five datasets, one for each strain (Farmington Hills, Hunter, Den Haag, New Orleans and Sydney) while it circulated. For each strain-specific dataset we truncated the data to the day before the subsequent strain emerged. This resulted in the exclusion of 2 outbreaks of Farmington Hills, 3 outbreaks of Hunter, and 30 outbreaks of Den Haag and 16 outbreaks of New Orleans (Figure 5-S3). Using maximum likelihood, we fit this model to each strain specific dataset to estimate an effective contact rate (β), seasonality parameters (α, ω), a reporting rate ($mt_x + b$) as well as the initial proportion of the population susceptible to a given strain at the time it emerges ($S_{j,t0}$). We assumed the monthly counts of strain specific GII.4 outbreaks were Poisson distributed with a mean equal to the model estimated strain specific incidence multiplied by the reporting rate. We calculated the negative log likelihood for each strain-specific model as follows:

$$\ln L_j\left(y_1 \ldots y_{n_j} \middle| \theta\right) = -n_j(\theta) - \sum_{x=1}^{n_j} \ln(y_x!) + \ln(\theta) \sum_{x=1}^{n_j} y_x!$$

Where j = 1,…,5 strains, and n_j is the number of observations for each strain-specific dataset, x is time in months, and y is the observed data and θ is the model estimated number of outbreaks. We

optimized an overall negative log likelihood which was the sum of the strain-specific negative log likelihoods.

Estimating cross-immunity

To estimate a lower bound of cross-immunity between two strains we assumed that the dynamics of a currently circulating strain only affect the proportion of the host population that is initially susceptible to the subsequent strain at the time it emerges ($S_{j,t0}$). As such, the lower bound of cross-immunity between GII.4 strains can be interpreted as the proportion of the population that would need to gain immunity to a future strain (strain *j*) given exposure to a currently circulating strain (strain *i*) to result in the proportion initially susceptible to the future strain (strain *j*) when it emerges (i.e., the value of $S_{j,t0}$ estimated through model fitting). For example, if there is perfect cross-immunity between a currently circulating strain and a future strain (σ = 1) then the proportion susceptible to the future strain and the currently circulating strain would be the same at the time the future strain emerges ($S_{j,t0} = S_{i,tf}$). If, however, the level of cross-immunity is low, we would expect that the proportion initially susceptible to a future strain would be higher than the proportion susceptible to a currently circulating strain at the time the future strain emerges ($S_{j,t0} > S_{i,tf}$). The model used to infer the lower bound of cross-immunity is as follows:

$$\frac{dS_i}{dt} = \mu N - \theta_t \beta I_i \frac{S_i}{N} - \mu S_i$$

$$\frac{dS_j}{dt} = \mu N - \sigma_{ij} \theta_t \beta I_i \frac{S_j}{N} - \mu S_j$$

$$\frac{dI_i}{dt} = \theta_t \beta I_i \frac{S_i}{N} - \nu I_i - \mu I_i$$

Where S_i and I_i are the number susceptible to or infected with the currently circulating strain, S_j is the number susceptible to the future strain, N is the total population size, μ is the birth and death rate, θ_t is the seasonal forcing parameter (as described above), β is the effective contact rate, v is the rate at which infected individuals recover and σ_{ij} is the level of cross immunity between strains i and j. We simulated from the time strain i emerges until the time strain j emerges and assumed that the initial proportion susceptible to both strains was equal to the estimated initial proportion susceptible to strain i ($S_{i,t0}$) at the time strain i emerges. We then simulated over the range of possible values for σ_{ij} (0, 1) and selected the values that resulted in the initial proportion susceptible to strain j ($S_{j,t0}$) that was previously estimated.

To estimate an upper bound of cross-immunity, we assumed that if strain i provides immunity to strain j, then strain j also provides immunity to strain i. We further assumed that the extinction of a historic strain (strain i) is driven by cross-immunity from a currently circulating strain (strain j) and that extinction occurs when the reproduction number of the historic strain is less than 1 ($R_{t,i} < 1$). Therefore the upper bound of cross-immunity was calculated as the value of σ_{ij} that drives the reproduction number (R_t) of the historic strain (strain i) below 1 after strain j emerged. We used the following model to track the proportion susceptible to a historic strain (strain i) given the dynamics of a currently circulating strain (strain j):

$$\frac{dS_i}{dt} = \mu N - \sigma_{ij}\theta_t \beta I_j \frac{S_i}{N} - \mu S_i$$

$$\frac{dS_j}{dt} = \mu N - \theta_t \beta I_j \frac{S_j}{N} - \mu S_j$$

$$\frac{dI_j}{dt} = \theta_t \beta I_j \frac{S_j}{N} - \nu I_j - \mu I_j$$

Where I_j is the number infected with strain j and all other parameters are as previously described. Based on the output from the model above we calculated the reproduction number for the historic strain over time as:

$$R_{i,t} = \frac{\beta \frac{S_{i,t}}{N}}{\nu + \mu}$$

We then simulated values of σ_{ij} from the estimated lower bound of cross-immunity to 1 (the maximum possible value of σ_{ij}) and selected the value of σ_{ij} that drove $R_{i,t}$ below 1.

Model simulation, fitting, and analysis were conducted in R version 3.5.2 using the *nloptr* and *deSolve* packages.[158,167,168]

RESULTS

Comparison of GII.4 strain datasets

As mentioned previously, to quantify the level of immune escape and cross-immunity between GII.4 norovirus strains, a long time series of incidence or outbreaks, ideally collected from a large population (e.g., national level surveillance), is required. The CDC's CaliciNet data and the PHE data are both collected from national surveillance. However, the CaliciNet data cover a relatively short time span, and contain data on only three strains and two strain transitions. Pandemic strains of norovirus have been rapidly evolving and circulating since the mid-1990s;[36,40] these CaliciNet data would only provide insight into population-level susceptibility for the most recent GII.4

strains. A longer time series that captures data of historic as well as contemporary GII.4 strains is necessary to understand changes in population-level susceptibility and cross-immunity over time.

The PHE data are a longer time series and represent a larger number of strains than the CaliciNet data. However, the proportion of typed GII.4 strain samples is quite sparse prior to 2011 (Figure 5-S2A). Based on our method to scale counts of typed samples to counts of lab reports, the resulting dataset had three strains co-circulating between 2006 and 2007: Hunter, Yerseke and Den Haag (Figure 5-S2B). Published surveillance data from several countries support the co-circulation of Yerseke and Den Haag; however Hunter showed little co-circulation with Yerseke and Den Haag.[40] As the GII.4 strain data were sparse in this dataset prior to 2011, and our method of interpolation produce patterns that contradict observed GII.4 strain trends, we did not use this dataset in our primary analyses (Text S1). Further, in our preliminary analysis of these data we found that a status based-model (detailed below) did not capture the inter-annual variability in the observed data (Text S1, Figure 5-S5).

Unlike the CaliciNet and PHE datasets, the data from Alberta, Canada are collected at the smaller, provincial scale. Additionally, these data show a biennial pattern of norovirus outbreaks, which is uncommon. While these factors may reduce the generalizability of results from analyses of this data, of all the available datasets, it is the longest, most complete time series; five GII.4 strains and four strain transitions are captured in these data (more than the CaliciNet data) and the reporting/observation of strains is more consistent over time than the PHE data. For these reasons, (in addition to our selection of model approach as detailed below) we selected the Alberta data for our primary modeling analysis.

Table 5-1. Description of time series data for GII.4 norovirus strains.

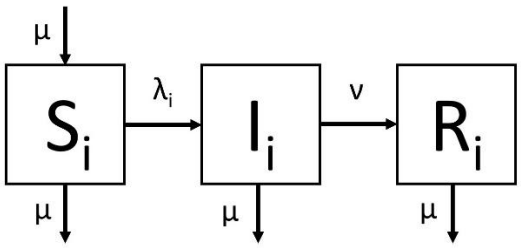
Data Source	Level of Surveillance	Time span	Data Type	GII.4 Strains represented	N[b]	Benefits	Limitations
CDC's CaliciNet	National (US)	September 2011-August 2018	Monthly outbreaks	New Orleans 2009, Sydney 2012, Sydney 2015	3,353	National level surveillance	Shorter time series with fewer strains represented
Public Health England	National (England and Wales)	January 2005 – November 2018	Monthly lab reports[a]	Hunter 2004, Den Haag 2006, Yerseke 2006, New Orleans 2009, Sydney 2010	5,761	National level surveillance and long time series with many strains represented	Data prior to 2011 sparse and requires scaling from typed strains to total lab reports
Published Studies[196,198,202]	Provincial (Alberta, Canada)	October 2002 – December 2015	Monthly outbreaks	Farmington Hills 2002, Hunter 2004, Den Haag 2006, New Orleans 2009, Sydney 2012	1,175	Long time series with many strains represented	Data are collected at a provincial level and show unusual biennial to annual pattern of disease

a. Monthly counts of strain specific lab reports were estimated by multiplying the proportion of typed GII.4 strain samples by the total number of GII.4 lab reports.

b. N represents either total number of outbreaks (CaliciNet, Published Alberta studies) or total number of lab reports (PHE data)

Based on preliminary analyses we found that the history-based model could not capture the dynamics of CDC's CaliciNet data and the status-based model was unable to capture the inter-annual variability of GII.4 strains of the PHE dataset (Figure 5-S4, Figure 5-S5, Text S1, Table 5-S1). As such, we selected the set of coupled single-strain models and the Alberta dataset for further analysis.

Figure 5-1. Coupled single-strain model structure. Births enter directly into the susceptible compartment, and susceptibles become infected by the force of infection (λ_i). Infected individuals gain immunity at rate ν; we assume immunity to a strain does not wane.



The best fit for the set of coupled single-strain models captured the GII.4 strain outbreak dynamics reasonably well (Figure 5-1B). A seasonal forcing of 6.3% of peak-to-mean amplitude provided the best fit to observed seasonal variation in monthly outbreaks across all strains (Table 5-2). The reporting rate was estimated to increase by 9.66×10^{-8} with each month, such that the initial reporting rate in October 2002 was 0.0012 and the final reporting rate in December 2015 was 0.0017. The estimated initial and final percent susceptible to each strain ranged from 17.5% to 20.0% and 17.5% to 18.4%, respectively (Table 5-3).

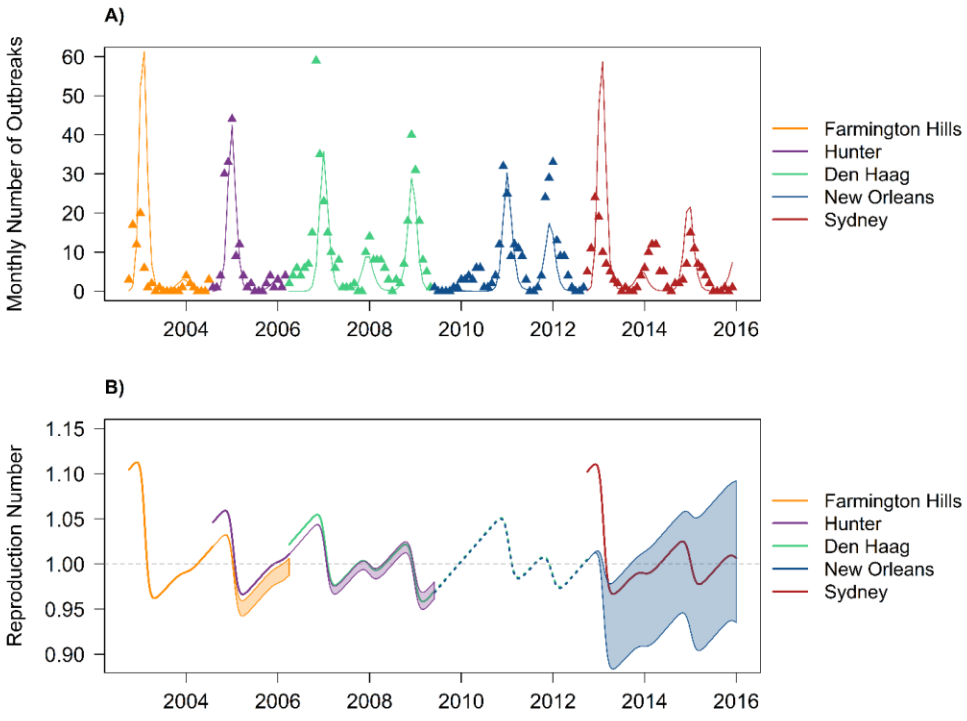
Population-level susceptibility increased slightly during the strain transitions for Farmington Hills to Hunter and Hunter to Den Haag (3% and 1% relative increase, respectively) while the transition from New Orleans to Sydney corresponded to a larger increase in susceptibility (9% relative increase). This model estimated a small decrease in susceptibility during the Den Haag to Sydney strain transition (0.13% decrease). The estimated effective reproduction numbers (R_e) for each strain ranged from 0.97 to 1.10 (Table 5-2, Figure 5-2) and the overall estimate for R_0 (assuming an entirely susceptible population) was 5.53. There were several norovirus seasons that the set of coupled single-strain models did not capture well. These models overestimated the monthly number of outbreaks for the first seasons of the Farmington Hills and Sydney strain, and underestimated the first and third seasons of the Den Haag and New Orleans strains.

