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:

_____ _____

Maria Nelson Date

Estimating Norovirus Seroprevalence in the US Population Using Finite Mixture Modeling on NHANES Stored Biologic Specimens

By

Maria Nelson

Master of Public Health

Epidemiology

Dr. Amy Kirby, PhD

Faculty Thesis Advisor

Estimating Norovirus Seroprevalence in the US Population Using Finite Mixture Modeling on NHANES Stored Biologic Specimens

By

Maria Nelson

B.S., University of Minnesota, 2013

Faculty Thesis Advisor: Dr. Amy Kirby, PhD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology

2015

Abstract

Estimating Norovirus Seroprevalence in the US Population Using Finite Mixture Modeling on NHANES Stored Biologic Specimens

By Maria Nelson

Background and Purpose- Norovirus is the most common form of viral gastroenteritis in the world. Previous studies have estimated that 23 million cases of norovirus occur in the United States every year.

Methods- Quanitative ELISA tests were run on 2,168 serum samples from 16-49 year olds in the 1999-2000 and 2003-2004 National Health and Nutrition Examination Survey (NHANES). Finite mixture modeling was used to estimate the seroprevalence of norovirus for GI.4, GII.3, and GII.4 US-95/96 strains of norovirus. Seroprevalences were modeled by norovirus strain, age group, and NHANES collection cycle.

Results- Predicted average seroprevalences of three strains of norovirus in the United Stated ranged between 55-60%. The seroprevalence of all three strains decreased significantly from the 1999-2000 NHANES collection to the 2003-2004 collection. GII.4 US-95/96 dropped from 61% to 54% between the two years (χ^2 $p<0.002$). It is hypothesized this decrease is partly because of the emergence of the Farmington Hills strain during this time frame. The most burdened age groups varied by strain, possibly due to differences in transmission routes.

Conclusions-Finite mixture modeling is a beneficial tool to estimate seroprevalence while decreasing sources of bias found in other serological studies.

Estimating Norovirus Seroprevalence in the US Population Using Finite Mixture Modeling on NHANES Stored Biologic Specimens

By

Maria Nelson

B.S., University of Minnesota, 2013

Faculty Thesis Advisor: Dr, Amy Kirby, PhD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology

2015

Acknowledgements:

First and foremost, I would like to express my deepest gratitude to Peter Teunis for trusting me with his model. He patiently handled my embarrassingly rudimentary R knowledge and guided me through my analysis.

I would also like to thank my faculty adviser Dr. Amy Kirby for mentoring me through this exciting project. She gave me freedom to work at my own pace, but also kept me on track. I'm grateful for her kindness and the way she broke the project down into manageable pieces for me to tackle. She was always encouraging, but had beneficial feedback.

Lastly, I'm grateful for my family and friends, especially my fiancé Westin, for allowing me to drop off the face of the world for days at a time while I lost myself in the world of norovirus.

Table of Contents

Literature Review

Norovirus structure and characteristics

Norovirus was previously called "Norwalk-like" virus after the city of Norwalk, Ohio, where the 1968 outbreak took place. The virus was not isolated and identified until 1972 (46, 47). Norwalk virus later became one species in the genus Norovirus, which is one of four genera in the Caliciviridae family (11). Sapovirus, Lagovirus, and Vesivirus are also part of the Calciviridae family (74). Of these genera, the Norovirus genus is the only one known to contain species pathogenic to humans (18). Sapovirus can infect humans, but it is unknown whether it causes disease.

Norovirus is a small, non-enveloped virus 27-38nm in diameter (57). It has a single-stranded positive sense RNA genome that is approximately 7.5kb long (11). This genome contains three open reading frames (ORFs). ORF1 encodes a non-structural protein that is cleaved into six proteins, including an RNA-dependent RNA polymerase and an NTPase (14). ORF2 encodes the major capsid protein (VP1) and ORF3 encodes the minor capsid protein (VP2) (11). The minor capsid protein is further divided into two domains—the shell (S) and protruding (P) domains. The P domain is comprised of the P1 and P2 sub-domains (11). P1 acts as a flexible hinge between the rigid S and protruding P2 domains (14). Motifs in the P2 domain bind to host cells, which contributes to the virus's antigenicity (72, 70).

The norovirus genus houses a wide array of diversity, containing six genogroups (GI-GVI) (69). GI, GII, and GIV norovirus genogroups are the only known to infect humans, while GIII and GV mainly infect cows and mice, respectively (11). It has been suggested that a zoonotic reservoir for human noroviruses could exist, but this has not yet been observed (71,73).The human genogroups can be further divided into at least 32 genotypes (11). A few exceptions to the general rules exist. In contrast to the rest of the GII genogroup, GII. 11 strains

infect pigs, but not humans. Also, GIV.1 viruses infect humans, but GIV.2 is a predominantly feline cluster (66).

It is hypothesized that each norovirus genogroup may be its own species or serotype due to the high capsid sequence variation. There is 60% variation between the genogroups and 57% variation amongst the human genogroups alone. This is much higher than that of other single-stranded RNA viruses, such as polio, which have closer to 20% variation in nucleotides (81).Unfortunately, since no neutralization test is available to determine species, conclusions must be made based on the association between phylogenetic and antigenic characterizations of the virus (11).

GI and GII are the two largest and most diverse genogroups (11). GI norovirus strains are thought to be more stable in water, which may contribute to the fact that they are the strain associated most often with waterborne outbreaks of human norovirus outbreaks (13, 33, 64). They can also contribute to foodborne outbreaks, but that is less common (40).North America is experiencing an increase in GI cases, though. In the United States, GI.6 incidence is increasing (62). Canada is having a more pronounced increase in GI incidence. In Alberta, Canada, GI activity climbed from 7.8% in the early 2000's to 37% in 2012 (75). Of the GI strains reported to the Medical Officer of Health in Canada, GI.6 and GI.7 were most predominant.

The majority of human cases of norovirus infection are caused by the GII strain, which is most frequently the cause of foodborne and person-to-person outbreaks (13). In fact, 55-85% of gastroenteritis cases across the world can be attributed to GII.4 noroviruses (58). Over the last three decades, global epidemics were caused by seven different GII.4 variants (3). GII. 4 is the most active norovirus cluster and is reported in one third of all outbreaks caused by a single genogroup and cluster in US cases (11, 13). Data from the Food-Borne Viruses in Europe

network reveal that from Jan 1999-Dec. 2004, GII.4 caused 52% of all human outbreaks of the virus. Other variants had much less of an effect: GII.b was implicated in 13% and GII.7 6% of outbreaks (28). Every two or three years, a new GII.4 variant emerges to displace the previous dominant variant (3). A 2000 waterborne outbreak in France was caused by GII.b. This was the first reported instance of this variant, which quickly spread to multi-pathogen and oyster-associated outbreaks (28).

Studying human noroviruses is hindered because of their resistance to growing in cell cultures, as well as the limited number of viable animal models (48, 74) Attempted animal models include non-human primates, pigs, calves, and mice, but none of these species adequately mimic the human response to norovirus (43). Chimp models show promise, but NIH has banned their use in medical research. Because of this, using a gnotobiotic pig as an immunological model may provide a better option. Their gut resembles that of humans, and the virus replicates in a similar way in both species (43).

In order to circumvent these challenges, Jiang et al. created virus like particles (VLPs), enclosed empty capsid particles (84). These particles are morphologically similar to a norovirus virion on the exterior, but lack genomic RNA and are unable to replicate (38). To create VLPs, ORF2 was expressed in a recombinant virus vector, either baculovirus or Venezuelan equine encephalitis virus (84). In this system, capsids comprised of 180 copies of the major capsid protein VP1 are formed (49). Since the initial expression in baculovirus, VLPs have also been created in other systems such as vesicular stomatitis virus, and *E. coli*, as well as plant, insect, and mammalian cells (15). It is unknown whether the VLPs created from various systems are biologically different, so researchers choose whichever system is most convenient for their laboratory and skill (38).

Similar to VLPs are virus replicon particles (VRPs), which are structures that can participate in one round of replication with the aid of replication constructs (38). These produce an even greater yield of protein than VLPs and can aid research until a more suitable research model is found for norovirus.

Clinical Attributes and Epidemiology

Once norovirus infects a person, it incubates for 12 to 48 hours. Symptoms include nausea, vomiting, diarrhea, and cramps. Some people refer to it as the "stomach flu" due to the occasional low fever and general body aches that may come with infection. Rarely, chronic diarrhea in immunocompromised patients also occurs (12). This is thought to contribute to viral evolution. Approximately 30% of norovirus infections are asymptomatic, but can still transmit the disease (50,57). Typically, symptoms self-limit after 1-3 days, but may linger up to a week, particularly in children, the elderly, and immunocompromised patients (63).

Not everyone is at the same risk of contracting norovirus; a genetic component plays a significant role. Experimental challenge studies showed that a group of individuals were not susceptible to norovirus infections, despite repeat infection (90, 91). A breakthrough occurred when, using samples from a 2002 norovirus outbreak in a Swedish hospital, Thorven et al. found that susceptibility to norovirus was linked to the expression of histo-blood group antigen (HBGA) in mucosal linings and their secretions, such as saliva (6). This "secretor status" is determined by the gene FUT2, which encodes a fucosyltransferase that adds a necessary component to a HBGA precursor molecule. In numerous recorded outbreaks, not a single non-secretor contracted norovirus (5, 6). Secretors, whether homozygous (SeSe) or heterozygous (Sese), have at least one functional copy of FUT2, and are thus able to express A and B blood

group antigens. This gene also gives them the ability to express H-type 1 and Lewis b (Le^b) antigens. It is thought that HBGAs, a group of carbohydrates found on the surface of epithelial cells, are the receptor for norovirus entry into host cells. When examining stool specimens from children with acute diarrhea in Xi'an, China, Liu et al. found that secretor homozygous and heterozygous children were significantly more susceptible to GII.3 and GII.4 infections (100). Norovirus susceptibility can be strongly associated with secretor status as well as ABO type. Individuals expressing type B antigen have a protective effect against GI strains, while those with type O are drastically more at risk (10, 61).

However, this is not applicable to all strains and situations: a non-secretor was infected with a GII.4 strain in a challenge study (8). In a waterborne outbreak among school children in England, blood type B was significantly associated with protection against norovirus infection, but secretor status did not have a similar statistically significant correlation (10). Some strains GII.4 strains have been found to infect completely independent of secretor, ABO type, or Lewis type (7, 9, 26) Different strains of the virus bind to different patterns of antigens (4).

Norovirus is easily transmissible, requiring as few as 10 viral particles for an infectious dose (ID_{50}=18 viruses) (51). Based on this estimation, a single gram of feces contains approximately 5 billion infectious doses (51). Shedding of the virus peaks 2-5 days post-infection, and typically persists for 4 weeks (17, 50). However, shedding can be detected as early as 18 hours after infection and lasts for up to 8 weeks for typical patients; those with chronic diarrhea can shed up to a year (17). In children, shedding has been observed for more than 100 days after infection (87). Shedding can occur before symptoms develop (17). Symptomatic individuals have peak titer levels significantly higher than asymptomatic individuals (88). Both symptomatic and asymptomatic people shed the virus at approximately the same time, but

asymptomatic people might do so at lower titers (17) However, a challenge study done by Kirby et al. showed no difference between shedding titers (88).