Table 5-2. Fixed and estimated parameter values of the model

Parameter Description	Symbol	Value	Source
Duration of illness	$1/\nu$	2 days	Devasia et al 2014[63]
Birth and mortality rate[a]	μ	3.56e-05/day	Alberta Government[203]
Transmission parameter	β	2.76	Estimated
Reporting rate[a]	ρ	$9.66e\text{-}8 \times t_x + 0.0012$	Estimated
Seasonal amplitude	α	0.063	Estimated
Seasonal Offset	ω	1.51	Estimated

a. The mortality rate is assumed to be equal to the birth rate in Alberta
b. The reporting rate is assumed to increase linearly by month (t_x)

Figure 5-2. Model fit and change in reproduction number over time (R_t). A) Comparison of model predicted GII.4 strain dynamics (lines) with reported GII.4 strain dynamics (triangles). B) Estimates of the reproduction number for each strain ($R_{i,t}$) over time. The top border of the shaded regions represents the estimated R_t based on the lower bound of cross-immunity (σ), while the bottom border of the shaded region represents the estimated R_t based on perfect cross-immunity (σ = 1) between the historic strain and the currently circulating strain (ranges of σ for each strain is presented in Table 5-1). The estimated R_t for Den Haag after the emergence of New Orleans (green dashed line) is based on the assumption of perfect cross-immunity between these strains (σ = 1). As detailed in the text, the initial proportion susceptible to GII.4 New Orleans at the time this strain emerges ($S_{NO,t0}$ = 0.1751) is lower than the existing proportion susceptible to GII.4 Den Haag ($S_{DH,tf}$ = 0.1754), thus even assuming Den Haag provides perfect cross-immunity to New Orleans (σ = 1) cannot account for the population level susceptibility at the time New Orleans emerges.



We estimated the range of cross-immunity between Farmington Hills and Hunter ($\sigma_{FH,H}$) to be 0.84 to 0.91. This upper bound of cross-immunity leads to $R_t < 1$ for Farmington Hills 5 months after Hunter emerged. In the observed data, 3 months after the emergence of Hunter, Farmington Hills was no longer circulating. We estimated cross-immunity between Hunter and Den Haag ($\sigma_{H,DH}$) ranged from 0.91 to 0.94. An upper bound of 0.94 results in $R_t < 1$ for Hunter approximately 10 months after Den Haag emerges; the observed data show that Hunter was no longer circulating 2 months after Den Haag emerged. For a brief period in 2008, R_t of Hunter climbs above 1, however, Hunter was no longer circulating by this time (Figure 5-2). Cross-immunity between New Orleans and Sydney ($\sigma_{NO,S}$) was estimated to be low ($\sigma_{NO,S}$ = 0.34, 0.73) (Table 5-3). The upper bound of cross-immunity between New Orleans (0.73) resulted in $R_t < 1$ for New Orleans approximately 3 months after Sydney emerges (Figure 5-2). In the data from Alberta, Canada, there were no reported outbreaks of New Orleans after Sydney emerged.

Table 5-3. Estimates of initial proportion susceptible, final proportion susceptible and range of cross-immunity for each strain. The initial susceptible to each strain at the time the strain emerges ($S_{i,t0}$) was estimated through model fitting. The final susceptible to each strain at the time step prior to a subsequent strain emerging ($S_{i,tf}$) was simulated based on the best fit model parameters.

GII.4 Strain	Initial Susceptible	Final Susceptible	Range of Cross-immunity
Farmington Hills	0.200	0.184	(0.84, 0.91)
Hunter	0.189	0.183	(0.91, 0.94)
Den Haag	0.185	0.1754	> 1[a]
New Orleans	0.1751	0.182	(0.34, 0.73)
Sydney	0.199	0.182	NA[b]

a. The initial proportion susceptible to GII.4 New Orleans at the time this strain emerges ($S_{NO,t0}$ = 0.175) is lower than the existing proportion susceptible to GII.4 Den Haag ($S_{DH,tf}$ = 0.177), thus even assuming Den Haag provides perfect cross-immunity to New Orleans ($\sigma = 1$) cannot account for the population level susceptibility at the time New Orleans emerges.

b. Cross-immunity could not be estimated as there was no data for strains after GII.4 Sydney

This model estimated that there was very little change in the proportion susceptible in the population during the transition from Den Haag to New Orleans ($S_{DH,tf}$ = 0.1754; $S_{NO,t0}$ = 0.1751). If Den Haag conferred partial immunity to New Orleans (i.e., $\sigma_{DH,NO} < 1$) we would expect that the proportion susceptible would increase during this strain transition (i.e., the percent susceptible to New Orleans at the time it emerges would be greater than the final percent susceptible to Den Haag; $S_{DH,tf} < S_{NO,t0}$). With perfect cross-immunity (i.e., $\sigma_{DH,NO} = 1$) we would expect that susceptibility would not change during this strain transition ($S_{DH,tf} = S_{NO,t0}$). However, the initial susceptible to New Orleans is slightly lower than the final susceptible to Den Haag. Thus even assuming Den Haag provides perfect cross-immunity to New Orleans ($\sigma_{DH,NO} = 1$) cannot account for a decrease in population-susceptibility when Sydney emerges. This suggests that cross-immunity from Den Haag alone cannot account for the initial proportion susceptible to GII.4 New Orleans at the time it emerges.

DISCUSSION

The results from this analysis provide insight into the degree of immune escape that have driven GII.4 strain dynamics over time. Our models estimated that population-level susceptibility to GII.4

strains as they emerged ranged between 17.5% - 20.0%, and that small changes in the fraction of the population susceptible are sufficient to allow new strains to emerge and, though cross-immunity, drive previous strain extinct. Based on the estimated changes in population-level susceptibility, we estimated that the level of cross-immunity between Farmington Hills and Hunter and Hunter and Den Haag was high, and low between New Orleans and Sydney. Interestingly, we found that population-level susceptibility decreased slightly during the transition from Den Haag to Sydney, suggesting factors other than cross-immunity that we did not account for in our model, may be affecting population-level susceptibility during this strain transition.

There are few studies that provide insight into population susceptibility to GII.4 norovirus. In two challenge studies of GII.4 norovirus, 43% and 58% of participants were susceptible to infection while 30% in both studies developed acute gastroenteritis given infection.[108,178] Data from a surrogate neutralization assay of human sera collected between 1979 and 2010 that 57.5%, 36%, and 28.2% of the population would be susceptible to the 1987 (MD145), 1995 (Grimsby), and 2001 (Houston) strains of GII.4.[191] Our estimates of population-level susceptibility to GII.4 norovirus strains are substantially lower. One potential explanation for this is that we did not account for individuals who may be genetically resistant to infection from GII.4 norovirus. Individuals who have non-functional FUT2 genes (i.e., non-secretors) do not express HBGA antigens in saliva and mucosa.[86] As such, non-secretors are almost completely resistant to GI.1 and GII.4 norovirus infection.[82,83,87,138,178] Approximately 20% of populations of European descent are non-secretors. Therefore, our values of susceptibility, among the genetically susceptible, may be underestimated.

Studies of antigenic variation, molecular epidemiology and evolution of GII.4 strains provide context for our estimates of cross-immunity between GII.4 norovirus strains. Four studies have reported that earlier GII.4 strains are cross-reactive with contemporary strains, suggesting that cross-immunity may persist through time.[118,191,204,205] Data from two studies using enzyme immunoassays show that monoclonal antibodies derived from GII.4 Farmington Hills and GII.4 Den Haag were cross-reactive with GII.4 Hunter.[94,205] This suggests that there may be cross-immunity between these strains and supports our finding of high cross-immunity for the Farmington Hills to Hunter and Hunter to Den Haag strain transitions. As mentioned above, even assuming perfect cross immunity between GII.4 Den Haag and GII.4 Sydney does not fully explain the decrease in population susceptibility to Sydney at the time it emerges. However, the genetic sequences of key antigenic regions of Den Haag and Sydney have been reported to be highly similar, which suggests there may be high levels of cross-protection between these strains.[206] Data from antibody binding and blockade assays show that recognition of key binding regions of GII.4 New Orleans and GII.4 Sydney are very different, which suggests that there may be little cross-immunity between these strains.[23] These data support our finding that cross-immunity between New Orleans and Sydney is low. However, our estimate for cross-immunity between New Orleans and Sydney must be interpreted with caution, as our model did not capture this strain transition well.

There are a number of limitations to this study. First, while our model provides reasonably good fits to the observed GII.4 strain dynamics, there are several norovirus seasons where the predictions overestimate (Farmington Hills in 2002; Sydney in 2013 and 2015) or underestimate the data (Den Haag in 2007). Our estimates of cross-immunity entirely depend on the estimated

change in the level of susceptibility at the time a strain emerges. As our model overestimated the number of outbreaks during the first season of Sydney (suggesting that population-level susceptibility was higher than observed), the corresponding value of cross-immunity for this strain transition ($\sigma_{NO,S}$) may be underestimated. Similarly, our model underestimated the first season of Den Haag (i.e., lower predicted susceptibility than observed) thus cross-immunity between Hunter and Den Haag ($\sigma_{NO,S}$) may be overestimated.

Second, we assumed that cross-immunity only occurs between two strains that are consecutive (e.g., Farmington Hills provides cross-immunity to Hunter but does not provide cross-immunity to Den Haag). Evidence from *in vitro* studies utilizing surrogate neutralization assays and enzyme immunoassays have indicated that some GII.4 strains, such as US 95/96 and Farmington Hills, are broadly cross-reactive with other GII.4 strains,[204,205] suggesting that cross-immunity between strains that are not adjacent in time. Further, we assumed that cross-immunity was symmetrical (i.e., if Farmington Hills provides a certain level of immunity to Hunter, then Hunter provides that same level of immunity to Farmington Hills), however this may not be true. In a recent study, Tamminen et al. demonstrated that sera from mice immunized with GII.4 1999 VLPs blocked binding of GII.4 2012 VLPs, while sera from mice immunized with GII.4 2012 VLPs did not block binding of GII.4 1999 VLPs.[105] However, VLPs are non-infectious particles, therefore immune responses elicited through natural infection may be different. Our assumptions of cross-immunity are certainly simplifications, yet a fundamental understanding of norovirus cross-immunity is lacking, thus we currently cannot determine which assumptions are more biologically relevant for noroviruses.

A final limitation is that the data from Alberta exhibit a biennial epidemic pattern that shifts to annual shortly after New Orleans emerges. Biennial patterns of norovirus are uncommon, though they have been documented in Sweden as well.[199] Thus our findings of population-level susceptibility and cross-immunity may not be generalizable to regions that have annual patterns of norovirus. Further, we did not attempt to explain the source of these biennial patterns as it was beyond the scope of this analysis. However, biennial patterns of disease may be driven by demographic factors such as lower birth rates and net migration rates (i.e., low influx or reduction of susceptible individuals in the population)[207] or factors associated with the virus (e.g., increased virulence). Ideally, we would analyze a dataset that shows a more characteristic, annual epidemic pattern of disease to make our results more generalizable; however, the other existing datasets of GII.4 strain dynamics with annual patterns of outbreaks have critical limitations as well (Text S1).

The extent of immune-escape and cross-immunity between GII.4 norovirus strains has important implications for the design of vaccines. If the level of cross-immunity between current and future strains of norovirus is high, then norovirus vaccines may not need to be reformulated regularly. If, however, the level of cross-immunity between current and future strains is low, then norovirus vaccines may need to be reformulated frequently to keep pace with GII.4 evolution. We found that certain strain transitions (e.g., Farmington Hills to Hunter, Hunter to Den Haag) are associated with high levels of cross-immunity. Our results add to the growing body of evidence that vaccine formulations containing carefully selected strains may provide protection against future strains GII.4 norovirus.[118,191,204] However, our results cannot provide insight into whether norovirus vaccines will need to be re-formulated to keep pace with GII.4 evolution. Future studies are required to further estimate cross-immunity between a broad range of GII.4 strains.

SUPPLEMENTAL

Text S1

Figure 5-S1. CDC's CaliciNet data of monthly counts of outbreaks by GII.4 strain in the US.

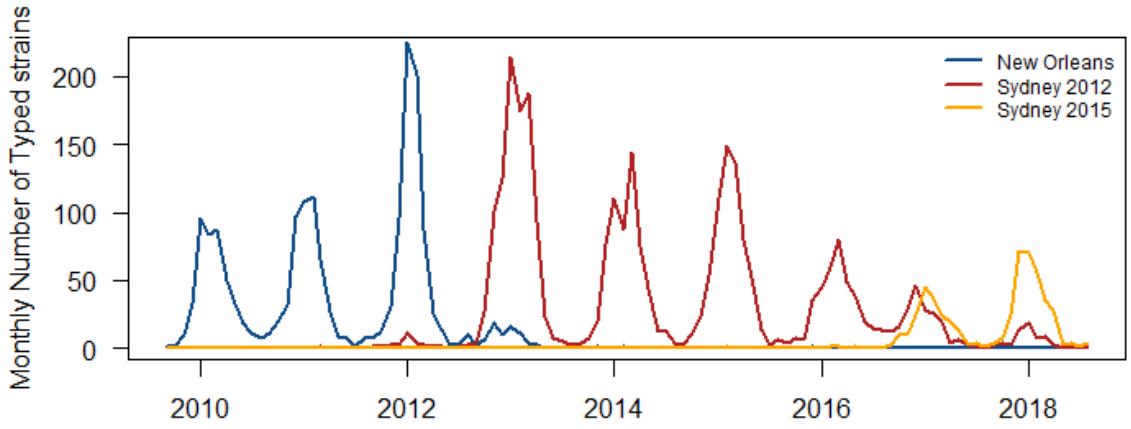


Figure 5-S2. Monthly counts of GII.4 reports to Public Health England. (A) Raw data of counts of sequenced samples of GII.4 norovirus. (B) Approximated monthly counts of GII.4 norovirus.

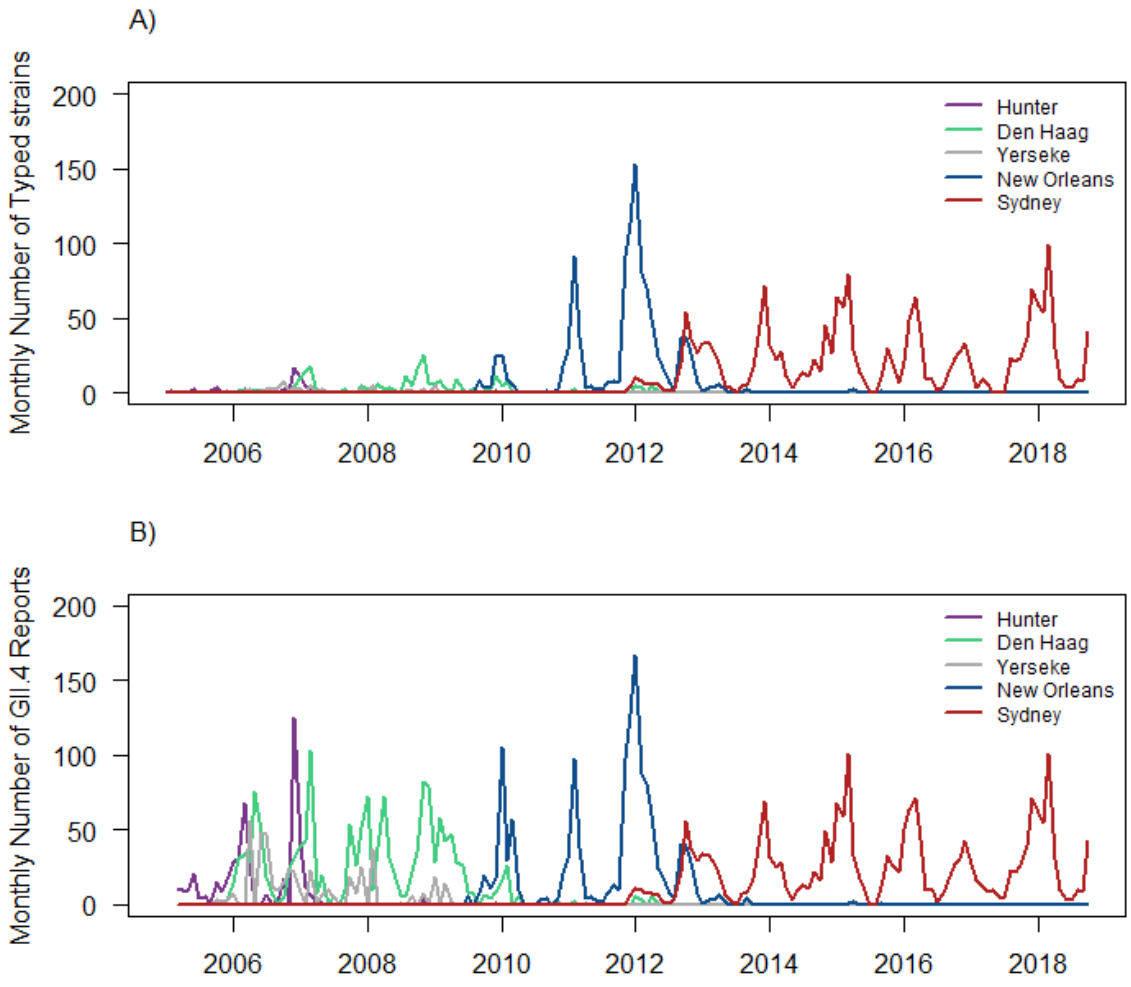
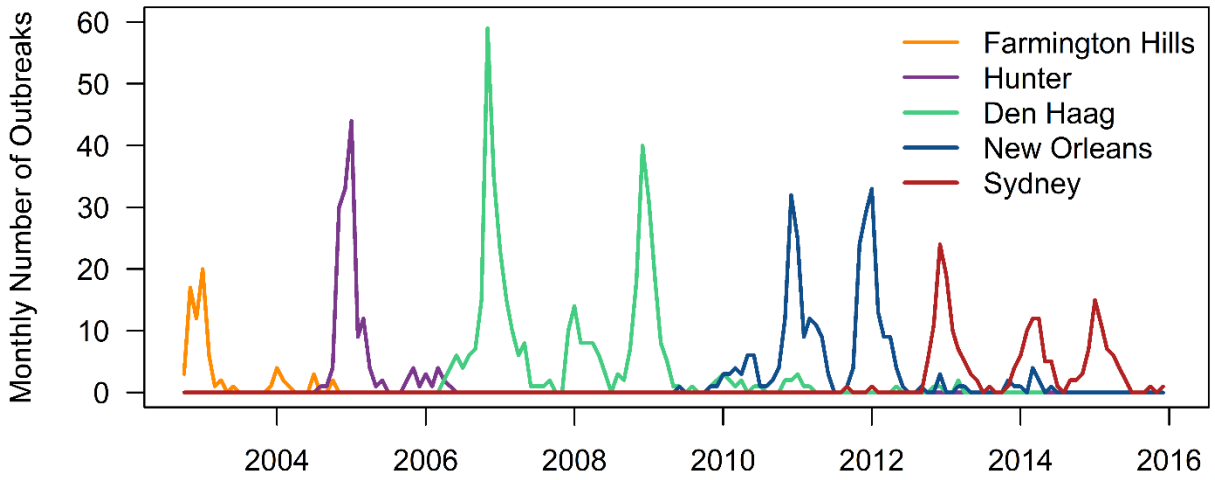


Figure 5-S3. Monthly counts of GII.4 outbreaks reported in Alberta, Canada.



Alternative multi-strain model approaches

For the first step of this analysis, we compared the ability of a history-based model and a status-based model to capture the dynamics of a single strain transition: GII.4 New Orleans to GII.4 Sydney in the US (CaliciNet data). For both models, we assumed that the two strains have the same natural history parameters and transmission parameter. The system of equations for this two strain history-based model follow those presented in the main text, however we included tracking of asymptomatic infection (A_i and B_i are the number of individuals with their first and second, respectively, infections with strain i) and assumed that asymptomatic infections contribute to transmission and are 5% as infectious as symptomatic infection. To simulate seasonality, we applied a seasonal forcing parameter ($\theta(t)$) that governs the peak-to-mean amplitude in transmissibility:

$$\theta_t = 1 + \alpha\,(cos * t + \omega)$$

where α represents seasonal amplitude, ω is the seasonal offset and t represents time in years.

Following the notation developed by Gog and Grenfell, 2009, our system of equations for the status-based model is as follows:

$$\frac{dS_i}{dt} = \mu N - \sum_j \sigma_{ij}\theta_t\beta\frac{I_j + \varepsilon A_j}{N}S_i - \mu S_i$$

$$\frac{dI_i}{dt} = \theta_t\beta\frac{I_i + \varepsilon A_i}{N}S_i - (\mu + \nu)I_i$$

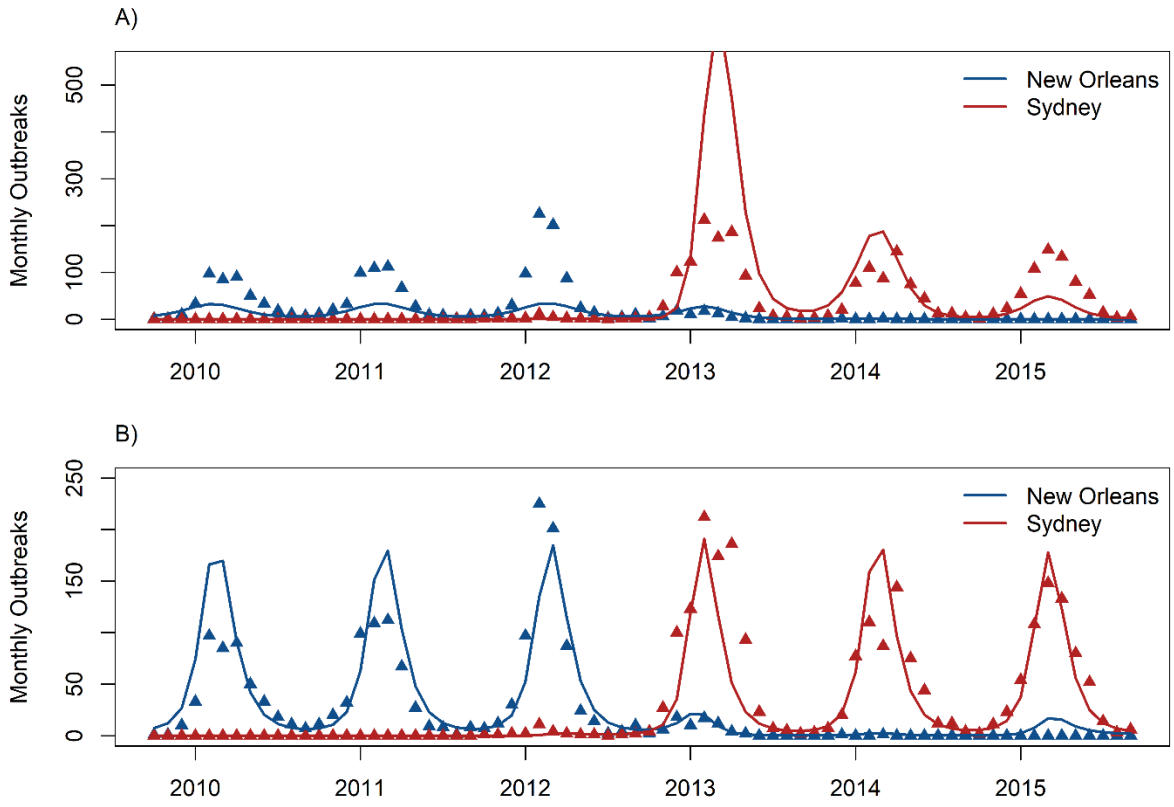
$$\frac{dA_i}{dt} = \nu I_i - (\mu + \eta)A_i$$

Where i=1,2, S_i are the number of people susceptible to strain i, I_i and A_i are the number symptomatically and asymptomatically infected with strain i, N is the total population size, μ is the birth and death rate, β is the effective contact rate for both strain, v is the rate at which infected individuals recover from symptomatic infection, v is the rate at which infected individuals recover from symptomatic infection, η is the rate at which infected individuals recover from asymptomatic infection, and σ is the level of cross-immunity (i.e., reduced infectivity).

To estimate the reported number of strain-specific outbreaks for each model, we multiplied the projected disease incidence by the probability a reported outbreak would occur given a case of norovirus (ρ). We fit the model to data from CDC's CaliciNet on monthly counts outbreaks in the US from GII.4 New Orleans and GII.4 Sydney using maximum likelihood to estimate the transmission (β), seasonality ($\theta(t)$), reporting rate (ρ), and cross-immunity (σ) parameters. We assumed the monthly numbers of outbreaks were Poisson distributed with mean equal to the model estimated strain specific incidence multiplied by the reporting rate.

The status-based model provided a better fit and was better able to capture the shift from GII.4 New Orleans and GII.4 Sydney from the observed Calicinet data (Figure 5-S3A, 5-S3B). The level of cross-protection (σ) was estimated to be high, such that 91% of those exposed to New Orleans do not contribute to transmission when infected with Sydney (Table 5-S1).

Figure 5-S4. Comparison of estimated GII.4 outbreaks (lines) from two strain history-based (A) and status-based (B) models. Triangles represent the observed monthly counts of GII.4 outbreaks reported to CDC's CaliciNet.