Transmission occurs through three routes: foodborne, waterborne, and person-to-person. Globally, most norovirus cases are foodborne in nature (54%) and occur in foodservice settings (35%), and the fewest number of outbreaks had environmental transmission (9%) (13). Although norovirus outbreaks can happen throughout the year, there is a seasonal pattern which peaks during the winter (79). Approximately half of all norovirus deaths, emergency department, and outpatient visits in the US occur during December-February (19,20)

Most people will acquire the virus many times over their lifetime; Hall et al. estimates that the average person will contract it 5 times (79). This is partly due to the virus's high variability. Studies suggest that noroviruses may evolve faster than other RNA viruses, creating new strains relatively quickly (11). The host immune system puts pressure on the hypervariable region of the P2 domain, encouraging evolution (80). This fast evolution contributes to its prevalence and adaptability (59). Immunity to the virus is also relatively short, ranging from as short as 6 months up to 5-6 years (16, 88). All of these factors contribute to the risk for multiple norovirus infections over a person's lifetime.

New variants of norovirus genotypes can emerge when genetic drift of the VP1 gene occurs, most often in GII.4 strains (102). Genetic drift is partly responsible for the emergence of a new dominant GII.4 strain every 2-3 years (3). Other contributors are greater person-to-person transmission amongst GII.4 strains, recombination, and the ability to escape from herd immunity (80).

Because of the virus's high infectivity and multiple transmission routes, it can be challenging to control outbreaks of norovirus. Proper hand hygiene may be the single most effective way to prevent and control norovirus (34, 52). Washing hands with soap and water for 20 seconds has been shown to reduce the amount of norovirus RNA by 0.7-1.2log 10 in an RT-PCR assay. Physical removal is an important part of hand washing; vigorous rubbing reduces the viral titer more than not rubbing (34). Alcohol-based hand sanitizers did not have any significant effect on norovirus contamination as measured by RNA levels (34). Because norovirus is not able to be cultivated *in vitro*, it is unknown whether the viral RNA detected in these studies by RT-PCR is viable.

Cultivable viruses similar to human norovirus, such as murine norovirus and feline calcivirus, are more susceptible to ethanol than to other alcohols (53). Ethanol works by disrupting the integrity of the capsid protein, but does not affect the RNA (68). Despite its position as the most effective alcohol, ethanol is still only weakly effective against norovirus (65). Phenolic compounds are also minimally effective (54,55). The standard for disinfection is chlorine bleach (sodium hypochlorite) which is extremely effective at disinfecting surfaces with human norovirus (60). It disrupts both the capsid protein and the viral RNA (68).

Its resistance to disinfection is due to the stability of norovirus, which can persist in the environment for long periods of time. *In vitro* studies show that norovirus may remain infectious in water for years (33).

Burden of Disease

Norovirus is a leading cause of gastroenteritis in all settings and age groups of populations studied (3). Approximately 12% of all diarrheal hospitalizations in children <5 years

of age worldwide are due to norovirus (83). The virus is more prevalent in developing countries than in developed, possibly influenced by differences in hygienic practices (21). Developing countries generally have higher seroprevalence rates and earlier age of acquisition than that of developing countries (21, 22, 24). Chilean individuals over the age of 20 had >90% seroprevalence rates and continued to increase with age (22). Whereas children in the US became seropositive between 5-15 years of age, children in Bangladesh acquired antibodies earlier in life (23,82). In a population-based norovirus birth cohort study, 80% of 1 year-old children had been infected with norovirus (87). Lower socioeconomic status, lower hygiene standards, crowding, and lower maternal education level may be risk factors for all age groups (22, 23). Even developed countries can show a high prevalence of antibody to norovirus, though (13, 19, 20). Outpatient incidence rates range from 54-64 visits per 10,000 persons in the United States and Europe (18, 19,76).

Norovirus has a profound effect on the United States. Between 570 and 800 deaths, more than 60,000 hospitalizations, and 19-21 million total illnesses happen every year (18, 76, 77). Of the estimated 9.4 million cases of foodborne illnesses that occur in the US each year, approximately 58% are caused by norovirus (1). This leads to a high economic burden of $2 billion for healthcare and loss of productivity for foodborne illnesses alone (86). Between 2009-2010, norovirus was responsible for 68% of single-etiology non-foodborne outbreaks (18). Noroviruses accounted for 78% of illnesses, 46% of hospitalizations, and 86% of deaths in these outbreaks (85). In a community-based study in Georgia, the incidence of acute gastroenteritis caused by norovirus is 4 times that of bacteria and more than 15 times that of parasites (18).

Healthcare facilities, including hospitals and nursing homes, are the most frequent settings of norovirus outbreaks in industrialized countries (40). They are so common that 49% of

norovirus outbreaks in the US occur in healthcare facilities (84). Infections in these settings are more likely to have severe outcomes. Globally, the tendency is to focus on gastroenteritis deaths in children under 5. However, most (83%) norovirus-related deaths in the US are in adults over the age of 65 (20). They have the highest mortality rate (20/1,000,000 person-years), more than 15-fold higher than that of the next most prevalent age group, children aged 1-4 years (1.3/ 1,000,000 person-years) (20).

Challenges of Estimating Burden of Disease

Despite the ubiquity of norovirus in the US, it is difficult to accurately determine the burden of disease in the population for several reasons. There is a lack of a sensitive and rapid assay for diagnosing norovirus cases (78). Current diagnostic tests rely on molecular methods such as RT-PCR, which are not common in clinical settings. Without an accurate assay, norovirus cases are generally not laboratory-confirmed, meaning that they are not reported in any surveillance system. Molecular methods are prone to false positives and require multiple primer pairs (36). Antibody enzyme-linked immunosorbent assay (ELISA) tests have been shown to have high sensitivity (83%) and specificity (95-100%), but are skewed by cross-reactivity of different norovirus strains (27, 29). As of yet, no clinical laboratory-based system exists for reporting isolated and sporadic norovirus cases (78). In addition, norovirus-specific disease reporting codes are only used when a laboratory-confirmed diagnosis is obtained. Most gastroenteritis sufferers elect to stay home, knowing that their symptoms are relatively mild and self-limiting. Since only 15% of all gastroenteritis patients seek medical care, and diagnostic testing is requested for only 13% of that group, the data is woefully insensitive (18). Norovirus surveillance is almost entirely based on passively reported outbreaks. All of these factors compile to make incidence extrapolation difficult.

Three methods have been used to generate norovirus incidence estimates in the US: 1) attributable proportion extrapolation, 2) laboratory-confirmed population-based surveillance, and 3) indirect attribution from regression modeling (79). Mead et al. and Scallan et al. both used data from active surveillance systems such as FoodNet--the US surveillance system for foodborne diseases--as well as passive surveillance systems such as the National Notifiable Disease Surveillance System, to reach estimates of 21 million and 23 million norovirus associated illnesses, respectively, occurring in the US annually. (1, 77). The authors recognized that in order for a case to be included in their laboratory-based surveillance methodology, the afflicted person must visit a medical professional. There, they must submit a sample for laboratory testing. The laboratory must then perform the necessary tests to correctly identify norovirus and they must then report the result to the public health surveillance system. Failure to complete any of these steps would lead to under-diagnosis (1)

Hall et al. was the first to come up with an estimate based on direct testing of acute gastroenteritis patients in the US (18). Fecal specimens that were submitted for routine clinical diagnostics from a known population based on HMO catchment were randomly sampled to be tested for norovirus. The prevalence in these samples was scaled-up to community incidence based on FoodNet population surveys regarding health care use rates. When expanded to the US population as a whole, an estimate of 19 million cases occur annually in the US (18). This study suffers from a lack of generalizability because it was conducted in a single, somewhat homogenous community in Georgia. Similarly, Payne et al. used active surveillance enrollment and laboratory testing in three pediatric hospitals belonging to the New Vaccine Surveillance Network to estimate the number of outpatient visits for children under the age of 5 (76).

Although previous studies used PCR for identifying norovirus, many studies are moving to using ELISAs. ELISA data for serum samples is continuous from a low value (low reactivity) for negative samples to higher values (high reactivity) for positive samples. This continuous spectrum of results poses a challenge for data analysis. Traditionally, a cut-off point is chosen to distinguish a positive sample from a negative sample. This works well for the truly positive samples; for instance, from someone who was recently infected. They are expected to have a strong response to a VLP of the same strain with which they were recently infected. However, as time passes and their immunity wanes, the response will not be as strong. A weak reaction to a VLP could be caused by a past infection to the infecting strain, but it could also be due to cross-reactivity of antibodies. Cross-reactivity between genotypes generally gives a comparatively weak to moderate signal, though, so high responses are unlikely to be caused by this phenomenon (29).

Exposed and unexposed populations may not have clearly discernible distributions. The traditional approach assumes that everyone in the population responds the same to an infection. Cut-off points are often chosen to be more specific than sensitive since they will be used in clinical settings, which will skew prevalence estimates (30). Without a "gold standard" test by which to measure the sensitivity and specificity, extreme bias may be introduced.

One method of avoiding such complications is using mixture modeling. These statistical models take into account the underlying antibody distribution in the population (i.e. unexposed, previously exposed, recently exposed, and currently ill). Such models have been used in multiple studies, such as to estimate the seroprevalence for measles, mumps, and rubella in Wales (30), as well as varicella in the Netherlands (31). Peter Teunis recently developed an extension of the

mixture model, called finite mixture modeling. Rather than assuming that once someone has antibodies, they retain those antibodies indefinitely, finite mixture modeling takes into account antibody decay. It models each sample concentration as a probability that it is truly positive or truly negative. Beyond that, it responds to individual variation in serum antibody response (Teunis, unpublished presentation). It can account for cross-reactivity of antibodies in ELISA tests, and, since it is based on a likelihood function, analysis of uncertainty is relatively simple.

The purpose of this project is to estimate the seroprevalence of norovirus in the US population, using finite mixture modeling on National Health and Nutrition Examination Survey (NHANES) samples. Every two years, about 5,000 people are interviewed and clinically examined as part of this National Center for Health Statistics project. Great pains are exercised to make it a sample nationally representative of the U.S. population at all ages. Each participant undergoes an interview that includes demographic, health, and diet-related questions. The clinical component includes medical, dental, and physiological examination. Laboratory tests are performed on biological samples collected from participants. These samples are placed in a bank for research use. Finite mixture modeling will be applied to quantitative ELISA data received from serum samples of a 1/3 subset of the 1999-2000 and 2003-2004 NHANES survey. In this nationally representative sample, the seroprevalence to three common genotypes of norovirus can be calculated while minimizing the bias introduced in other study methods. Changes in seroprevalence between age group and between years can be calculated. All of these will contribute to a more accurate picture of the burden norovirus places on the United States population.

Materials and Methods

Research Design

This study is a cross-sectional study of 2,168 participants of the 1999-2000 and 2003-2004 NHANES cohorts. These collection cycles were chosen to sample from the population before and after the introduction of the Farmington Hills GII.4 pandemic strain. Changes in immunity could be assessed for both time points.