As the status-based model captured this strain transition best, we expanded this model from two strains to five strains. Following the notation from Gog and Swinton, the system of equations for this 5 strain status-based model are:

$$\frac{dS_J}{dt} = \mu\delta_{J,\emptyset}N - \sum_{i,K} C(K,J,i)\theta_t\beta\frac{I_i}{N}S_K - \sum_{i\notin J}\theta_t\beta\frac{I_i}{N}S_J - \mu S_J$$

$$\frac{dI_i}{dt} = \theta_t\beta\frac{I_i}{N}S_J - (\mu+\nu)I_i$$

Where i= 1,…,5 strains, J represents all possible sets of strains that individuals have immunity to, S_J is the number of individuals immune to the set of strains J and susceptible to all other strains, I_i is the number of individuals infected with strain i, and N, μ, β, $\theta(t)$ and v are as defined for the two-strain history- and status-based models. Individuals are born susceptible to all strains, thus the birth rate ($\mu\delta_{J,\emptyset}$) is:

$$\mu\delta_{J,\emptyset} = \begin{cases} 1 & \text{if } J = \emptyset \\ 0 & \text{else} \end{cases}$$

$C(K,J,i)$ represents the number of individuals who become immune to strains in set J after infection with strain i and had immunity to strains in set K:

$$C(K,J,i) = \begin{cases} \prod_{j \in J \setminus K} M_{i,j} \prod_{j \notin J} (1 - M_{i,j}) & \text{if } i \notin K \text{ and } \mathrm{K} \subset \mathrm{J} \\ 0 & \text{else} \end{cases}$$

Where M is the matrix of cross-immunity ($M_{i,j}$) (i.e., infection with strain i provides cross-immunity to strain j):

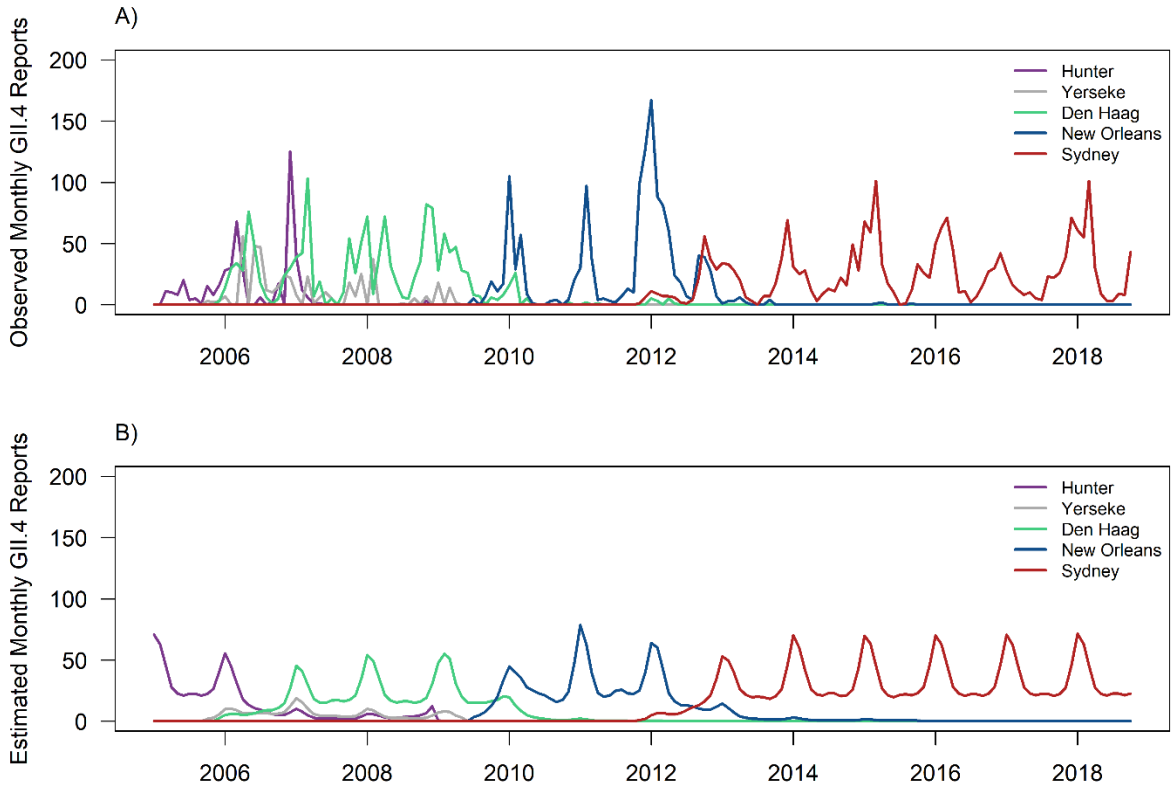
$$M = \begin{pmatrix} 1 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} & \sigma_{1,5} \\ \sigma_{2,1} & 1 & \sigma_{2,3} & \sigma_{2,4} & \sigma_{2,5} \\ \sigma_{3,1} & \sigma_{3,2} & 1 & \sigma_{3,4} & \sigma_{3,5} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & 1 & \sigma_{4,5} \\ \sigma_{5,1} & \sigma_{5,2} & \sigma_{5,2} & \sigma_{5,4} & 1 \end{pmatrix}$$

Cross-immunity was assumed to be symmetrical, such that $\sigma_{ij} = \sigma_{ji}$. We fit this model to data from PHE's data on monthly counts of GII.4 laboratory reports using maximum likelihood to estimate

the transmission (β), seasonality ($\theta(t)$), reporting rate (ρ), and cross-immunity (σ_{ij}) parameters. We assumed the monthly number of lab reports were Poisson distributed with mean equal to the model estimated strain specific incidence multiplied by the reporting rate (ρ).

The best fit for this model did not capture the inter-annual variability of the data (Figure 5-S4B). Further, the estimated value for the effective contact rate (β) results in an estimated $R_0 = 57.3$, which is much higher than R_0 values estimated from other published norovirus transmission models (range of 1.1 – 7.2).[60] Generally, the estimates of cross-immunity were quite high, and cross-immunity was estimated to decline as the amount of time increased between the strains considered (e.g., cross-immunity for Hunter to Sydney and Yerseke to Sydney were 0.683 and 0.579, respectively).

Figure 5-S5. A) Observed monthly counts of GII.4 laboratory reports submitted to Public Health England. B) Estimates of monthly counts of GII.4 reports from status-based model.



Comparison of cross-immunity estimates from all multi-strain models

Despite having different assumptions of cross-immunity, different datasets and different model structures, the five strain status-based model and the set of coupled single-strain models provided similar estimates of cross-immunity for the Hunter to Den Haag (0.972 vs. 0.91-0.94, respectively) and Den Haag to New Orleans (0.999 vs. >1) strain transitions. The set of coupled single-strain models however, estimated a much lower level cross-immunity for the New Orleans to Sydney

strain transition (range: 0.34, 0.73) relative to both the two and five strain status-based model (0.916 and 0.995, respectively). The discrepancy between these estimates of cross-immunity may indicate that there were true differences in the New Orleans to Sydney strain transition in Alberta, Canada relative to the same strain transition in England and Wales (PHE data). However, given the differences between these models we cannot determine whether this discrepancy reflects true differences between countries or if it is the result in differences in assumptions, goodness of fit and/or model structure.

Table 5-S1. Comparison of data, strains, parameters, assumptions and estimates from all multi-strain modeling approaches.

	2 Strain History-Based Model	**2 Strain Status-Based Model**	**5 Strain Status-Based Model**	**5 Strain Coupled Single-Strain Models**
Data	CaliciNet	CaliciNet	PHE	Alberta
Strains modeled	New Orleans, Sydney	New Orleans, Sydney	Hunter, Yerseke, Den Haag, New Orleans, Sydney	Farmington Hills, Hunter, Den Haag, New Orleans, Sydney
Fixed parameters				
Daily birth/death rate	3.39×10^{-5}	3.39×10^{-5}	3.18×10^{-5}	3.56×10^{-5}
Duration of symptomatic infection	2 days	2 days	2 days	2 days
Duration of asymptomatic infection	10 days	10 days	NA	NA
Relative infectiousness of asymptomatics	0.05	0.05	NA	NA
Estimated parameters				
Effective contact rate (β)	1.46	17.13	28.65	2.76
Basic Reproduction Number (R_0)	2.91	34.26	57.30	5.53

Seasonal amplitude	0.117	0.018	0.045	0.063
Seasonal offset	4.04	3.65	4.24	1.51
Reporting rate	$5.74X10^{-7}$	$1.64x10^{-4}$	$2.18x10^{-7}$	$9.66e\text{-}8 \times t_m + 0.0012$
Cross-Immunity Assumption	Reduces susceptibility	Reduces infectivity	Reduces infectivity	Reduces susceptibility
Farmington Hills, Hunter	NA	NA	NA	(0.84, 0.91)
Hunter, Yerseke	NA	NA	0.982	NA
Hunter, Den Haag	NA	NA	0.972	(0.91, 0.94)
Hunter, New Orleans	NA	NA	0.843	NA
Hunter, Sydney	NA	NA	0.683	NA
Yerseke, Den Haag	NA	NA	0.995	NA
Yerseke, New Orleans	NA	NA	0.985	NA
Yerseke, Sydney	NA	NA	0.579	NA
Den Haag, New Orleans	NA	NA	0.999	$> 1^a$
Den Haag, Sydney	NA	NA	0.984	NA
New Orleans, Sydney	0.221	0.916	0.995	(0.34, 0.73)

a. The initial proportion susceptible to GII.4 New Orleans at the time this strain emerges ($S_{NO,t0} = 0.175$) is lower than the existing proportion susceptible to GII.4 Den Haag ($S_{DH,tf} = 0.177$), thus even assuming Den Haag provides perfect cross-immunity to New Orleans ($\sigma = 1$) cannot account for the population level susceptibility at the time New Orleans emerges.

6 Conclusion

Noroviruses are becoming the primary cause of acute gastroenteritis globally, associated with an estimated 140 million to 677 million AGE cases and 71,000 to 212,000 deaths annually.[5–7] Norovirus vaccines are currently in development; however, key biological and epidemiological factors of norovirus could present challenges to vaccine development and implementation. The overall goal of this dissertation was to provide insight into key aspects of norovirus epidemiology that could pose challenges to vaccine development and implementation. Specifically, this research examined how norovirus transmission varies between age groups and how this variability could affect vaccine strategies with implications for implementation (Aim1); characterized norovirus transmission across different settings for norovirus outbreaks (Aim 2); and quantified that level of immune escape and cross-immunity between multiple GII.4 norovirus strains (Aim 3). This research provided insight into the variability of norovirus transmission and changes in population-level susceptibility and cross-immunity as new GII.4 strains emerged, and highlighted areas for further research.

6.1 Contribution of Aim 1

Norovirus vaccine evaluations have predominantly been trialed among adults; however, noroviruses affect all ages and young children and the elderly disproportionately suffer the more severe outcomes of norovirus infections. Further, observational studies suggest that young children may be an important driver of norovirus transmission.[16,61] Through Aim 1, we provide evidence for the benefits of age-targeted vaccination strategies by employing a better understanding of how transmission of norovirus varies by age. Ultimately, these results can be used to guide norovirus

vaccine development in private industry, inform policy makers on the potential impacts of alternative norovirus vaccine programs, and provide insight into the impact that vaccination programs could have on the population-level dynamics of norovirus transmission in the US.

In Aim 1, we had the unique opportunity to influence the direction of vaccine development by estimating the potential impact of different, age-targeted vaccination strategies with a dynamic transmission model of norovirus. At the time that we were conducting analyses for Aim 1, a bivalent, intramuscular VLP vaccine developed by Takeda Pharmaceuticals, was on track to enter Phase IIb/III efficacy studies among adults.[12] Through this work, we provided evidence that targeting pediatric populations for vaccination can maximize population-level impacts due to reductions in disease transmission. These findings are in line with impacts observed from other pediatric vaccines. The introduction of pediatric rotavirus and pneumococcal vaccines in the US led to large indirect benefits among unvaccinated populations through reductions in the overall force of infection.[163,164]

Further, a recently published second model of norovirus transmission and vaccination in Germany also found targeting the pediatric age group for vaccination provides population-level benefits; however, these impacts were not as great as those we estimated. This discrepancy is due to differences in assumptions of vaccine action. We assumed that vaccination does not prevent infection but disease, while Gaythorpe et al. examined many assumptions of vaccine action (e.g., reduction in susceptibility, prevention of symptoms).[193] Despite different assumptions of vaccine action, the findings from our study and Gaythorpe et al.'s study clearly argue for refocusing on a

clinical development plan that will deliver a vaccine with a safety and efficacy profile suitable for use in children. Future studies modeling studies should focus on estimating which vaccine schedules among young children will lead to the greatest reductions of disease.

From our analyses we found that the elderly contribute little to norovirus transmission, and subsequently, vaccinating this age group provided minimal population-level benefits. However, the epidemiological data we used for model fitting are estimates of US hospitalization rates informed by community incidence rates from a UK study (as norovirus surveillance in the US only captures outbreaks). The incidence estimates among older age groups from this UK study were low and may have been biased downwards.[16] As such, an important avenue for future research will be to improve the surveillance of norovirus among the elderly. Improving surveillance in this age group we will lead to better estimates of the burden of disease and subsequently we will be better able to determine the population-level value of vaccination within this group.

Our finding that the model estimated impact of both pediatric and elderly vaccine strategies are strongly sensitive to the assumed duration of vaccine immunity and age-specific contribution to transmission highlights the need for future studies that better estimate these parameters. Currently, we have limited empirical data to inform these parameters because norovirus transmission is largely unobservable,[121] and thus far there are published no vaccine studies have included long-term follow-up for clinical outcomes.[111] Observational studies of transmission (such as secondary household transmission studies) and clinical trials that can estimate the duration of vaccine protection, particularly among children <5 and the elderly, will be vital to improving our estimates of vaccine impact.

Our work in Aim 1 provides an example of the value of using mathematical models to estimate vaccine impacts prior to the licensing and distribution of vaccines, and further encourages open communication and collaboration with industry partners to best achieve public health goals. The results from our model challenged the direction of norovirus vaccine development, which at the time was directed towards adults. Since presenting and subsequently publishing our Aim 1 study findings, Takeda has conducted at least two safety and immunogenicity studies among young children (<5 years old) and older adults (60+).[208,209] Continued interactions with vaccine developers and public health decision makers will be crucial to maximizing the success of norovirus vaccines.

6.2 Contribution of Aim 2

As demonstrated through Aim 1, harnessing a better understanding of how norovirus transmission varies by age allowed us to better understand which vaccination strategies will maximize population-level benefits. However, public health interventions (such as vaccination strategies) required to mitigate or prevent outbreaks may differ substantially from population-level interventions, as outbreaks generally affect sub-populations (e.g., food handlers, immunocompromised, healthcare workers) whose risk for transmission and disease may be different from what is observed at a population-level.[150] The findings from Aim 2 provide insight into which factors (e.g., setting, season and geographic region) are associated with increased transmissibility of norovirus at the outbreak-level.

In contrast to our finding that the elderly contribute little to transmission from Aim 1 (and thus targeting them for vaccination would provide minimal population-level benefits), our analysis of norovirus transmission at the outbreak-level indicates that transmission in long-term care/assisted living facilities (where the majority of individuals are elderly) is higher than other common settings (e.g., schools/colleges/universities, hospitals). These results provide evidence that while the elderly may not be important drivers of transmission at the population-level, transmission among elderly populations during outbreaks in long-term care/assisted living facilities pose an important public health burden. Further, the elderly suffer the vast majority of norovirus related deaths,[15] therefore vaccination of the elderly may be considered if to reduce norovirus associated mortality and/or mitigate the burden of outbreaks in long-term care/assisted living facilities. Additionally, our finding that transmission of norovirus is higher for outbreaks that occur in winter suggests that one possible vaccine strategy could be a seasonal vaccine policy that provides prophylactic vaccination within outbreak settings prior to winter as is done with other pathogens with strong seasonality, such as influenza.[210] However more research is required to determine which seasonal vaccine policies would be most effective.

This analysis highlights that uncertainties surrounding norovirus transmission in outbreak settings are driven both by a lack of understanding of biological factors of norovirus as well as programmatic factors related to surveillance and reporting. First considering biological factors, there is a large amount of asymptomatic transmission of norovirus, yet we have very limited understanding of the relative roles that symptomatic and asymptomatic infections in transmission.[121,187] Further, asymptomatic transmission, which we did account for in this analysis, generally goes undetected in surveillance and can limit the effectiveness of traditional control

methods focused on ill individuals. Additionally, we have little understanding of the level of acquired immunity and susceptibility in populations prior to outbreaks, which influence the extent of transmission and subsequently the severity of outbreaks. A better understanding of the role of asymptomatic transmission and levels of acquired immunity and susceptibility in populations will help guide our understanding of which interventions, and the level of control measures, are required to mitigate/prevent outbreaks.

Among programmatic and surveillance factors, the exposed population size, which is important for estimating transmission and evaluating the extent of outbreaks, is difficult to quantify and is not consistently reported to NORS. Further, there is tremendous variability in outbreak reporting between states (approximately 100-fold difference between the highest and lowest reporting states) [120] and variability in the quality of reporting between outbreaks (i.e., smaller, probable norovirus outbreaks may be less well investigated than confirmed outbreaks, with lower rates of case ascertainment). These uncertainties, both from biological and programmatic factors, emphasize the need for improvements in norovirus surveillance and reporting, in concert with studies that better characterize the role and prevalence of asymptomatic transmission and population-level susceptibility to norovirus.

6.3 Contribution of Aim 3

The rapid evolution of GII.4 noroviruses poses a potentially significant challenge for vaccine development and implementation, yet we have little understanding of how susceptibility changes over time and whether cross-immunity exists between different strains. In Aim 3 we compared three existing GII.4 norovirus time series datasets and three multi-strain modeling approaches to

determine which dataset and method were best suited to describe GII.4 strain dynamics, estimate changes in population-level susceptibility over time, and estimate the level of cross-immunity between different GII.4 strains. Ours is the first norovirus transmission modeling study to incorporate GII.4 strain evolution.

6.3.1 Comparison of GII.4 norovirus datasets

Our review of GII.4 norovirus time series data highlights a critical need for better surveillance of norovirus, particularly of the distribution of GII.4 norovirus over time. We were only able to identify, and subsequently obtain, three datasets of GII.4 norovirus strain distributions over time and each of the three datasets had critical limitations. The CaliciNet data represent a relatively short time span and only capture the most recent strains, thus would only provide insight of susceptibility and cross-immunity for contemporary strains. In the PHE data there were large inconsistencies in reporting over time. As such, we used a method to scale these data to overall GII.4 reports and interpolated over missing data. Uncertainty surrounding the number and distribution of GII.4 strain reports in these early years subsequently led to uncertainties in the estimates of cross-immunity from these early strains. Finally, the Alberta data contained few observations (due to collection in a small population) and exhibited an uncharacteristic biennial-to-annual epidemic pattern of norovirus, thus limiting generalizability of analysis from these data to regions where norovirus has an annual epidemic pattern. Moving forward, norovirus surveillance systems should focus on improving reporting rates of norovirus and incorporate consistent sequencing of noroviruses at the genogroup-, genotype- and strain-level. Further, if there are existing case and/ outbreak samples from historic surveillance data, retrospective sequence analyses should be conducted to determine norovirus genotype- and strain-distributions.

6.3.2 Comparison of multi-strain models

This work also contributes to the growing body of research into methods for modeling multi-strain pathogens. In our preliminary analyses, we found that a history-based model could not capture norovirus strain transmission dynamics. This finding is consistent with a previous modeling study that found a history-based model failed to capture influenza strain transmission dynamics.[148] Further, we showed that while a status-based approach was able to capture the average trend of strain dynamics over time, it failed to capture observed inter-annual variability associated with the emergence of new strains. An important consideration for multi-strain models is striking the balance between biological realism and computational tractability. History-based models employ assumptions of cross-immunity that are more biologically relevant, however are computationally intractable when considering many strains. On the other hand, status-based models are tractable for considering many strains, however this computationally efficiency is gained through assumptions of cross-immunity that may be biologically unreasonable (i.e., exposure to a previous strain reduces an individual's ability transmit a subsequent strain given infection, reduced infectivity).[200,201] To overcome these issues, we developed a set of coupled single-strain models to capture the dynamics of GII.4 strain transitions and estimate changes in population-level susceptibility over time. This approach is tractable when modeling many strains and does not require assumptions that lack biological realism. A limitation of this approach however is that it can only be used when there is little to no co-circulation of strains, as was observed in the Alberta dataset.

6.3.3 Estimates of susceptibility to and cross-immunity between GII.4 norovirus strains

The results from our set of coupled single-strain models provide insight into the degree of immune escape that has driven GII.4 strain dynamics over time. Our models estimated that population-level susceptibility to GII.4 strains ranged between 17.5% - 20.0%, and that small changes in the fraction of the population susceptible are sufficient to allow new strains to emerge and, though cross-immunity, drive previous strains extinct. There are very few existing studies that provide context for our findings of population-level susceptibility, though data from challenge studies and a surrogate neutralization study suggest that population-level susceptibility may be higher than our estimates.[108,178,191] However, our model does not account for genetic resistance (i.e., 20% of Caucasian populations are genetically resistant to GI.1 and GII.4 infections) which may explain why our estimates of susceptibility are lower than challenge and surrogate neutralization studies. To better determine changes in population-level susceptibility to GII.4 norovirus over time, serological studies should be conducted, particularly focusing on times when new GII.4 strains have emerged.

Using our set of coupled single-strain models, we estimated that the level of cross-immunity between Farmington Hills to Hunter and Hunter to Den Haag was high, and low between New Orleans and Sydney. Our analysis indicated that population-level susceptibility decreased slightly during the transition from Den Haag to Sydney, suggesting factors other than cross-immunity that we did not account for in our model, may be affecting population-level susceptibility during this strain transition. Despite having different assumptions of cross-immunity, different datasets and different model structures, our five-strain status-based model and set of coupled single-strain models similarly predicted high cross-immunity between the Hunter to Den Haag (0.972 vs. 0.91-

0.94, respectively) and Den Haag to New Orleans (0.999 vs. >1) strain transitions. Further, data from existing studies utilizing enzyme immunoassays and surrogate neutralization assays have suggested the existence of cross-immunity between the Farmington Hills to Hunter, Hunter to Den Haag, and Den Haag to Sydney strain transitions.[94,205,206] The high-levels of cross-immunity for these strains suggest that vaccines that contain these strains may be broadly protective. However, further research of cross-immunity between these strains should be conducted, as these *in vitro* studies relied on non-infectious VLPs, therefore immune responses elicited through natural infection may be different. Importantly, future vaccine efficacy studies should test heterotypic challenge with GII.4 norovirus to determine the level of protection against GII.4 strains that are not part of the vaccine formulation.

The results from our set of coupled single-strain models and five strain status-based model were contradictory on the level of cross-immunity between New Orleans and Sydney. The linked single-strain model estimated a much lower level cross-immunity for the New Orleans to Sydney strain transition (range: 0.34, 0.73) relative to both the status-based model (0.916 and 0.995, respectively); however given the differences in data, model structures, and goodness of fit we cannot determine what drives the difference between these estimates. Notably the linked single-strain model overestimated population-level susceptibility during the first season of Sydney, thus cross-immunity is likely underestimated. Data from *in vitro* antibody binding and blockade assays demonstrate that recognition of key binding regions of GII.4 New Orleans and GII.4 Sydney are very different, which suggests that there may be little cross-immunity between these strains.[23] However, in many countries the emergence of GII.4 Sydney did not coincide with increases in outbreak activity, which suggests that the level of cross-immunity was higher than what was

estimated from *in vitro* studies. If the level of cross-immunity between these strains is low, this raises concerns that norovirus vaccines my need to be reformulated. However, evidence from *in vitro* studies utilizing surrogate neutralization assays and enzyme immunoassays have indicated that some GII.4 strains, such as US 95/96 and Farmington Hills, are broadly cross-reactive with other GII.4 strains,[204,205] suggesting that cross-immunity between strains that are not consecutive. Again, these findings provide further support that future vaccine efficacy studies should quantify vaccine protection against heterotypic exposure with GII.4 noroviruses. Further, long-term global surveillance will be required to track the emergence of GII.4 strains and characterize changes in GII.4 strain distributions over time to better understand how GII.4 evolution may influence vaccine impact.

6.4 Summary

The results of this dissertation contribute a better understanding of the variability in norovirus transmission and susceptibility and cross-immunity to rapidly evolving GII.4 noroviruses. With a better understanding of how norovirus transmission varies across the age range we can leverage that knowledge to target specific age groups for vaccination to maximize population-level benefits and reduce the burden of severe norovirus disease outcomes. Further, estimates of how norovirus transmission varies by outbreak settings and season can help us identify where and when norovirus outbreaks could be more severe, and perhaps consider whether the level and type of control measures should differ by outbreak context. Finally, a better understanding of GII.4 norovirus strain dynamics and the extent of cross-protection between strains provides insight into whether norovirus vaccines will need to be reformulated to keep pace with viral evolution.

7 References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet (London, England)* **392,** 1736–1788 (2018).

2. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Globa. *Lancet (London, England)* **392,** 1859–1922 (2018).

3. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for. *Lancet (London, England)* **392,** 1789–1858 (2018).

4. Ahmed, S. M. *et al.* Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet Infect. Dis.* **14,** 725–730 (2014).

5. Pires, S. M. *et al.* Aetiology-specific estimates of the global and regional incidence and mortality of diarrhoeal diseases commonly transmitted through food. *PLoS One* **10,** 1–17 (2015).

6. Lanata, C. F. *et al.* Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. *PLoS One* **8,** e72788 (2013).

7. Troeger, C. *et al.* Estimates of the global, regional, and national morbidity, mortality, and

aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect. Dis.* **18,** 1211–1228 (2018).