NHANES biological samples are not available for children under the age of 5. Because NHANES does not sample from institutionalized individuals, including those in the hospital or nursing homes, the elderly population is also highly limited. Due to these factors, for this study, NHANES serum samples were used to create a nationally representative sample of individuals between the ages of 16 and 49 years old (**Figure 1**). Previous studies have shown persons between the ages of 16 and 60 are most likely to interact with people from all age groups (98). This group was chosen to ensure that all transmission pathways were represented in the sample. Additionally, by restricting the sample to this age group, potential age-related confounders such as maternal antibodies and immune status will be minimized (97). Due to funding constraints, a 1/3 subset of the complete 1999-2000 and 2003-2004 cohorts were used. Based on the high previously reported seroprevalence levels, this subset will be more than sufficiently powered to estimate norovirus seroprevalence in both study periods. The lowest reported adult seroprevalence, 59%, would require 366 subjects to determine the proportion of seropositive individuals with a precision of 1% on the 95% confidence interval (92). The literature more often reports a higher seroprevalence, though, that is over 80% (92-97). If the seroprevalence truly is that high, the results of this study will be powered to stratify by age group as well.

Years of Collection	Age group	Number of Samples
1999-2000	16-19	341
	20-29	253
	30-39	228
	40-49	242
2003-2004	16-19	339
	20-29	260
	30-39	266
	40-49	239
Total		2,168

Table 1. The collection cycle and age groups of 2,168 participants in the NHANES study. The numbers of samples are a 1/3 subset of each age group from the 1999-2000 and 2003-2004 collection cycle.

VLP creation

Norovirus virus-like particle (VLP) creation was all done as part of the USDA's Norovirus Collaborative for Outreach, Research, and Education (NoroCORE) program. Researchers from Baylor College of Medicine, the United States Centers for Disease Control and Prevention (CDC), and Emory University created norovirus VLPs for all genotypes that infect humans. All VLPs were expressed in baculovirus, one of the most common expression systems. The most pervasive strains were chosen from each genotype to more accurately assess the prevalence of the virus (**Table 2**). These genotypes were chosen based on the findings of CDC surveillance data and Verhoef et al.'s analysis of bivalve mollusk monitoring data and outbreak data from Europe (28). Three different GII.4 VLPs were made with the pandemic strains GII.4 US-95/96, GII.4 Farmington Hills, and GII.4 New Orleans. Although GI.1 viruses rarely contribute to outbreaks, the VLP was included in this study because it is often used as a reference strain. For the purpose of this thesis, GI.4, GII.3, and GII.4 US-95/96 were analyzed.

Norovirus Strain	Pandemic Era
GI.1 Norwalk	NA
GI.4	**NA**
GII.3	**NA**
GII.4 US-95/96	**1995-2002**
GII.4 Farmington Hills	2002-2005
GII.4 New Orleans	2009-2012
GIV.1	NA

Table 2. VLPs created in a baculovirus expression system. Bolded strains are strains used in this study.

Quantitative IgG ELISA

96- well plates were coated with 0.2 μg VLP suspended in PBS. A standard curve was generated using human IgG ranging from a concentration of 2000ng/mL to 31.25 μg/mL. PBS was used as a blank. After incubating for more than 1 hour, the plates were washed with PBS-Tween and blocked with 5% Blotto (block). Serum samples were diluted 1:50 in the block, added to duplicate wells, and incubated at room temperature for 1 hour. Antibody complexes were detected with goat anti-human IgG-alkaline phosphatase and pNPP developer. Plates were read at 405nm. Antibody concentrationis were calculated by comparsion to an IgG standard curve.

Data analysis

Values above the level of detection were set to a value slightly above the highest average concentration recorded. Likewise, values below the level of detection were set to a value marginally lower than the lowest average concentration recorded. The minimum values were set to 24.34, 84.34, and 115.54 for GI.4, GII.3, and GII.4 US-95/96, respectively. The maximum values were set to 823,419, 40,785,452, and 2,967,252 for GI.4, GII.3, and GII.4 US-95-96, respectively. Non-normal data with two distinct distributions are needed for mixture modeling to detect those who are seropositive and seronegative in a population. To check for this requirement, IgG concentrations were log-transformed and plotted in a histogram using SAS

version 9.4. The histograms also provided preliminary starting values for μ_1, μ_2, and prevalence for the mixture model. Correlations between the three genotypes were examined.

All finite mixture modeling was run on R version 3.0.2. Each genotype's ELISA data was analyzed separately. The log-transformed concentrations were again plotted in a histogram. The preliminary starting values for the 5 necessary parameters were input. These parameters are the mean and standard deviation for concentration in the seronegative population (μ_1 and SD_1), mean and standard deviation for concentration in the seropositive population (μ_2 and SD_2), as well as the estimated seroprevalence. They were inversely transformed before their maximum likelihood estimates were calculated. After transforming back to the original values, the prevalence was calculated by the program. These were then graphed as a density function. The model parameters were modified to optimize the sensitivity and specificity while also minimizing deviance.

Mixture models were created for the aggregated data for each genotype. Then, each genotype was stratified on the NHANES sample collection cycle years, 1999-2000 and 2003-2004. Each genotype was then also stratified by age groups. Ages were categorized as 16-19, 20-29, 30-39, and 40-49 years old.

Because of the large proportion of values above the ELISA's detection level, the previous analyses were repeated with the extreme values deleted. To determine whether these high values biased the finite mixture modeling, the models with the extreme values included and the model with them excluded were compared to what was expected using the traditional cut-off method. Serological concentration data from a GI.I norovirus human challenge study were log-transformed (32). The anti-norovirus IgG concentration value at the 75^{th} percentile of this data of uninfected subjects, 10.99, was set as a cut-off point. This provides a biologically-derived cut-off point, rather than an arbitrary value. This cut-off point was then applied to the NHANES data.

Anti-norovirus IgG concentrations above 10.99 were categorized as seropositive, and values below it were categorized as seronegative. These seroprevalence values were compared to the finite mixture modeling results with and without extreme concentrations.

Results

Validation with GI.I Challenge Data

The data used in this study had a large proportion of values above the level of detection. To assess whether all of these high values on the finite mixture modeling, the model was applied to the full dataset, as well as the dataset with the concentrations above the level of detection removed. In order to find which method would provide the most accurate results, the traditional cut-off approach to serological analysis was used as a gold standard. The anti-norovirus concentration value at the 75% percentile of uninfected individuals from a previous human challenge study were used to derive a cut-off value with a biological basis(32). When this cut-off of value of 10.99 was applied to the complete NHANES dataset, resulting seroprevalence values ranged from 32-72%, varying by genotype, collection cycle, and age group. Finite mixture modeling with no exclusions estimated seroprevalence values between 37-98%. Excluding the concentrations above the level of detection for finite mixture modeling led to generally higher seroprevalence values, ranging from 69-84%. A multiple comparison ANOVA test revealed that excluding values above the detection level was statistically different from both the traditional cut-off method and using finite mixture modeling on the complete dataset ($F=19.35$, $p<0.0001$). Because the finite mixture modeling without exclusions agreed more with the values calculated using the traditional method, this dataset was used for the rest of the analyses.

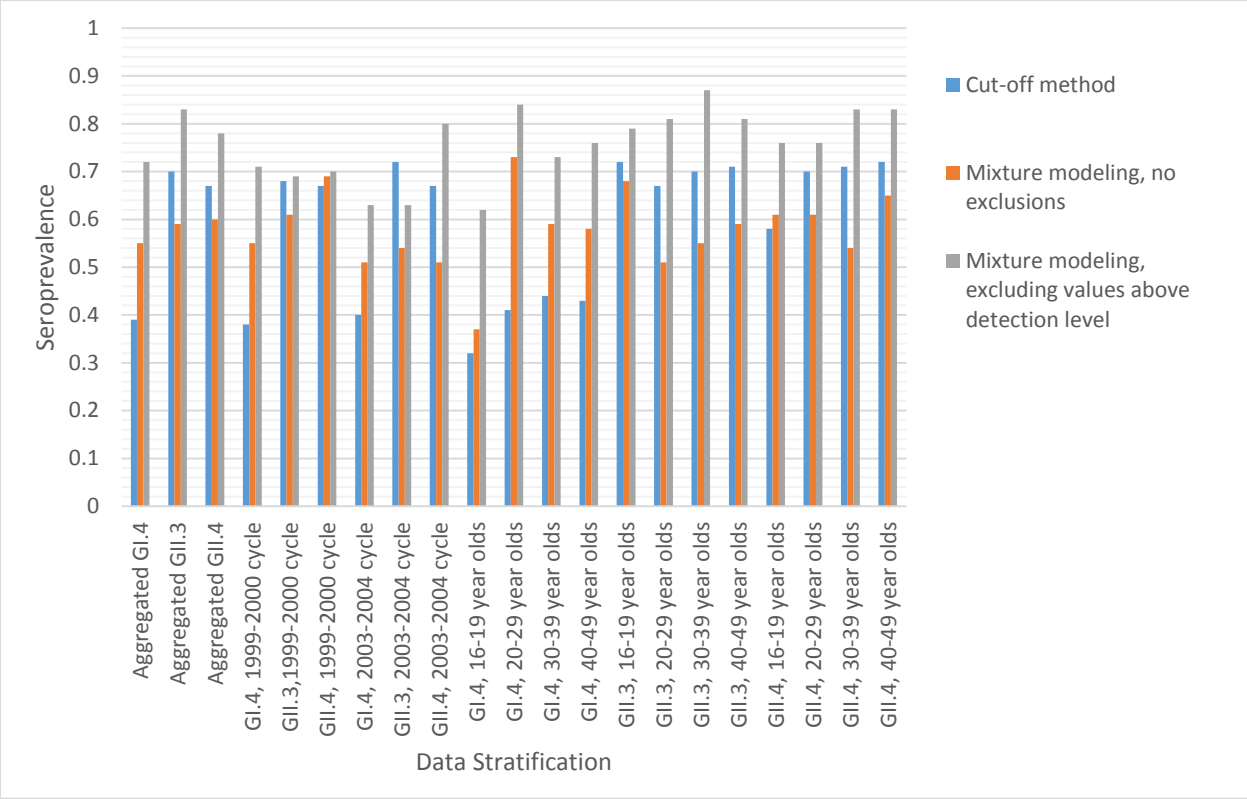


Figure 1: A concentration cut-off value of 10.99 was derived from previously published GI.I challenge data (32). Seroprevalences were calculated using the traditional cut-off method, with all log-transformed anti-norovirus concentration values above 10.99 being categorized as positive. These seroprevalence values were compared to values computed by finite mixture modeling using all data, as well as excluding values above the level of quantification.