8. Yen, C. *et al.* Rotavirus vaccines. *Hum. Vaccin. Immunother.* **10,** 1436–1448 (2014).

9. Payne, D. C. *et al.* Norovirus and Medically Attended Gastroenteritis in U.S. Children. *N. Engl. J. Med.* **368,** 1121–1130 (2013).

10. Patel, M. M. *et al.* Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerg. Infect. Dis.* **14,** 1224–31 (2008).

11. Hall, A. J. *et al.* Norovirus disease in the united states. *Emerg. Infect. Dis.* **19,** 1198–1205 (2013).

12. Ramani, S., Atmar, R. L. & Estes, M. K. Epidemiology of human noroviruses and updates on vaccine development. *Curr. Opin. Gastroenterol.* **30,** 25–33 (2014).

13. Gastañaduy, P. A., Hall, A. J., Curns, A. T., Parashar, U. D. & Lopman, B. A. Burden of norovirus gastroenteritis in the ambulatory setting - United States, 2001-2009. *J. Infect. Dis.* **207,** 1058–1065 (2013).

14. Lopman, B. A., Hall, A. J., Curns, A. T. & Parashar, U. D. Increasing rates of gastroenteritis hospital discharges in US adults and the contribution of norovirus, 1996-2007. *Clin. Infect. Dis.* **52,** 466–74 (2011).

15. Hall, A. J., Curns, A. T., McDonald, L., Parashar, U. D. & Lopman, B. A. The roles of clostridium difficile and norovirus among gastroenteritis- associated deaths in the United States, 1999-2007. *Clin. Infect. Dis.* **55,** 216–223 (2012).

16. Phillips, G. *et al.* Community incidence of norovirus-associated infectious intestinal disease

in England: improved estimates using viral load for norovirus diagnosis. *Am. J. Epidemiol.* **171,** 1014–1022 (2010).

17. Ahmed, S. M., Lopman, B. A. & Levy, K. A systematic review and meta-analysis of the global seasonality of norovirus. *PLoS One* **8,** 1–7 (2013).

18. Greer, A. L., Drews, S. J. & Fisman, D. N. Why 'winter' vomiting disease? seasonality, hydrology, and norovirus epidemiology in Toronto, Canada. *Ecohealth* **6,** 192–199 (2009).

19. Lopman, B., Armstrong, B., Atchison, C. & Gray, J. J. Host, weather and virological factors drive norovirus epidemiology: Time-series analysis of laboratory surveillance data in England and Wales. *PLoS One* **4,** (2009).

20. Duizer, E. *et al.* Inactivation of caliciviruses. *Appl. Environ. Microbiol.* **70,** 4538–4543 (2004).

21. Doultree, J. C., Druce, J. D., Birch, C. J., Bowden, D. S. & Marshall, J. A. Inactivation of feline calicivirus, a Norwalk virus surrogate. *J. Hosp. Infect.* **41,** 51–57 (1999).

22. Donaldson, E. F., Lindesmith, L. C., Lobue, A. D. & Baric, R. S. Viral shape-shifting: norovirus evasion of the human immune system. *Nat. Rev. Microbiol.* **8,** 231–41 (2010).

23. Debbink, K. *et al.* Emergence of new pandemic GII.4 Sydney norovirus strain correlates with escape from herd immunity. *J. Infect. Dis.* **208,** 1877–1887 (2013).

24. Lindesmith, L. C. *et al.* Conformational occlusion of blockade antibody epitopes, a novel mechanism of GII.4 human norovirus immune evasion. (2018). doi:10.1128/mSphere.00518-17

25. Kroneman, A. *et al.* Proposal for a unified norovirus nomenclature and genotyping. *Arch*

Virol **158,** 2059–2068 (2013).

26. Vinjé, J. Advances in laboratory methods for detection and typing of norovirus. *J. Clin. Microbiol.* **53,** 373–381 (2015).

27. Hoa Tran, T. N., Trainor, E., Nakagomi, T., Cunliffe, N. a & Nakagomi, O. Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: global distribution of genogroups, genotypes and GII.4 variants. *J. Clin. Virol.* **56,** 185–93 (2013).

28. Noel, J. S., Fankhauser, R. L., Ando, T., Monroe, S. S. & Glass, R. I. Identification of a distinct common strain of 'Norwalk-like viruses' having a global distribution. *J. Infect. Dis.* **179,** 1334–1344 (1999).

29. Parra, G. I. *et al.* Static and evolving norovirus genotypes: implications for epidemiology and immunity. *PLOS Pathog.* **13,** e1006136 (2017).

30. Siebenga, J. J. *et al.* Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. *J. Virol.* **81,** 9932–41 (2007).

31. Siebenga, J. J. *et al.* Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. *PLoS Pathog.* **6,** 1–13 (2010).

32. Graaf, M. De *et al.* Emergence of a novel GII.17 norovirus – End of the GII.4 era? *Euro Surveill* **20,** (2015).

33. Lopman, B. *et al.* Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. *Lancet* **363,** 682–688 (2004).

34. Tan, M. *et al.* Mutations within the P2 domain of norovirus capsid affect binding to human

histo-blood group antigens: evidence for a binding pocket. *J. Virol.* **77,** 12562–12571 (2003).

35. Chen, R. *et al.* Inter- and intragenus structural variations in caliciviruses and their functional implications. *J. Virol.* **78,** 6469–6479 (2004).

36. Lindesmith, L. C. *et al.* Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* **5,** e31 (2008).

37. Boon, D. *et al.* Comparative evolution of GII.3 and GII.4 norovirus over a 31-year period. *J. Virol.* **85,** 8656–8666 (2011).

38. Bull, R. A., Eden, J. S., Rawlinson, W. D. & White, P. A. Rapid evolution of pandemic noroviruses of the GII.4 lineage. *PLoS Pathog.* **6,** 1–10 (2010).

39. Bull, R. A. & White, P. A. Mechanisms of GII.4 norovirus evolution. *Trends Microbiol.* **19,** 233–240 (2011).

40. Siebenga, J. J. *et al.* Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001-2007. *J. Infect. Dis.* **200,** 802–12 (2009).

41. Kroneman, A. *et al.* Analysis of integrated virological and epidemiological reports of norovirus outbreaks collected within the Foodborne Viruses in Europe network from 1 July 2001 to 30 June 2006. *J. Clin. Microbiol.* **46,** 2959–65 (2008).

42. Eden, J.-S., Tanaka, M. M., Boni, M. F., Rawlinson, W. D. & White, P. A. Recombination within the pandemic norovirus GII.4 lineage. *J. Virol.* **87,** 6270–82 (2013).

43. Cannon, J. L. *et al.* Herd immunity to GII.4 noroviruses is supported by outbreak patient sera. *J. Virol.* **83,** 5363–5374 (2009).

44. Lindesmith, L. C. *et al.* Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. *PLoS Pathog.* **8,** e1002705 (2012).

45. Ruis, C. *et al.* The emerging GII.P16-GII.4 Sydney 2012 norovirus lineage is circulating worldwide, arose by late-2014 and contains polymerase changes that may increase virus transmission. *PLoS One* **12,** e0179572 (2017).

46. Lindesmith, L. C. *et al.* Antigenic characterization of a novel recombinant GII.P16-GII.4 Sydney norovirus strain with minor sequence variation leading to antibody escape. *J. Infect. Dis.* **217,** 1145–1152 (2018).

47. Glass, R. I., Parashar, U. D. & Estes, M. K. Norovirus gastroenteritis. *N. Engl. J. Med.* **361,** 1776–85 (2009).

48. Geun, W. P., Boston, D. M., Kase, J. A., Sampson, M. N. & Sobsey, M. D. Evaluation of liquid- and fog-based application of sterilox hypochlorous acid solution for surface inactivation of human norovirus. *Appl. Environ. Microbiol.* **73,** 4463–4468 (2007).

49. Atmar, R. L. *et al.* Determination of the 50% human infectious dose for Norwalk virus. *J. Infect. Dis.* **209,** 1016–1022 (2014).

50. Teunis, P. F. *et al.* Norwalk virus: how infectious is it? *J. Med. Virol.* **80,** 1468–1476 (2008).

51. Aoki, Y. *et al.* Duration of norovirus excretion and the longitudinal course of viral load in norovirus-infected elderly patients. **75,** 42–46 (2010).

52. Kirby, A. E., Shi, J., Montes, J., Lichtenstein, M. & Moe, C. L. Disease course and viral shedding in experimental Norwalk virus and Snow Mountain virus infection. *J. Med. Virol.*

86, 2055–2064 (2014).

53. Atmar, R. L. *et al.* Norwalk virus shedding after experimental human infection. *Emerg. Infect. Dis.* **14,** 1553–7 (2008).

54. Kirby, A. E. *et al.* Vomiting as a symptom and transmission risk in norovirus illness: evidence from human challenge studies. *PLoS One* **11,** 1–11 (2016).

55. Teunis, P. F. M. *et al.* Shedding of norovirus in symptomatic and asymptomatic infections. *Epidemiol. Infect.* **143,** 1–8 (2014).

56. Mathijs, E. *et al.* A review of known and hypothetical transmission routes for noroviruses. *Food Environ. Virol.* **4,** 131–52 (2012).

57. Lopman, B. *et al.* Environmental transmission of norovirus gastroenteritis. *Curr. Opin. Virol.* **2,** 96–102 (2012).

58. Thornley, C. N., Emslie, N. A., Sprott, T. W., Greening, G. E. & Rapana, J. P. Recurring norovirus transmission on an airplane. *Clin. Infect. Dis.* **53,** 515–520 (2011).

59. Evans, M. R. *et al.* An outbreak of viral gastroenteritis following environmental contamination at a concert hall. *Epidemiol. Infect.* **129,** 355–360 (2002).

60. Gaythorpe, K. A. M., Trotter, C. L., Lopman, B., Steele, M. & Conlan, A. J. K. Norovirus transmission dynamics: a modelling review. *Epidemiol Infect* **146,** 147–158 (2018).

61. de Wit, M. *et al.* Sensor, a population-based cohort study on gastroenteritidis in the Netherlands: incidence and etiology. *Am J Epidemiol* **154: 666–6,** 666–674 (2001).

62. Lee, R. M. *et al.* Incubation periods of viral gastroenteritis: a systematic review. *BMC Infect. Dis.* **13,** 446 (2013).

63. Devasia, T., Lopman, B., Leon, J. & Handel, A. Association of host, agent and environment characteristics and the duration of incubation and symptomatic periods of norovirus gastroenteritis. *Epidemiol. Infect.* **FirstView,** 2308–2314 (2015).

64. Hall, A. J. *et al.* Epidemiology of foodborne norovirus outbreaks, United States, 2001 - 2008. *Emerg. Infect. Dis.* **18,** 1566–1573 (2012).

65. Atmar, R. L. & Estes, M. K. The epidemiologic and clinical importance of norovirus infection. *Gastroenterol. Clin. North Am.* **35,** 275–90, viii (2006).

66. Wyatt, R. G. *et al.* Comparison of three agents of acute infectious nonbacterial gastroenteritis by cross-challenge in volunteers. *J. Infect. Dis.* **129,** 709–714 (1974).

67. Dedman, D., Laurichesse, H., Caul, E. O. & Wall, P. G. Surveillance of small round structured virus (SRSV) infection in England and Wales, 1990–5. in *Epidemiology and infection.* 139–149 (Cambridge University Press, 1998).

68. Rockx, B. *et al.* Natural history of calicivirus infection: a prospective cohort study. *Clin.* **35,** 246–253 (2002).

69. Kaplan, J. E. *et al.* Epidemiology of Norwalk Gastroenteritis and the Role of Norwalk Virus in Outbreaks of Acute Nonbacterial Gastroenteritis. *Ann. Intern. Med.* **96,** 756 (1982).

70. Lopman, B. A., Reacher, M. H., Vipond, I. B., Sarangi, J. & Brown, D. W. G. Clinical manifestation of norovirus gastroenteritis in health care settings. *Clin. Infect. Dis.* **39,** 318–24 (2004).

71. Barreira, D. M. P. G. *et al.* Viral load and genotypes of noroviruses in symptomatic and asymptomatic children in Southeastern Brazil. *J. Clin. Virol.* **47,** 60–64 (2010).

72. Kawada, J.-I. *et al.* Clinical characteristics of norovirus gastroenteritis among hospitalized children in Japan. *Microbiol Immunol* **56,** 756–759 (2012).

73. de Andrade, J. da S. R. *et al.* Noroviruses associated with outbreaks of acute gastroenteritis in the State of Rio Grande do Sul, Brazil, 2004–2011. *J. Clin. Virol.* **61,** 345–352 (2014).

74. Wu, T.-C., Liu, H.-H., Chen Yann-Jang, Hwang, B.-T. & Yuan, H.-C. Comparison of clinical features of childhood norovirus and rotavirus gastroenteritis in Taiwan. *J Chin Med Assoc* **71,** 566–570 (2008).