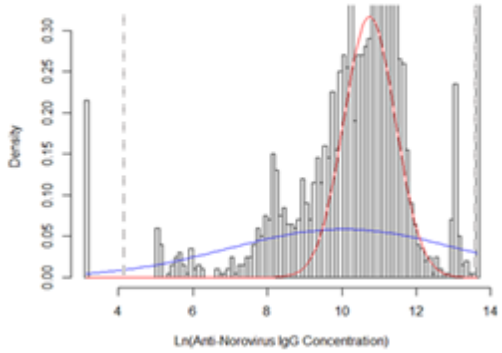
Aggregated Results

Seroprevalence estimates for the three genotypes over both collection cycles ranged from 55-60% (**Table 3**). GI.4 was the least prevalent, with an estimated seroprevalence of 55% (μ_1=10.05, μ_2=10.75). The predicted seropositive and seronegative populations had similar mean concentration values. This was not the case for the GII.3 and GII.4 genotypes. Both had seropositive means close to 11, (11.10 and 11.07, respectively), but had seronegative means that were much higher. The GII.3 had a seronegative mean of 17.74 (SD=0.69), which was beyond the level of quantification in this study. GII.4 also had a high mean for the seronegative individuals at 14.72 (SD=5.73). GII.3 and

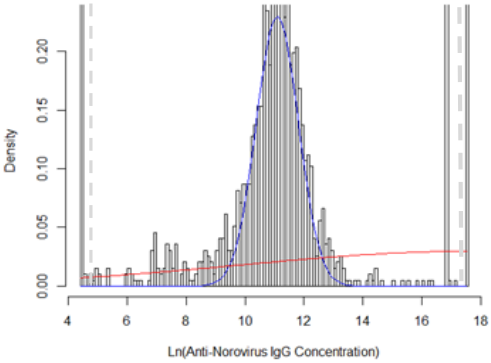
GII.4 had a higher proportion of values censored at the high end of the spectrum, which pulled the seronegative mean higher than that of the seropositive population (**Figure 2**).

	VLP		
Parameter	**GI.4**	**GII.3**	**GII.4 US-95/96**
μ_1 (SD)	10.05 (3.07)	11.10 (0.72)	11.07 (0.56)
μ_2 (SD)	10.75 (0.69)	17.74 (7.85)	14.72 (5.73)
Seroprevalence	0.55	0.59	0.60

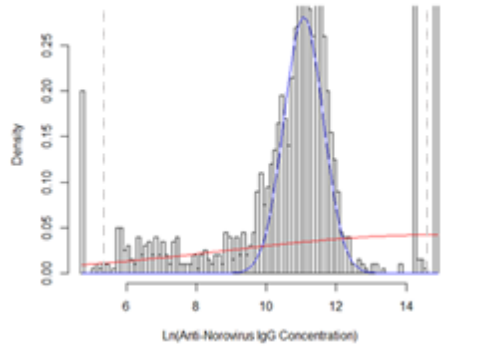
Table 3: Mixture modeling parameters and seroprevalence estimation for three Norovirus genotypes. Quantitative ELISAs were run on serological samples from the 16-49 year olds in the 1999-2000 and 2003-2004 NHANES collection cycles.
μ_1: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-positive population
μ_2: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-negative population



GI.4



GII.3



GII.4 US- 95/96

Figure 2- Finite mixture models of norovirus seropositivity in a 1/3 subset of 16-49 year olds from the 1999-2000 and 2003-2004 NHANES samples. Dotted lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals.

Collection cycle comparison

All three genotypes had a higher seroprevalence in the 1999-2000 collection cycle than in the 2003-2004 **(Table 4).** GI.4 seroprevalence decreased significantly from 55% in the 1999-2000 samples to 51% in 2003-2004 (χ^2 p< 0.0077). GII.3 also decreased significantly between the two collection cycles from 61% to 54% (χ^2 p< 0.002). GII.4 US-95/96 decreased the most, dropping from 69% to 51% over the time period (χ^2 p< 0.0001).

	VLP					
Parameter	**GI.4**		**GII.3**		**GII.4 US-95/96**	
NHANES Cycle	1999-2000	2003-2004	1999-2000	2003-2004	1999-2000	2003-2004
μ_1 (SD)	9.35(3.33)	10.65(2.50)	11.31(0.77)	10.93(0.68)	11.16(0.46)	10.96(0.66)
μ_2 (SD)	10.76(0.73)	10.75(0.63)	15.56(8.48)	19.60(5.92)	14.48(6.96)	14.92(3.83)
Prevalence	0.55	0.51	0.61	0.54	0.69	0.51
χ^2, (p-value)	7.1023(P<0.0077)		9.807(P< 0.002)		67.448(p<0.0001)	

Table 4: Mixture modeling parameters and seroprevalence estimation for three Norovirus genotypes. Quantitative ELISAs were run on serological samples from the 16-49 year olds in the 1999-2000 and 2003-2004 NHANES collection cycles. A chi-square test of proportions was run to test for significance at the 95% confidence level.
μ_1: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-positive population
μ_2: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-negative population

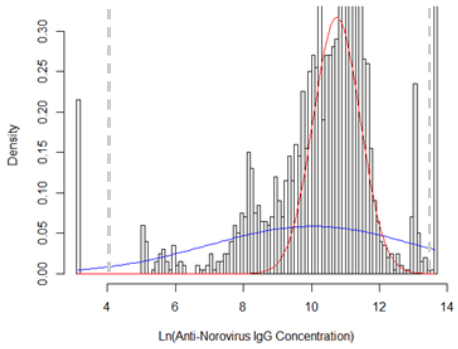
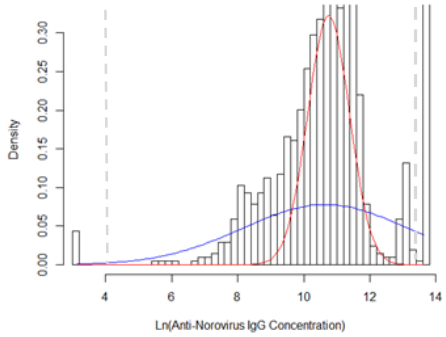
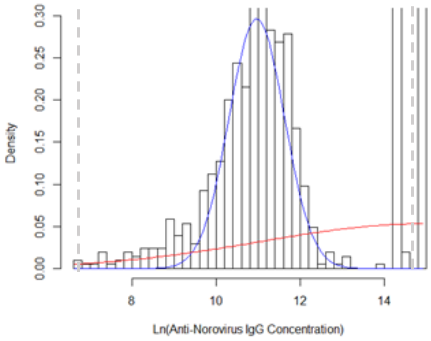
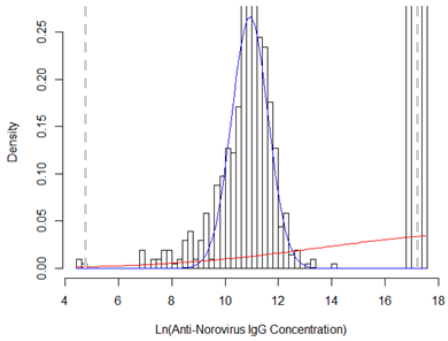
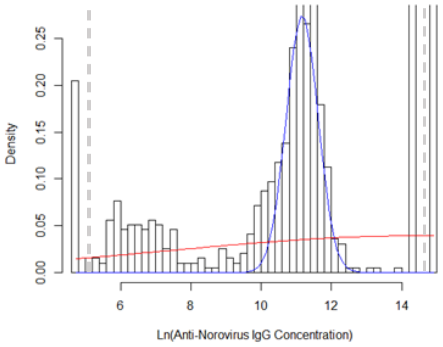
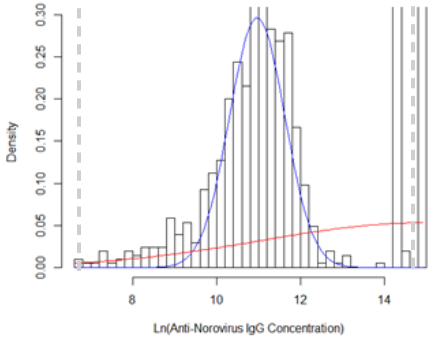
	1999-2000	2003-2004
GI.4		
GII.3		
GII.4-US 95/96		

Figure 3- Finite mixture models of norovirus seropositivity in a 1/3 subset of 16-49 year olds from the 1999-2000 and 2003-2004 NHANES samples. Dashed lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals

Age Group Comparisons

The highest GI.4 seroprevalence was among 20-29 year olds, with 73% of them being categorized as seropositive (μ_1=9.62, μ_2=10.66; **Table 5**). The lowest predicted seroprevalence in this genotype, 37%, was among the 16-19 year olds (μ_1=10.21, μ_2=10.66). These findings are reversed in the GII.3 genotype. In this genotype, 16-19 year olds have a seroprevalence of 68%, which is the highest of all of the age groups (μ_1=11.04, μ_2=18.39). The 20-29 year old group is the lowest seroprevalence for GII.3, with 51% predicted to be seropositive (μ_1=11.09, μ_2=17.30). At a seroprevalence rate of 65%, 40-49 year olds were predicted to be the most frequently seropositive to GII.4 US-95/96 (μ_1=11.19, μ_2=15.46). The strain was predicted to be least prevalent in 30-39 year olds (μ_1=11.15, μ_2=15.56). As with previous analyses, the GI.4 genotype had distributions less-affected by values beyond the ability to quantify.

	VLP											
	GI.4				**GII.3**				**GII.4 US-95/96**			
Age (years)	16-19	20-29	30-39	40-49	16-19	20-29	30-39	40-49	16-19	20-29	30-39	40-49
μ_1	10.21	9.62	9.89	10.09	11.04	11.09	11.09	11.15	10.83	11.18	11.15	11.19
(SD)	(1.04)	(3.94)	(3.35)	(2.95)	(0.67)	(0.74)	(0.68)	(0.76)	(0.58)	(0.51)	(0.54)	(0.53)
μ_2	10.66	10.66	10.92	10.86	18.39	17.30	17.11	17.74	13.91	14.51	15.56	15.46
(SD)	(3.66)	(0.95)	(0.59)	(0.62)	(8.31)	(8.38)	(6.81)	(7.65)	(5.50)	(5.55)	(6.13)	(5.92)
Preval -ence	0.37	0.73	0.59	0.58	0.68	0.51	0.55	0.59	0.61	0.61	0.54	0.65

Table 5. Mixture modeling parameters and seroprevalence estimation for three Norovirus genotypes, stratified by age group. Quantitative ELISAs were run on serological samples from 16-49 year olds in the 1999-2000 and 2003-2004 NHANES collection cycles.
μ_1: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-positive population
μ_2: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-negative population