75. Ternhag, A., Törner, A., Svensson, A., Ekdahl, K. & Giesecke, J. Short- and long-term effects of bacterial gastrointestinal infections. *Emerg. Infect. Dis.* **14,** 143–8 (2008).

76. Frange, P. *et al.* Prevalence and clinical impact of norovirus fecal shedding in children with inherited immune deficiencies. *J. Infect. Dis.* **206,** 1269–1274 (2012).

77. Munir, N. *et al.* Norovirus infection in immunocompromised children and children with hospital-acquired acute gastroenteritis. *J. Med. Virol.* **86,** 1203–1209 (2014).

78. Wingfield, T. *et al.* Chronic norovirus infection in an HIV-positive patient with persistent diarrhoea: A novel cause. *J. Clin. Virol.* **49,** 219–222 (2010).

79. Mattner, F. *et al.* Risk groups for clinical complications of norovirus infections: an outbreak investigation. *Clin. Microbiol. Infect.* **12,** 69–74 (2006).

80. Hutson, A. M., Atmar, R. L., Marcus, D. M. & Estes, M. K. Norwalk virus-like particle hemagglutination by binding to h histo-blood group antigens. *J. Virol.* **77,** 405–15 (2003).

81. Hutson, A. M. *et al.* Norwalk virus infection associates with secretor status genotyped from sera. *J. Med. Virol.* **77,** 116–120 (2005).

82. Hutson, A. M., Atmar, R. L., Graham, D. Y. & Estes, M. K. Norwalk virus infection and disease is associated with ABO histo–blood group type. *J. Infect. Dis.* **185,** 1335–1337 (2002).

83. Lindesmith, L. *et al.* Human susceptibility and resistance to Norwalk virus infection. *Nat. Med.* **9,** (2003).

84. Parrino, T. A., Schreiber, D. S., Trier, J. S., Kapikian, A. Z. & Blacklow, N. R. Clinical immunity in acute gastroenteritis caused by norwalk agent. *N. Engl. J. Med.* (1977).

85. Graham, D. Y. *et al.* Norwalk virus infection of volunteers: new insights based on improved assays. *J. Infect. Dis.* **170,** 34–43 (1994).

86. Marionneau, S. *et al.* ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* **83,** 565–573 (2001).

87. Marionneau, S., Airaud, F., Bovin, N. V., Pendu, J. Le & Ruvoën-Clouet, N. Influence of the combined ABO, FUT2, and FUT3 polymorphism on susceptibility to Norwalk virus attachment. *J. Infect. Dis.* **192,** 1071–1077 (2005).

88. Severine Marionneau *et al.* Norwalk virus binds to histo-blood group antigens present on gastroduodenal epithelial cells of secretor individuals. *Gastroenterology* **122,** 1967–1977 (2002).

89. Johnson, P. C., Mathewson, J. J., DuPont, H. L. & Greenberg, H. B. Multiple-challenge study of host susceptibility to Norwalk gastroenteritis in US adults. *J. Infect. Dis.* **161,** 18–21 (1990).

90. Lindesmith, L. *et al.* Cellular and humoral immunity following Snow Mountain virus

challenge. *J. Virol.* **79,** 2900 (2005).

91. Simmons, K., Gambhir, M., Leon, J. & Lopman, B. Duration of immunity to norovirus gastroenteritis. *Emerg. Infect. Dis.* **19,** 1260–7 (2013).

92. Lindesmith, L. C. *et al.* Heterotypic humoral and cellular immune responses following Norwalk virus infection. *J. Virol.* **84,** 1800–1815 (2010).

93. Rockx, B., Baric, R. S., Grijs, I. De, Duizer, E. & Koopmans, M. P. Characterization of the homo-and heterotypic immune responses after natural norovirus infection. *J. Med. Virol.* **77,** 439–446 (2005).

94. Lindesmith, L. C., Donaldson, E. F. & Baric, R. S. Norovirus GII.4 strain antigenic variation. *J. Virol.* **85,** 231–242 (2011).

95. LoBue, A. D. *et al.* Multivalent norovirus vaccines induce strong mucosal and systemic blocking antibodies against multiple strains. *Vaccine* **24,** 5220–5234 (2006).

96. Reeck, A. *et al.* Serological correlate of protection against norovirus-induced gastroenteritis. *J. Infect. Dis.* **202,** 1212–1218 (2010).

97. Treanor, J. J., Paul Madore, H. & Dolin, R. Development of an enzyme immunoassay for the Hawaii agent of viral gastroenteritis. *J. Virol. Methods* **22,** 207–214 (1988).

98. Yen, C. *et al.* Impact of an emergent norovirus variant in 2009 on norovirus outbreak activity in the United States. *Clin. Infect. Dis.* **53,** 568–571 (2011).

99. Leshem, E. *et al.* Effects and clinical significance of GII.4 Sydney norovirus, United States, 2012-2013. *Emerg. Infect. Dis.* **19,** 1231–1238 (2013).

100. Vinjé, J., Altena, S. A. & Koopmans, M. P. G. The incidence and genetic variability of small

round-structured viruses in outbreaks of gastroenteritis in The Netherlands. *J. Infect. Dis.* **176,** 1374–1378 (1997).

101. Hall, A. J. *et al.* Acute gastroenteritis surveillance through the National Outbreak Reporting System, United States. *Emerg. Infect. Dis.* **19,** 1305–9 (2013).

102. National Outbreak Reporting System (NORS). *Centers for Disease Control and Prevention* Available at: https://www.cdc.gov/nors/data.html. (Accessed: 22nd October 2017)

103. Burke, R. M. *et al.* The norovirus epidemiologic triad: predictors of severe outcomes in US norovirus outbreaks, 2009-2016. *J. Infect. Dis.* (2018). doi:10.1093/infdis/jiy569

104. Tamminen, K., Lappalainen, S., Huhti, L., Vesikari, T. & Blazevic, V. Trivalent combination vaccine induces broad heterologous immune responses to norovirus and rotavirus in mice. *PLoS One* **8,** 1–14 (2013).

105. Tamminen, K. *et al.* A comparison of immunogenicity of norovirus GII-4 virus-like particles and P-particles. *Immunlogy* **135,** 89–99 (2012).

106. Tan, M. & Jiang, X. Norovirus P particle: a subviral nanoparticle for vaccine development against norovirus, rotavirus and influenza virus. *Nanomedicine* **7,** 889–897 (2012).

107. Velasquez, L. S. *et al.* Intranasal delivery of Norwalk virus-like particles formulated in an in situ gelling, dry powder vaccine. *Vaccine* **29,** 5221–31 (2011).

108. Bernstein, D. I. *et al.* Norovirus vaccine against experimental human GII.4 virus illness: A challenge study in healthy adults. *J. Infect. Dis.* **211,** 870–878 (2015).

109. Atmar, R. L. *et al.* Norovirus vaccine against experimental human Norwalk Virus illness. *N. Engl. J. Med.* **365,** 2178–87 (2011).

110. Richardson, C., Bargatze, R. F., Goodwin, R. & Mendelman, P. M. Norovirus virus-like particle vaccines for the prevention of acute gastroenteritis. *Expert Rev. Vaccines* **12,** 155–67 (2013).

111. Treanor, J. J. *et al.* A novel intramuscular bivalent norovirus virus-like particle vaccine candidate--reactogenicity, safety, and immunogenicity in a phase 1 trial in healthy adults. *J. Infect. Dis.* **210,** 1763–1771 (2014).

112. El-Kamary, S. S. *et al.* Adjuvanted intranasal Norwalk virus-like particle vaccine elicits antibodies and antibody-secreting cells that express homing receptors for mucosal and peripheral lymphoid tissues. *J. Infect. Dis.* **202,** 1649–58 (2010).

113. Ettayebi, K. *et al.* Replication of human noroviruses in stem cell–derived human enteroids. *Science* **353,** 1387–1393 (2016).

114. Ball, J. M. *et al.* Recombinant Norwalk virus – like particles given orally to volunteers: phase I study. *Gastroenterology* **117,** 40–48 (1999).

115. Vinjé, J. A norovirus vaccine on the horizon? *J. Infect. Dis.* **202,** 1623–5 (2010).

116. Parra, G. I. *et al.* Immunogenicity and specificity of norovirus Consensus GII.4 virus-like particles in monovalent and bivalent vaccine formulations. *Vaccine* **30,** 3580–6 (2012).

117. Sundararajan, A. *et al.* Robust mucosal-homing antibody-secreting B cell responses induced by intramuscular administration of adjuvanted bivalent human norovirus-like particle vaccine. *Vaccine* **33,** 568–576 (2015).

118. Lindesmith, L. C. *et al.* Broad blockade antibody responses in human volunteers after immunization with a multivalent norovirus VLP candidate vaccine: immunological

analyses from a phase I clinical trial. *PLoS Med.* **12,** e1001807 (2015).

119. Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in Humans. *Science* **314,** 1898–1903 (2006).

120. Hall, A. J., Wikswo, M. E., Pringle, K., Gould, L. H. & Parashar, U. D. Vital signs: foodborne norovirus outbreaks—United States, 2009–2012. *MMWR Morb Mortal Wkly Rep* **63,** 491–495 (2014).

121. Sukhrie, F. H. *et al.* Nosocomial transmission of norovirus is mainly caused by symptomatic cases. *Clin. Infect. Dis.* **54,** 931–7 (2012).

122. Ajami, N. J. *et al.* Seroepidemiology of norovirus-associated travelers' diarrhea. *J. Travel Med.* **21,** 6–11 (2014).

123. Riddle, M. S. *et al.* Epidemic infectious gastrointestinal illness aboard U.S. Navy ships deployed to the Middle East during peacetime operations – 2000–2001. *BMC Gastroenterol.* **6,** 9 (2006).

124. Rha, B., Lopman, B. A., Alcala, A. N., Riddle, M. S. & Porter, C. K. Incidence of norovirus-associated medical encounters among active duty United States military personnel and their dependents. *PLoS One* **11,** e0148505 (2016).

125. Halloran, M. E. & Struchiner, C. J. Study Designs for Dependent Happenings. *Epidemiology* **2,** 331–338

126. Ross, R. An application of the theory of probabilities to the study of a priori pathometry. Part I. *Proceedings of the Royal Society of London.* **92,** 204–230

127. Nokes, D. J. & Anderson, R. M. The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programmes. *Epidemiol. Infect.* **101,** 1–20 (1988).

128. Anderson, R. M. & May, R. M. Vaccination and herd immunity to infectious diseases. *Nature* **318,** 323–329 (1985).

129. Viboud, C. *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312,** 447–51 (2006).

130. Dushoff, J., Plotkin, J. B., Levin, S. A. & Earn, D. J. D. Dynamical resonance can account for seasonality of influenza epidemics. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16915–6 (2004).

131. Fraser, C. *et al.* Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* **324,** 1557–61 (2009).

132. Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. doi:10.1038/nature04795

133. Cauchemez, S., Valleron, A.-J., Boëlle, P.-Y., Flahault, A. & Ferguson, N. M. Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* **452,** 750–754 (2008).

134. Bootsma, M. C. J. & Ferguson, N. M. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 7588–93 (2007).

135. Towers, S. & Feng, Z. Pandemic H1N1 influenza: predicting the course of a pandemic and assessing the efficacy of the planned vaccination programme in the United States. *Euro Surveill* **14,** (2009).

136. O'Dea, E. B., Pepin, K. M., Lopman, B. A. & Wilke, C. O. Fitting outbreak models to data from many small norovirus outbreaks. *Epidemics* **6,** 18–29 (2014).

137. Vanderpas, J., Louis, J., Reynders, M., Mascart, G. & Vandenberg, O. Mathematical model for the control of nosocomial norovirus. *J. Hosp. Infect.* **71,** 214–222 (2009).

138. Lopman, B., Simmons, K., Gambhir, M., Vinjé, J. & Parashar, U. Epidemiologic implications of asymptomatic reinfection: a mathematical modeling study of norovirus. *Am. J. Epidemiol.* **179,** 507–512 (2014).

139. Papafragkou, E. *et al.* Challenges of culturing human norovirus in three-dimensional organoid intestinal cell culture models. *PLoS One* **8,** e63485 (2013).

140. Duizer, E. *et al.* Laboratory efforts to cultivate noroviruses. *J. Gen. Virol.* **85,** 79–87 (2004).

141. Herbst-Kralovetz, M. M. *et al.* Lack of norovirus replication and histo-blood group antigen expression in 3-dimensional intestinal epithelial cells. *Emerg. Infect. Dis.* **19,** 431–438 (2013).

142. Matsuyama, R., Miura, F. & Nishiura, H. The transmissibility of noroviruses: Statistical modeling of outbreak events with known route of transmission in Japan. *PLoS One* **12,** e0173996 (2017).

143. Andreasen, V., Lin, J. & Levin, S. A. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *J. Math. Biol.* **35,** 825–842 (1997).