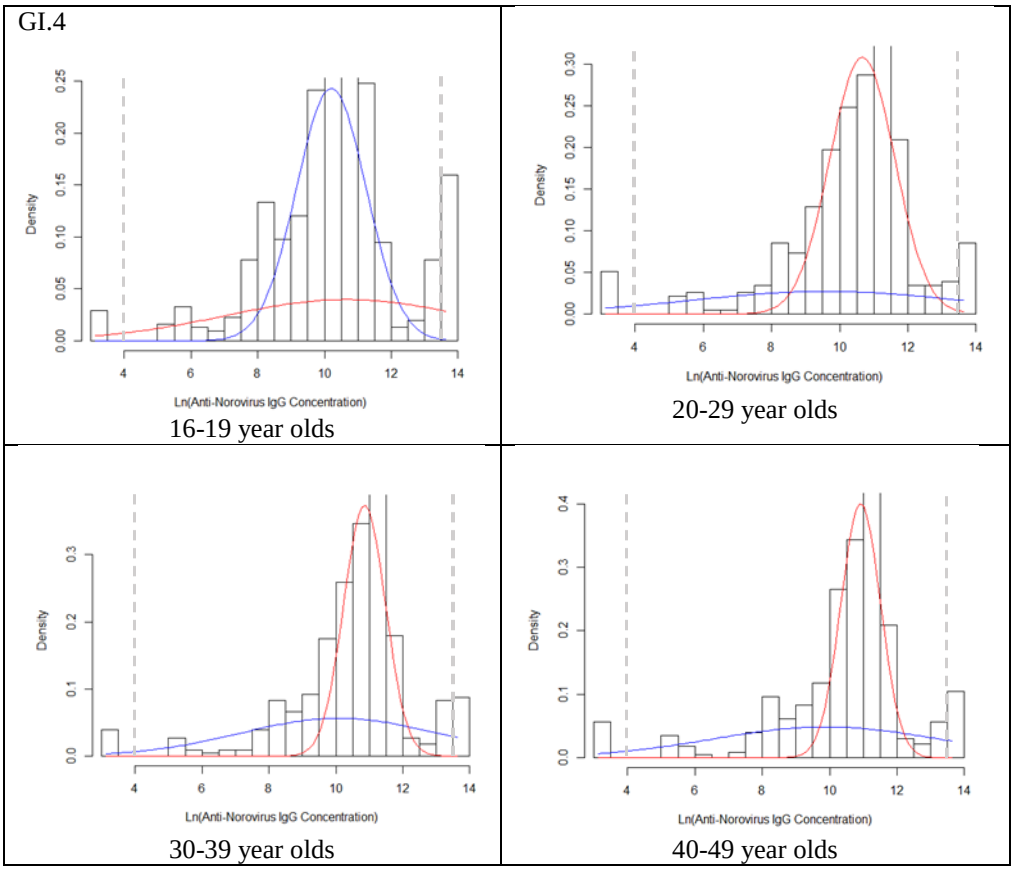


Figure 4- Finite mixture models of GI.4 norovirus seropositivity in a 1/3 subset from the 1999-2000 and 2003-2004 NHANES samples, stratified by age groups. Dotted lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals

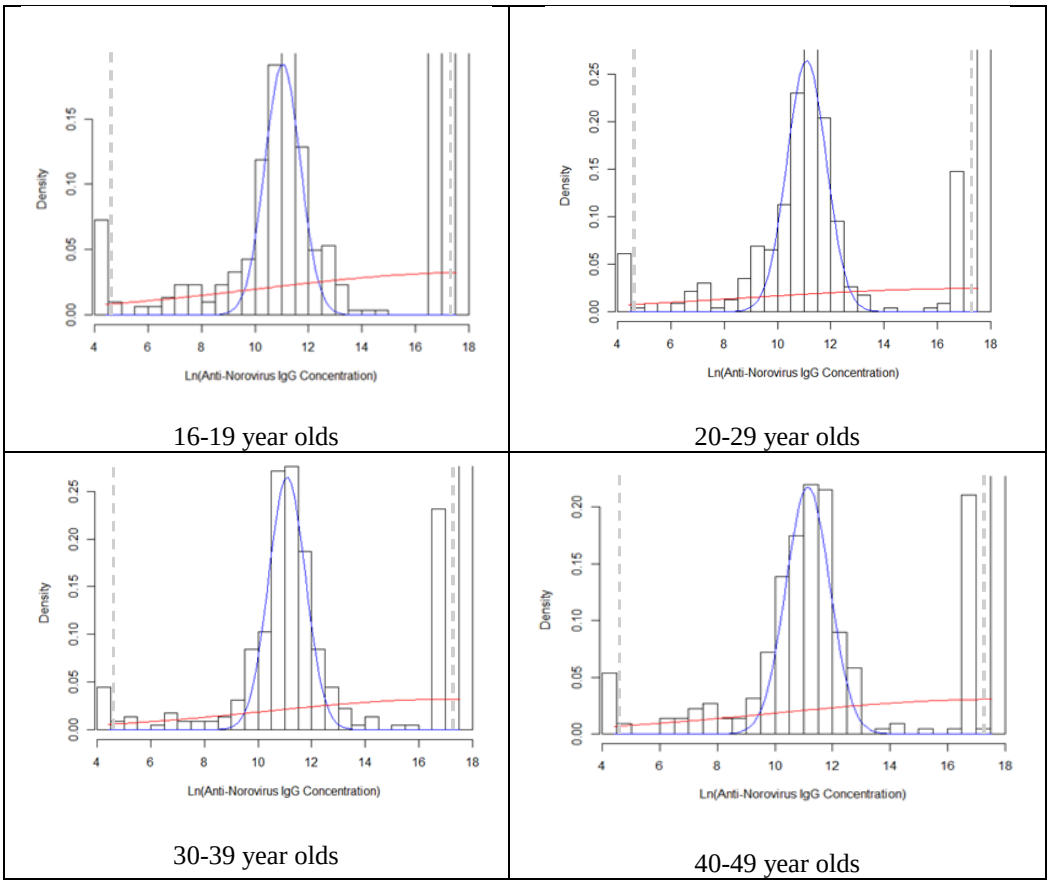


Figure 5- Finite mixture models of GII.3 norovirus seropositivity in a 1/3 subset from the 1999-2000 and 2003-2004 NHANES samples, stratified by age groups. Dotted lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals

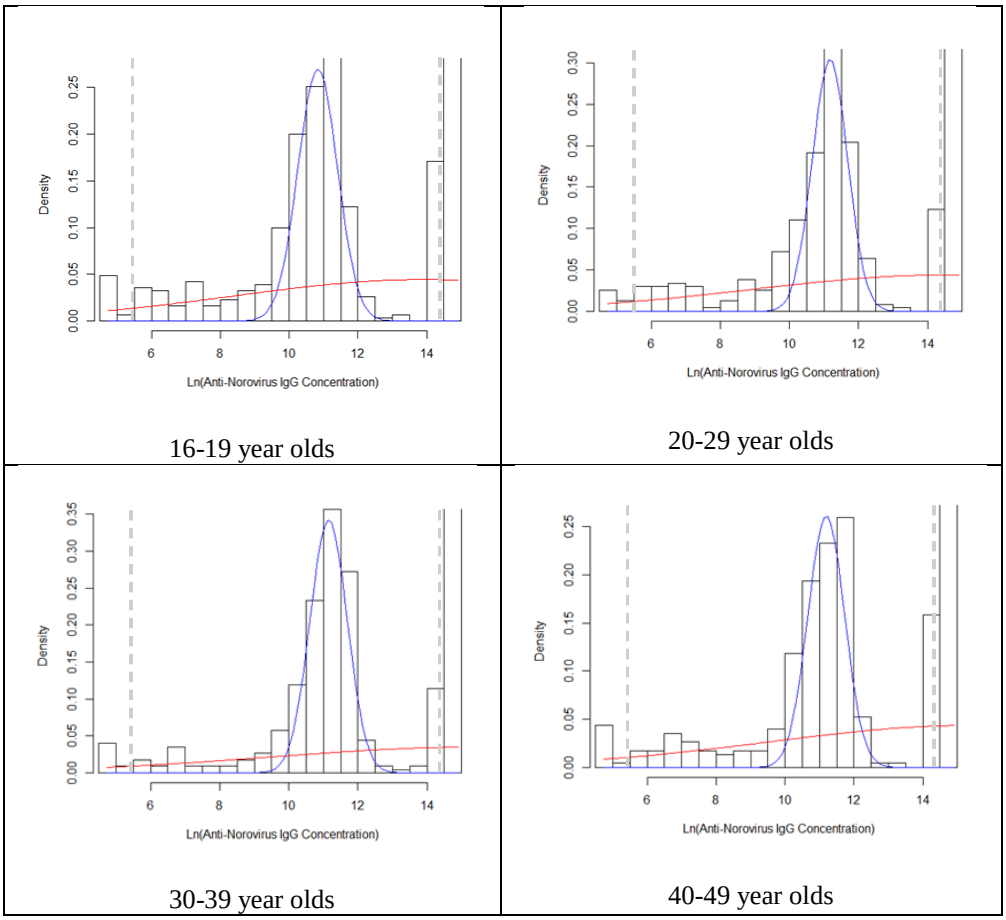


Figure 6- Finite mixture models of norovirus GII.4 norovirus seropositivity in a 1/3 subset from the 1999-2000 and 2003-2004 NHANES samples, stratified by age groups. Dotted lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals

Discussion

Finite mixture modeling was used to estimate the seroprevalence of three norovirus genotypes in the United States population. Quantitative ELISAs were run on 2,168 serological samples from the 1999-2000 and 2003-2004 National Health and Nutrition Examination Survey (NHANES) with participants being between the ages of 16 and 49. While finite mixture modeling has been used to model *Campylobacter jejuni* and *Salmonella* seroprevalence in the Netherlands, this study is the first time that the model has been applied to norovirus serological data (Teunis, unpublished data). Traditionally, seroprevalence is determined by setting a fixed cut-off point. Any value above the cut-off is categorized as seropositive. The cut-off values chosen are often arbitrary, not having a biological significance. This method also fails to take into account the individual variation in immunogenic response. In contrast, finite mixture modeling distinguishes two different populations of individuals with heterogenous responses. Each concentration value is modeled as the probability of being truly seropositive or seronegative. By doing so, it also addresses cross-reactivity of genotypes. As shown by the model, higher concentrations are much more likely to be due to an infected individual, rather than strong reactivity in an uninfected individual. Similarly, lower concentrations are much more likely to be found in the seronegative population. Some uncertainty does exist, though, as evidenced by where the distributions overlap.

In theory, the finite mixture model should detect two populations—the seronegative and seropositive. However, discrimination between these two populations was poor in this study, partially because of the high number of concentration values above the level of detection. Between the abundance of values above the level of detection, and the relatively smaller amount below the level of quantification, the GII.3 and GII.4 US-95/96 had 37% and 36% of their values

censored. These censored points created a third, "super-positive" population that was unexpected.

The "super-positive" distribution could represent multiple things. First, it could represent individuals who were very recently infected. Their serological titres would be expected to be generally much than people who had been infected less recently. It could also be caused by a homologous response, which is when the VLP is from the same strain that infected the patient. The moderate positive distribution could be the heterologous, or cross-reactive reaction.

In the stratifications with large proportions of data in the "super-positive" category above the level of quantification, the model struggled to distinguish between the seronegative and seropositive populations. This resulted in some of the models having an extremely broad seronegative distribution, and thus, a higher seropositive mean (μ_1) than seronegative mean (μ_2) concentration. The GII.3 and GII.4 US-95/96 genotypes were especially prone to this phenomenon.