144. Gog, J. R. & Swinton, J. A status-based approach to multiple strain dynamics. *Math. Biol.* **44,** 169–184 (2002).

145. Gog, J. R. & Grenfell, B. T. Dynamics and selection of many-strain pathogens. *Proc. Natl.*

Acad. Sci. U. S. A. **99,** 17209–14 (2002).

146. Kryazhimskiy, S., Dieckmann, U., Levin, S. A. & Dushoff, J. On state-space reduction in multi-strain pathogen models, with an application to antigenic drift in influenza A. *PLoS Comput. Biol.* **3,** e159 (2007).

147. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. *Evolution and Ecology of Influenza A Viruses*. *Microbiological Reviews* **56,** (1992).

148. Ballesteros, S., Vergu, E. & Cazelles, B. Influenza A gradual and epochal evolution: Insights from simple models. *PLoS One* **4,** e7426 (2009).

149. Bartsch, S. M., Lopman, B. A., Hall, A. J., Parashar, U. D. & Lee, B. Y. The potential economic value of a human norovirus vaccine for the United States. *Vaccine* **30,** 7097–104 (2012).

150. Lopman, B. A., Steele, D., Kirkwood, C. D. & Parashar, U. D. The vast and varied blobal burden of norovirus: prospects for prevention and control. *PLoS Med.* **13,** 1–12 (2016).

151. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5,** e74 (2008).

152. Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals*. (Princeton University Press, 2007).

153. Diekmann, O. & Heesterbeeck, J. A. P. *Mathematical Epidemiology of Infectious Diseases*. (Wiley, 2000).

154. Heffernan, J. M., Smith, R. J. & Wahl, L. M. Perspectives on the basic reproductive ratio. *J. R. Soc. Interface* **2,** 281–293 (2005).

155. LoBue, A. D., Lindesmith, L. C. & Baric, R. S. Identification of cross-reactive norovirus CD4+ T cell epitopes. *J. Virol.* **84,** 8530–8538 (2010).

156. Centers for Disease Control and Prevention (CDC). Updated norovirus outbreak management and disease prevention guidelines. *Morb. Mortal. Wkly. Report, Recomm. Reports* **60,** 1–18 (2011).

157. Gray, J. J., Jiang, X., Morgan-Capner, P., Desselberger, U. & Estes, M. K. Prevalence of antibodies to Norwalk virus in England: detection by enzyme-linked immunosorbent assay using baculovirus-expressed Norwalk virus capsid antigen. *J. Clin. Microbiol.* **31,** 1022–1025 (1993).

158. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

159. White, L. J. *et al.* Understanding the transmission dynamics of respiratory syncytial virus using multiple time series and nested models. *Math. Biosci.* **209,** 222–239 (2007).

160. Smith, P. G., Rodrigues, L. C. & Fine, P. E. M. Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *Int. J. Epidemiol.* **13,** (1984).

161. Centers for Disease Control and Prevention (CDC). National, state, and local area vaccination coverage among children aged 19-35 months--United States, 2011. *Morb. Mortal. Wkly. Rep.* **61,** 689–96 (2012).

162. Lu, P. *et al.* Surveillance of influenza vaccination coverage--United States, 2007-08 through 2011-12 influenza seasons. *MMWR. Surveill. Summ.* **62,** 1–28 (2013).

163. Lexau, C. A. *et al.* Changing epidemiology of invasive pneumococcal disease among older adults in the era of pediatric pneumococcal conjugate vaccine. *Jama* **294,** 2043–2051 (2005).

164. Tate, J. E. *et al.* Uptake, impact, and effectiveness of rotavirus vaccination in the United States: review of the first 3 years of postlicensure data. *Pediatr. Infect. Dis. J.* **30,** S56–S60 (2011).

165. Arinaminpathy, N. *et al.* Impact of cross-protective vaccines on epidemiological and evolutionary dynamics of influenza. *Proc. Natl. Acad. Sci.* **109,** 3173–3177 (2012).

166. Shioda, K., Kambhampati, A., Hall, A. J. & Lopman, B. A. Global age distribution of pediatric norovirus cases. *Vaccine* 3–6 (2015). doi:10.1016/j.vaccine.2015.05.051

167. Soetaert, K., Petzoldt, T. & Setzer, R. W. Package deSolve: solving initial value differential equations in R. *J. Stat. Softw.* **33,** 1–25 (2010).

168. Johnson, S. G. The NLopt nonlinear-optimization package, http://ab-initio.mit.edu/nlopt.

169. Carnell, R. Package 'lhs.' Available at: http://cran.r-project.org/web/packages/lhs/lhs.pdf.

170. Pujol, G., Iooss, B. & Janon, A. Package 'sensitivity'. (2015). Available at: http://cran.r-project.org/web/packages/sensitivity/sensitivity.pdf.

171. Eames, K. T. D., Tilston, N. L., Brooks-Pollock, E. & Edmunds, W. J. Measured dynamic social contact patterns explain the spread of H1N1v influenza. *PLoS Comput. Biol.* **8,** 1–8 (2012).

172. Pitzer, V. E. *et al.* Direct and indirect effects of rotavirus vaccination: Comparing predictions from transmission dynamic models. *PLoS One* **7,** (2012).

173. Wikswo, M. E. *et al.* Outbreaks of acute gastroenteritis transmitted by person-to-person contact, environmental contamination, and unknown modes of transmission--United States, 2009–2013. **64,** (2015).

174. Rothman, K., Greenland, S. & Lash, T. *Modern epidemiology*. (2008).

175. Kermack, W. D. & McKendrick, A. G. Mathematical models of the spread of infection. *Proc. R. Soc. A* **115,** 700–721 (1927).

176. Becker, N. G. *Analysis of Infectious Disease Data*. (Chapman and Hall, 1989).

177. Ajami, N. J. *et al.* Antibody responses to norovirus genogroup GI.1 and GII.4 proteases in volunteers administered Norwalk virus. *Clin. Vaccine Immunol.* **19,** 1980–1983 (2012).

178. Frenck, R. *et al.* Predicting susceptibility to norovirus GII.4 by use of a challenge model involving humans. *J. Infect. Dis.* **206,** 1386–93 (2012).

179. Leon, J. S. *et al.* Randomized, double-blinded clinical trial for human norovirus inactivation in oysters by high hydrostatic pressure processing. *Appl. Environ. Microbiol.* **77,** 5476–5482 (2011).

180. Bartsch, S. M., Huang, S. S., Wong, K. F., Avery, T. R. & Lee, B. Y. The spread and control of norovirus outbreaks among hospitals in a region: a simulation model. *Open Forum Infect. Dis.* **1,** ofu030-ofu030 (2014).

181. Steele, M. K. *et al.* Targeting pediatric versus elderly populations for norovirus vaccines: a model-based analysis of mass vaccination options. *Epidemics* **17,** 42–49 (2016).

182. Milbrath, M. O., Spicknall, I. H., Zelner, J. L., Moe, C. L. & Eisenberg, J. N. S. Heterogeneity in norovirus shedding duration affects community risk. *Epidemiol. Infect.*

141, 1572–84 (2013).

183. Lawrence, L., Kerrod, E., Gani, R. & Leach, S. London: Food Standards Agency. Microbiological risk assessment for norovirus infection contribution to the overall burden afforded by foodborne infections. *London Food Stand. Agency.* (2004).

184. Shah, M. P. *et al.* Near real-time surveillance of U.S. norovirus outbreaks by the Norovirus Sentinel Testing and Tracking Network — United States, August 2009–July 2015. *MMWR. Morb. Mortal. Wkly. Rep.* **66,** 185 (2017).

185. Reporting and Surveillance for Norovirus: NoroSTAT. *Centers for Disease Control and Prevention* Available at: https://www.cdc.gov/norovirus/reporting/noroSTAT/index.html.

186. Becker, N. Martingale methods for the analysis of epidemic data. *Stat. Methods Med. Res.* **2,** 93–112 (1993).

187. Sukhrie, F. H., Siebenga, J. J., Beersma, M. F. C. & Koopmans, M. Chronic shedders as reservoir for nosocomial transmission of norovirus. *J. Clin. Microbiol.* **48,** 4303–5 (2010).

188. Dolin, R. *et al.* Biological properties of Norwalk agent of acute infectious nonbacterial gastroenteritis. *Exp. Biol. Med.* **140,** 578–583 (1972).

189. Seitz, S. R. *et al.* Norovirus infectivity in humans and persistence in water. *Appl. Environ. Microbiol.* **77,** 6884–6888 (2011).

190. Debbink, K., Donaldson, E. F., Lindesmith, L. C. & Baric, R. S. Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. *J. Virol.* **86,** 1214–1226 (2012).

191. Sharma, S. *et al.* Human sera collected between 1979 and 2010 possess blocking-antibody

titers to pandemic GII.4 noroviruses isolated over three decades. *J. Virol.* **91,** (2017).

192. Mallory, M. L., Lindesmith, L. C., Graham, R. L. & Baric, R. S. GII.4 human norovirus: surveying the antigenic landscape. *Viruses* **11,** (2019).

193. Gaythorpe, K. A. M., Trotter, C. L. & Conlan, A. J. K. Modelling norovirus transmission and vaccination. *Vaccine* **36,** 5565–5571 (2018).

194. Vega, E. *et al.* Novel surveillance network for norovirus gastroenteritis outbreaks, United States. *Emerg. Infect. Dis.* **17,** 1389–1395 (2011).

195. Public Health England. Norovirus and rotavirus: summary of surveillance 2018 to 2019. Available at: https://www.gov.uk/government/statistics/norovirus-and-rotavirus-summary-of-surveillance-2018-to-2019.

196. Hasing, M. E. *et al.* Emergence of a new norovirus GII.4 variant and changes in the historical biennial pattern of norovirus outbreak activity in Alberta, Canada, from 2008 to 2013. (2013). doi:10.1128/JCM.00663-13

197. Hasing, M. E. *et al.* Changes in norovirus genotype diversity in gastroenteritis outbreaks in Alberta, Canada: 2012–2018. *BMC Infect. Dis.* **19,** 177 (2019).

198. Pang, X. L., Preiksaitis, J. K., Wong, S., Li, V. & Lee, B. E. Influence of novel norovirus GII.4 variants on gastroenteritis outbreak dynamics in Alberta and the Northern Territories, Canada between 2000 and 2008. *PLoS One* **5,** e11599 (2010).

199. Prag, C., Prag, M. & Fredlund, H. Proton pump inhibitors as a risk factor for norovirus infection. *Epidemiol. Infect.* **145,** 1617–1623 (2017).

200. Kucharski, A. J., Andreasen, V. & Gog, J. R. Capturing the dynamics of pathogens with

many strains. *J. Math. Biol.* **72,** 1–24 (2016).

201. Wikramaratna, P. S. *et al.* Five challenges in modelling interacting strain dynamics. *Epidemics* **10,** 31–34 (2015).

202. Hasing, M. E., Hazes, B., Lee, B. E., Preiksaitis, J. K. & Pang, X. L. Detection and analysis of recombination in GII.4 norovirus strains causing gastroenteritis outbreaks in Alberta. *Infect. Genet. Evol.* **27,** 181–192 (2014).

203. Vital Statistics (Births and Deaths) - Alberta, Census Divisions and Economic Regions. *Alberta Governement, Treasury Board and Finance.* Available at: https://open.alberta.ca/dataset/vital-statistics-births-and-deaths-alberta-census-divisions-economic-regions.

204. Tamminen, K. *et al.* Immunological cross-reactivity of an ancestral and the most recent pandemic norovirus GII.4 variant. *Viruses* **11,** 91 (2019).

205. Lindesmith, L. C. *et al.* Monoclonal antibody-based antigenic mapping of norovirus GII.4-2002. *J. Virol.* **86,** 873–883 (2012).

206. Zakikhany, K., Allen, D. J., Brown, D. & Iturriza-Gómara, M. Molecular evolution of GII-4 norovirus strains. *PLoS One* **7,** e41625 (2012).

207. Pitzer, V. E. *et al.* Influence of birth rates and transmission rates on the global seasonality of rotavirus incidence. *J. R. Soc. Interface* **8,** 1584–1593 (2011).

208. Takeda Pharmaceuticals. Safety and immunogenicity of norovirus GI.1/GII.4 bivalent virus-like particle vaccine in an elderly population. *NIH. U.S. National Library of Medicine* (2018). Available at:

https://clinicaltrials.gov/ct2/show/NCT02661490?term=Takeda%2C+Vaccine&age=2&rank=4. (Accessed: 1st June 2019)

209. Takeda Pharmaceuticals. Safety and immunogenicity of norovirus GI.1/GII.4 bivalent virus-like oarticle (VLP) vaccine in children. *NIH. U.S. National Library of Medicine* (2019). Available at: https://clinicaltrials.gov/ct2/show/results/NCT02153112?term=Takeda%2C+Vaccine&age=0&rank=9. (Accessed: 1st June 2019)

210. Grohskopf, L. A. *et al.* Prevention and control of seasonal influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices-United States, 2018-19 influenza season. *MMWR. Recomm. reports Morb. Mortal. Wkly. report. Recomm. reports* **67,** 1–20 (2018).