Between the two collection cycles, the seroprevalence of GII.4- US 95/96 decreased significantly. In the samples from 1999-2000, the seroprevalence was 69%. This number dropped to 51% in 2003-2004, representing a significant change between the collection cycles (χ^2 $p<0.0001$). Dominant GII.4 strains are replaced by new strains every 2-3 years (3). The US 95/96 strain (previously known as Grimsby) was the dominant GII.4 strain in 1999 and 2000, being the first identified pandemic GII.4 strain (103). The Farmington Hills strain emerged in 2002 and replaced US 95/96 as the most dominant strain. This would account for the decline in US 95/96 observed in this study. It is also consistent with current theories of immunity-driven pandemic cycling of the GII.4 genotype (15, 70, 80). The GII.4 genotype has found a beneficial balance of replication and mutation rate that puts it at an evolutionary advantage over other genotypes (14).

This, coupled with other factors such as high capsid diversity, allow new strains to escape herd immunity (14).

Limitations

The seroprevalence predictions modeled here were lower than what was expected for the United States population. Son et al. found that the seroprevalence for GI.4, GII.3, and GII.4 genotypes was 84%, 76%, and 94%, respectively (42). In Italy, 98.6% seroprevalence was found against GII.4 (99). South Africa and China also found high seroprevalence rates (93, 96).

One reason for the discrepency observed in this study could be that NHANES does not sample from institutionalized populations, which includes hospitalized individuals and those in nursing homes. Healthcare facilities, including hospitals and nursing homes, are the most frequent settings of norovirus outbreaks in industrialized countries (40). They are so common that 49% of norovirus outbreaks in the US occur in healthcare facilities (84). Infections in these settings are more likely to have severe outcome. Globally, the tendency is to focus on gastroenteritis deaths in children under 5. However, most (83%) norovirus-related deaths in the US are in adults over the age of 65 (20). Additionally, people below the age of 16 were not included in this study. Children under the age of 5 is the age group with the highest norovirus prevalence, after adults over the age of 65 (20). By excluding this group, the seroprevalence would naturally drop.

This study was also limited by the large amount of data above the level of quantification. Although censoring is accounted for in the model, the density of high values pulled the seronegative distribution up into a higher range, which was expected to contain the seropositive population. A tri-modal distribution was created where only a bi-modal was expected. The model had difficulties optimizing the parameters to find the seronegative population, so it was often stretched to accommodate the high values.

Conclusion

Finite mixture modeling has been shown to be a useful tool for analyzing serological data. By applying it to NHANES samples, it was found that the seroprevalence of norovirus is, on average, between 55-60% for 16-49 year olds. The seroprevalence of GI.4 and GII.3 genotypes decreased significantly between the 1999-2000 and 2003-2004 cycle, but not as markedly as the GII.4 US-95/96 genotype. This agrees with current theories of immunity-driven GII.4 evolution. Additionally, the seroprevalence for each age group varied by genotype, possibly due to differences in transmission patterns. This study was limited by the large amount of data above the level of detection.

Future Directions

Because of the large proportion of data that was above the level of quantification, it was difficult to model the distributions of the seropositive and seronegative populations. For more accurate results, future studies should re-run the ELISAs of any samples that were above or below the level of quantification using either a more dilute or more concentrated standard curve. This would give a clearer depiction of the tri-modal distribution hinted at in the current study. It would also result in more accurate seroprevalence estimations, as the model would not be as encumbered by the extreme values.

Additionally, a study that was expanded to include all ages represented in the NHANES would be beneficial to elucidate the differences in seroprevalence rates between age groups. In the current study, only participants between the ages of 16 and 49 were included. Although this sample is nationally representative for that section of the population, it cannot be used to extrapolate the seroprevalence rates for those outside of the age range. And because children and the elderly are the groups most burdened by norovirus, including them could yield drastically higher overall seroprevalence rates. Although NHANES does not have biological samples for children under the age of 5, and the elderly population would be limited by institutionalized individuals, expanding the study ages would add more information to an overall seroprevalence estimation.

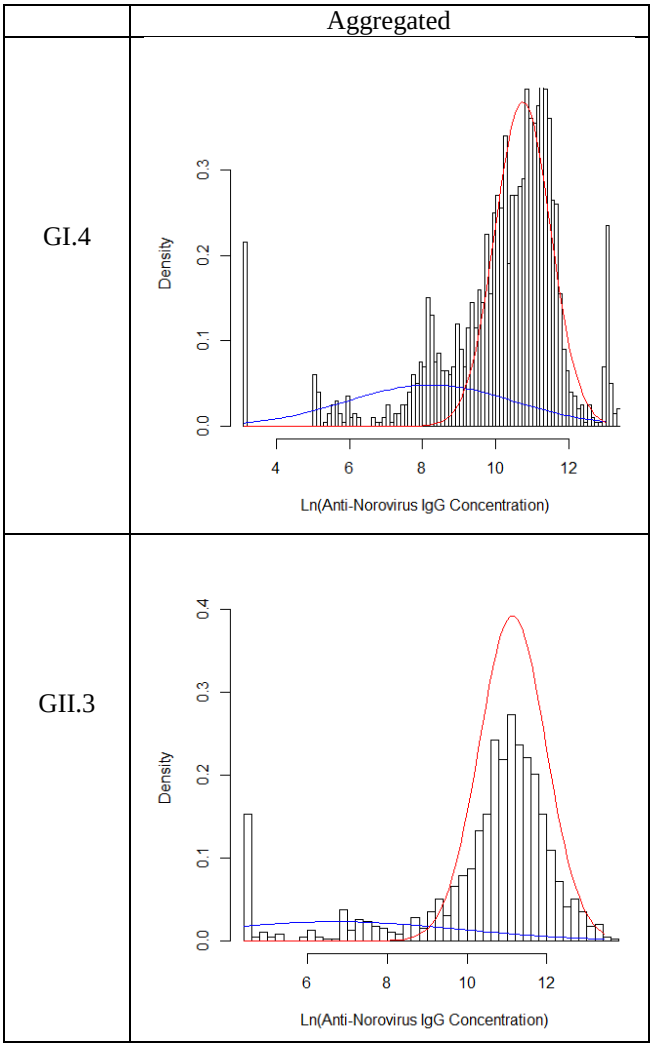
Finite mixture modeling can be applied to a variety of future studies involving serological data. Mixture modeling has already been used for varicella and measles, mumps, and rubella (30, 31). Finite mixture modeling has only been applied to unpublished data thus far. Rather than using the traditional cut-off method for analyzing serological data, finite mixture modeling could become the new standard for all seroprevalence studies.

The NHANES samples in this study should also be tested against the other VLP particles created by NoroCORE. By doing so, the seroprevalence of each genotype can be estimated in the United States population. This information would be beneficial for many future studies, including the development of a norovirus vaccine. A norovirus vaccine is currently in Phase 2 clinical trials, but in order to know how large of a sample size to use in the Phase 3 trial, the disease prevalence is needed. Completing all of the VLP analyses will give results that can be used to know which proportion of a population needs to be vaccinate to prevent epidemics.

Appendix

	VLP		
Parameter	**GI.4**	**GII.3**	**GII.4 US-95/96**
μ_1 (SD)	8.22 (2.27)	6.71 (3.09)	7.31 (2.17)
μ_2 (SD)	10.74 (0.76)	11.14 (0.83)	11.07 (0.63)
Seroprevalence	0.72	0.82	0.78

Table 1: Mixture modeling parameters and seroprevalence estimation for three Norovirus genotypes. Quantitative ELISAs were run on serological samples from the 16-49 year olds in the 1999-2000 and 2003-2004 NHANES collection cycles.
μ_1: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-positive population
μ_2: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-negative population



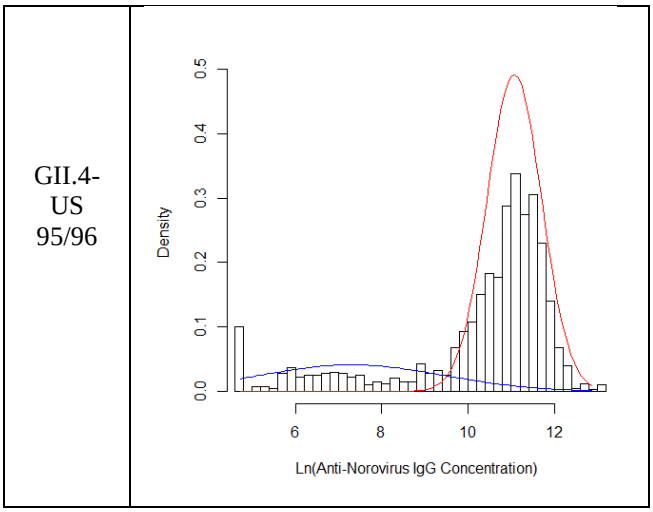


Figure 1- Finite mixture models of norovirus seropositivity in a 1/3 subset of 16-49 year olds from the 1999-2000 and 2003-2004 NHANES samples. Dotted lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals.

	VLP					
Parameter	**GI.4**		**GII.3**		**GII.4 US-95/96**	
NHANES Cycle	1999-2000	2003-2004	1999-2000	2003-2004	1999-2000	2003-2004
μ_1 (SD)	7.52 (2.50)	9.34 (1.65)	6.97 (4.16)	9.34 (1.65)	6.27 (1.65)	9.47 (1.17)
μ_2 (SD)	10.73 (0.79)	10.84 (0.63)	11.37 (0.88)	10.84 (0.63)	11.10 (0.59)	11.10 (0.63)
Prevalence	0.71	0.63	0.69	.63	0.70	0.80
Proportion Deleted	0.07	0.1	0.31	0.43	0.38	0.34

Table 2: Mixture modeling parameters and seroprevalence estimation for three Norovirus genotypes. Quantitative ELISAs were run on serological samples from the 16-49 year olds in the 1999-2000 and 2003-2004 NHANES collection cycles. A chi-square test of proportions was run to test for significance at the 95% confidence level.
μ_1: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-positive population
μ_2: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-negative population

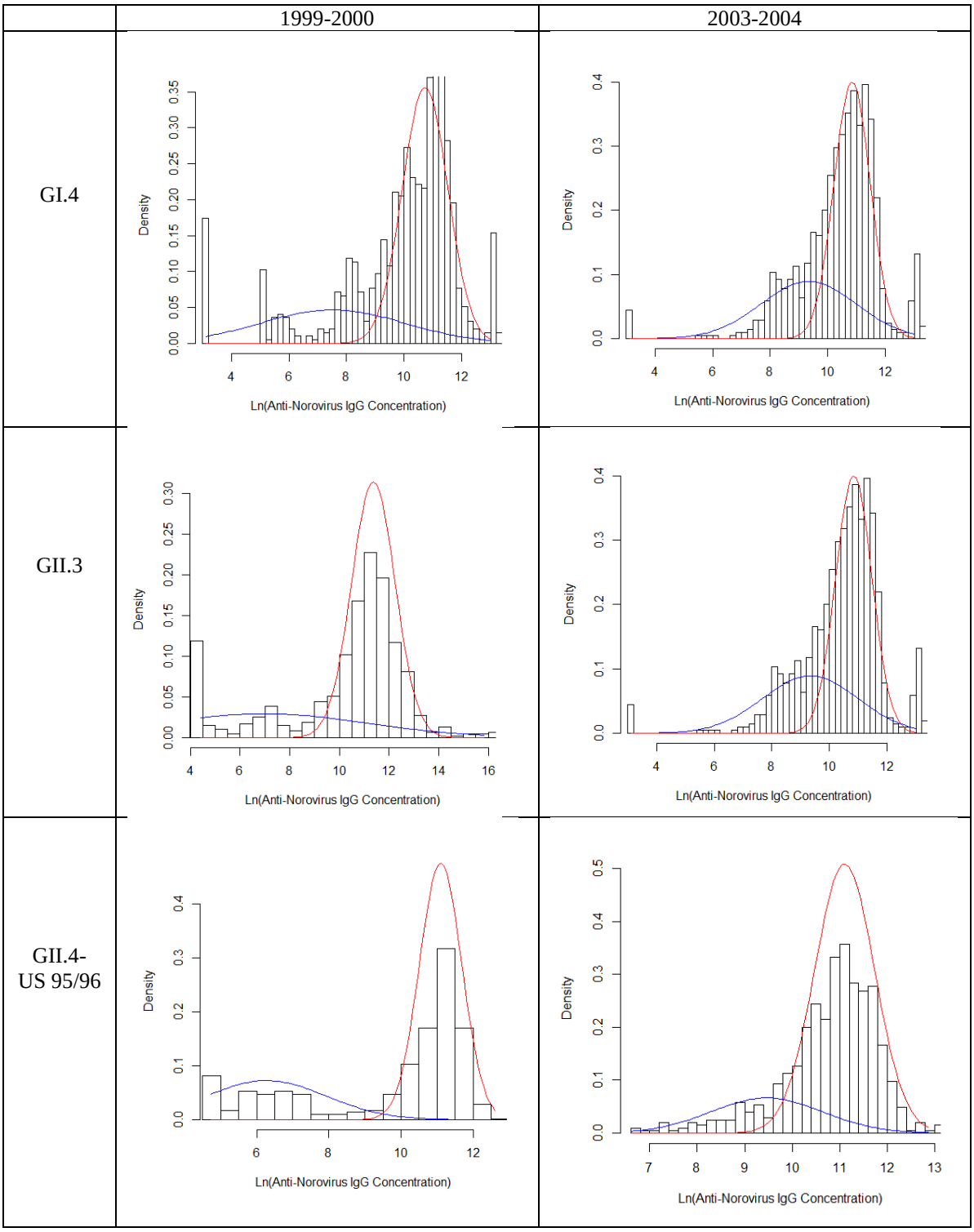


Figure 2- Finite mixture models of norovirus seropositivity in a 1/3 subset of 16-49 year olds from the 1999-2000 and 2003-2004 NHANES samples. Dashed lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals

	VLP											
	GI.4				GII.3				GII.4 US-95/96			
Age (years)	16-19	20-29	30-39	40-49	16-19	20-29	30-39	40-49	16-19	20-29	30-39	40-49
μ_1	8.50	7.17	8.10	8.29	7.55	7.18	6.36	7.02	7.13	7.67	7.21	7.22
(SD)	(1.89)	(2.17)	(2.52)	(2.27)	(3.13)	(2.96)	(3.10)	(2.97)	(2.01)	(2.04)	(2.37)	(2.16)
μ_2	10.54(	10.64	10.91	10.83	11.13	11.12	11.14	11.18	10.85	11.17	11.14	11.20
(SD)	0.74)	(0.95)	(0.64)	(0.69)	(0.78)	(0.80)	(0.83)	(0.82)	(0.65)	(0.60)	(0.60)	(0.60)
Prevalence	0.62	0.84	0.73	0.76	0.79	0.81	0.87	0.81	0.76	0.76	0.83	0.88
Proportion deleted	0.11	0.06	0.08	0.08	0.43	0.30	0.35	0.39	0.33	0.34	0.34	0.43

Table 3. Mixture modeling parameters and seroprevalence estimation for three Norovirus genotypes, stratified by age group. Quantitative ELISAs were run on serological samples from 16-49 year olds in the 1999-2000 and 2003-2004 NHANES collection cycles.

μ_1: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-positive population

μ_2: estimated mean log-transformed anti-norovirus IgG concentration for the norovirus-negative population

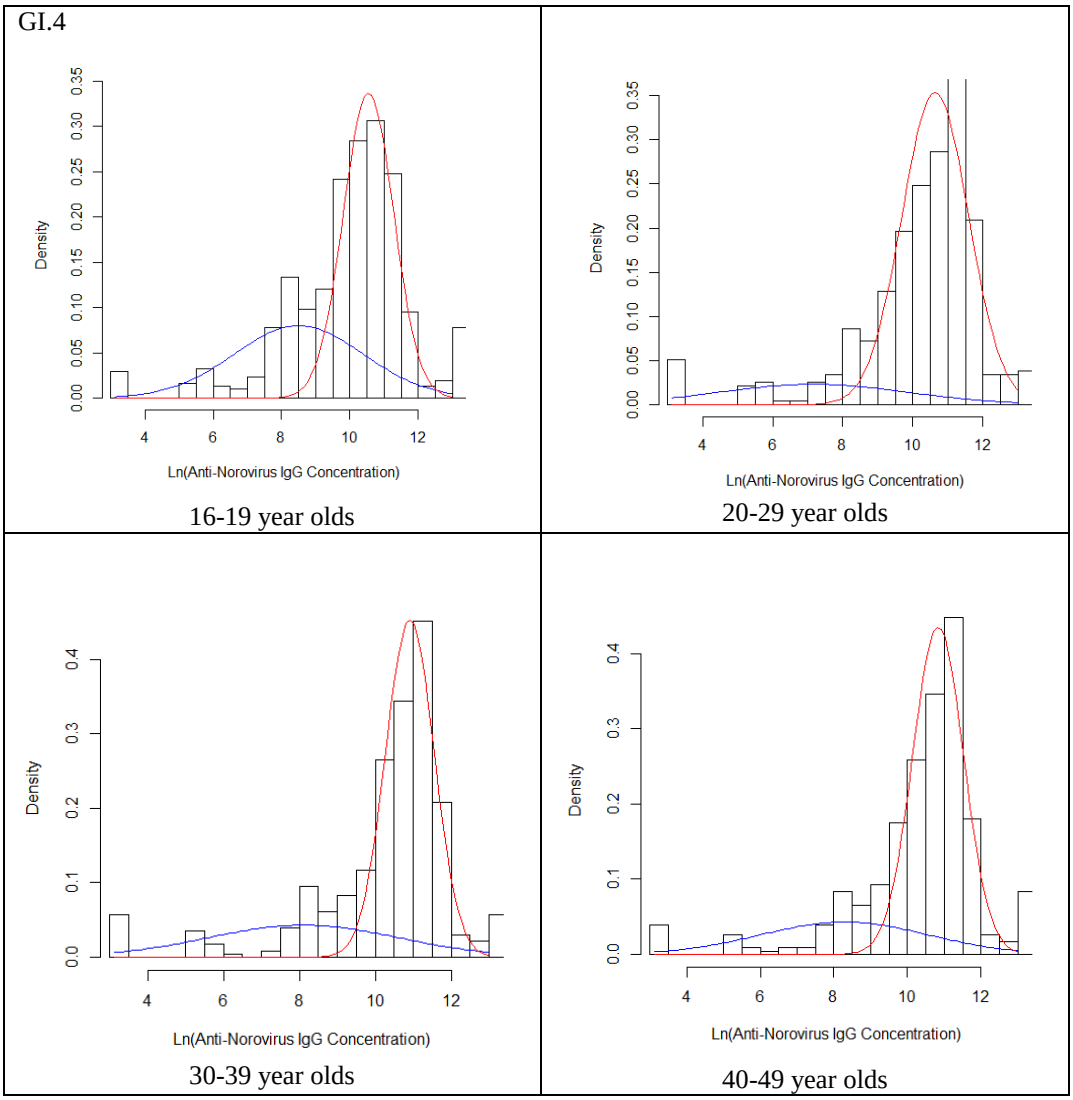


Figure 4- Finite mixture models of norovirus seropositivity in a 1/3 subset of 16-49 year olds from the 1999-2000 and 2003-2004 NHANES samples. Dashed lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals

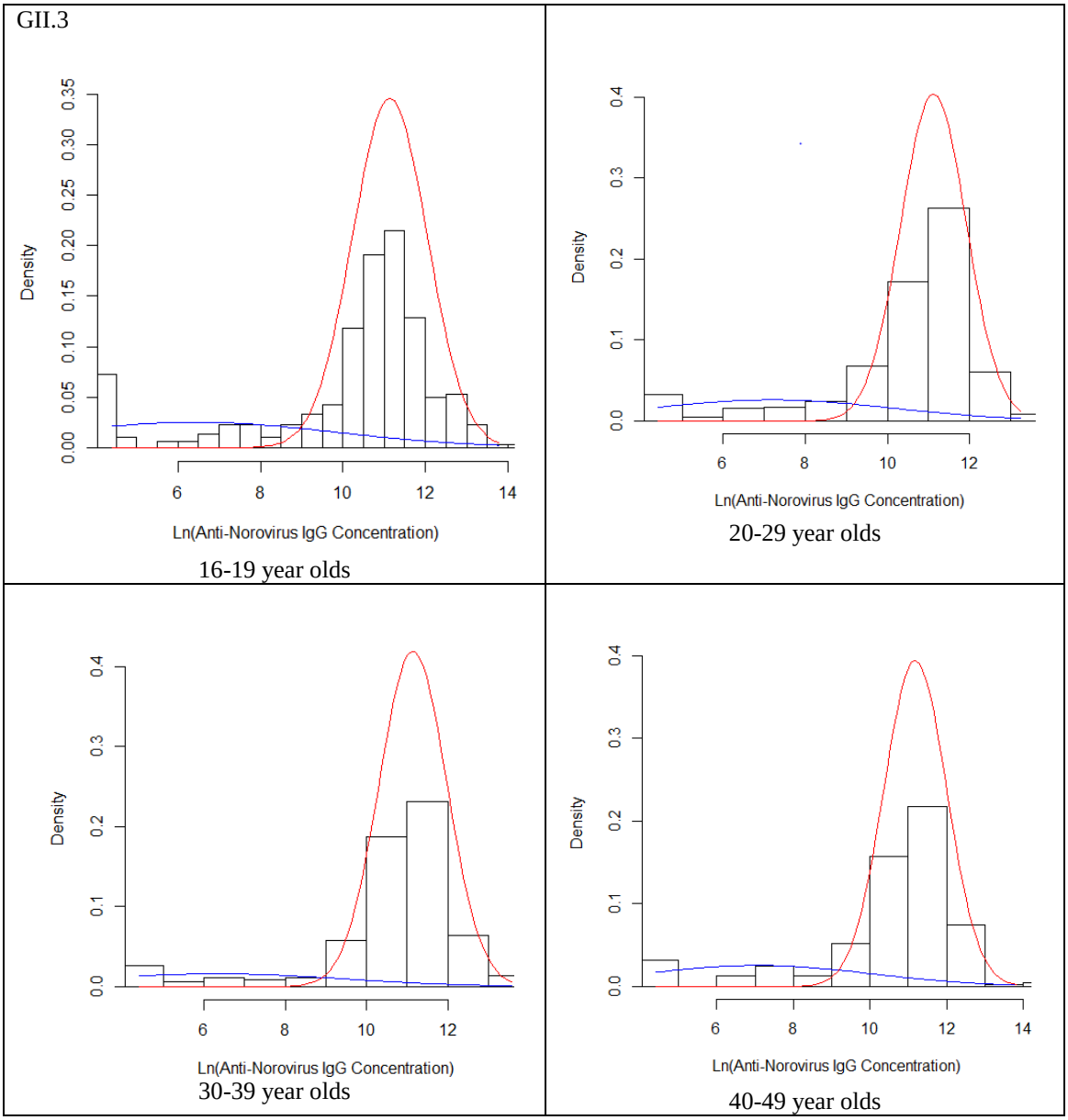


Figure 5- Finite mixture models of norovirus seropositivity in a 1/3 subset of 16-49 year olds from the 1999-2000 and 2003-2004 NHANES samples. Dashed lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals

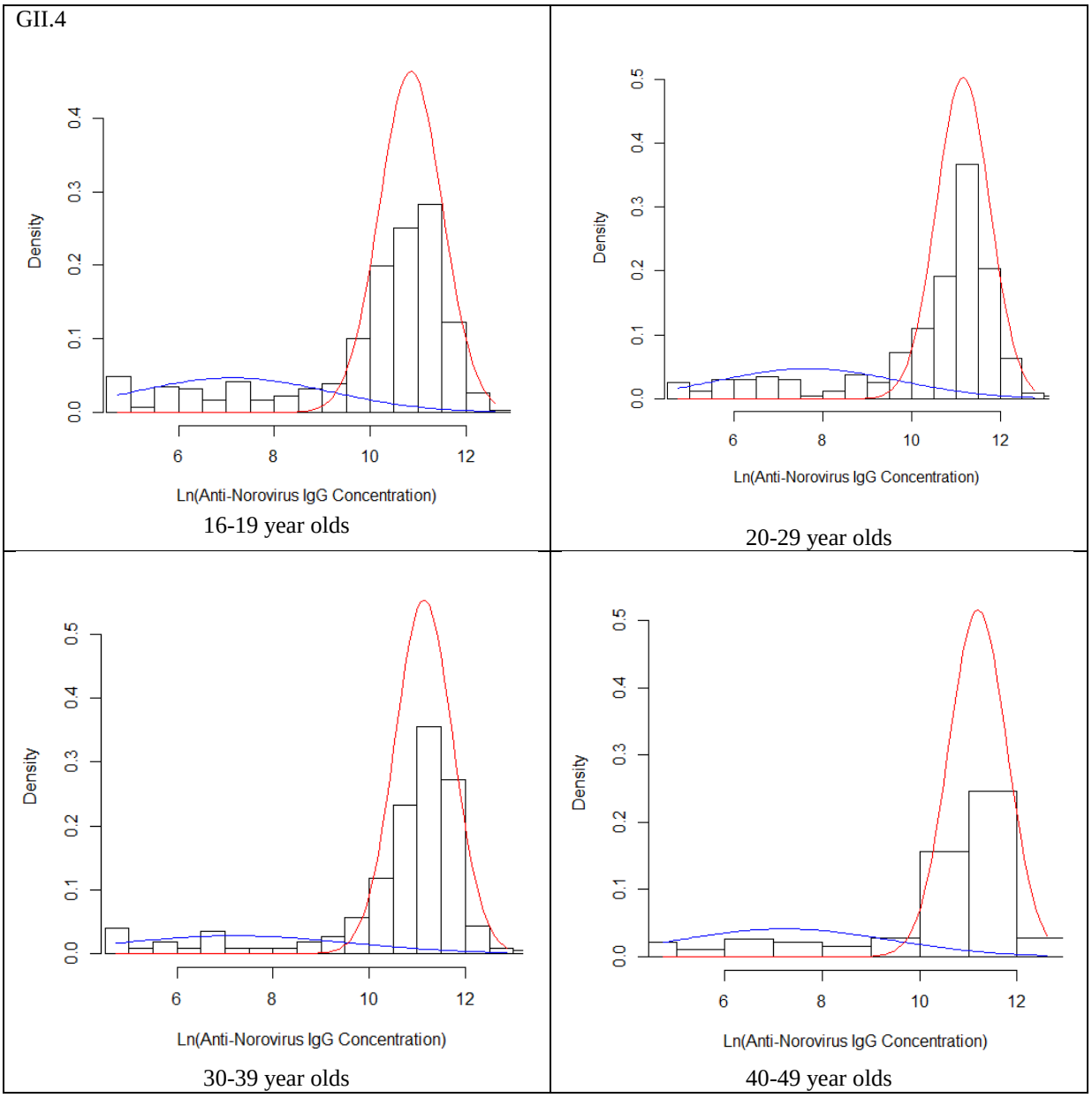


Figure 6- Finite mixture models of norovirus seropositivity in a 1/3 subset of 16-49 year olds from the 1999-2000 and 2003-2004 NHANES samples. Dashed lines represent the censoring threshold. Blue lines represent the estimated distribution of seronegative individuals. Red lines represent the estimated distribution of seropositive individuals

References

1. Scallan E, Hoekstra RM, Angulo FJ, et al. Foodborne Illness Acquired in the United States—Major Pathogens. *Emerg Infect Dis*. 2011;17(1):7-15. doi:10.3201/eid1701.P11101.

2. Mattner F, Sohr D, Heim A, Gastmeier P, Vennema H, Koopmans M. Risk groups for clinical complications of norovirus infections: an outbreak investigation. *Clin Microbiol Infect*. 2006;12(1):69-74. doi:10.1111/j.1469-0691.2005.01299.x.

3. Ramani S, Atmar RL, Estes MK. Epidemiology of human noroviruses and updates on vaccine development. *Curr Opin Gastroenterol*. 2014;30(1):25-33. doi:10.1097/MOG.0000000000000022.Epidemiology.

4. Huang P, Farkas T, Zhong W, et al. Norovirus and Histo-Blood Group Antigens : Demonstration of a Wide Spectrum of Strain Specificities and Classification of Two Major Binding Groups among Multiple Binding Patterns. *J Virol*. 2005;79(11):6714-6722. doi:10.1128/JVI.79.11.6714.

5. Kindberg E, Akerlind B, Johnsen C, et al. Host genetic resistance to symptomatic norovirus (GGII.4) infections in Denmark. *J Clin Microbiol*. 2007;45(8):2720-2722. doi:10.1128/JCM.00162-07.

6. Thorven M, Grahn A, Hedlund K-O, et al. A homozygous nonsense mutation (428G-->A) in the human secretor (FUT2) gene provides resistance to symptomatic norovirus (GGII) infections. *J Virol*. 2005;79(24):15351-15355. doi:10.1128/JVI.79.24.15351-15355.2005.

7. Carlsson B, Kindberg E, Buesa J, et al. The G428A nonsense mutation in FUT2 provides strong but not absolute protection against symptomatic GII.4 Norovirus infection. *PLoS One*. 2009;4(5):e5593. doi:10.1371/journal.pone.0005593.

8. Lindesmith L, Moe C, Lependu J, Frelinger J a, Treanor J, Baric RS. Cellular and humoral immunity following Snow Mountain virus challenge. *J Virol*. 2005;79(5):2900-2909. doi:10.1128/JVI.79.5.2900-2909.2005.

9. Nordgren J, Kindberg E, Lindgren PE, Matussek A, Svensson L. Norovirus gastroenteritis outbreak with a secretor-independent susceptibility pattern, Sweden. *Emerg Infect Dis*. 2010;16(1):81-87. doi:10.3201/eid1601.090633.

10. Rockx BHG, Vennema H, Hoebe CJP a, Duizer E, Koopmans MPG. Association of histo-blood group antigens and susceptibility to norovirus infections. *J Infect Dis*. 2005;191(5):749-754. doi:10.1086/427779.

11. Zheng D-P, Ando T, Fankhauser RL, Beard RS, Glass RI, Monroe SS. Norovirus classification and proposed strain nomenclature. *Virology*. 2006;346(2):312-323. doi:10.1016/j.virol.2005.11.015.

12. Hall A, Vinjé J, Lopman B, Park G. Updated norovirus outbreak management and disease prevention guidelines. *MMWR*. 2011;60(3). http://www.cdc.gov/mmwr/preview/mmwrhtml/rr6003a1.htm?s. Accessed November 12, 2014.

13. J. E. Matthews,, B. W. Dickey, R. D. Miller, J. R. Felzer, B. P. Dawson, A. S. Lee JJ, Rocks, J. Kiel1, J. S. Montes, C. L. Moe, J. N. S. Eisenberg and JSL. The epidemiology of published norovirus outbreaks: a systematic review of risk factors associated with attack rate and genogroup. *Epidemiol Infect*. 2013;140(7):1161-1172. doi:10.1017/S0950268812000234.The.

14. Bull R a, Eden J-S, Rawlinson WD, White P a. Rapid evolution of pandemic noroviruses of the GII.4 lineage. *PLoS Pathog*. 2010;6(3):e1000831. doi:10.1371/journal.ppat.1000831.

15. Debbink K, Lindesmith LC, Donaldson EF, Baric RS. Norovirus immunity and the great escape. *PLoS Pathog*. 2012;8(10):e1002921. doi:10.1371/journal.ppat.1002921.

16. Simmons KA. What is the duration of immunity to norovirus? A mathematical modeling study. *Emory Univ Electron Thesis Database*. 2012. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0. Accessed November 19, 2014.

17. Atmar RL, Opekun AR, Gilger M a, et al. Norwalk virus shedding after experimental human infection. *Emerg Infect Dis*. 2008;14(10):1553-1557. doi:10.3201/eid1410.080117.

18. Hall AJ, Rosenthal M, Gregoricus N, et al. Incidence of acute gastroenteritis and role of norovirus, Georgia, USA, 2004-2005. *Emerg Infect Dis*. 2011;17(8):1381-1388. doi:10.3201/eid1708.101533.

19. Gastañaduy P a, Hall AJ, Curns AT, Parashar UD, Lopman B a. Burden of norovirus gastroenteritis in the ambulatory setting--United States, 2001-2009. *J Infect Dis*. 2013;207(7):1058-1065. doi:10.1093/infdis/jis942.

20. Hall AJ, Curns AT, McDonald LC, Parashar UD, Lopman B a. The roles of Clostridium difficile and norovirus among gastroenteritis-associated deaths in the United States, 1999-2007. *Clin Infect Dis*. 2012;55(2):216-223. doi:10.1093/cid/cis386.

21. Nakata S, Honma S, Numata K. Prevalence of human calicivirus infections in Kenya as determined by enzyme immunoassays for three genogroups of the virus. *J Clin …*.

1998;36(11):3160-3163. http://jcm.asm.org/content/36/11/3160.short. Accessed November 22, 2014.

22. O'Ryan M, Víal P, Mamani N. Seroprevalence of Norwalk virus and Mexico virus in Chilean individuals: assessment of independent risk factors for antibody acquisition. *Clin Infect Dis.* 1998:789-795. http://cid.oxfordjournals.org/content/27/4/789.short. Accessed November 22, 2014.
23. Peasey AE, Ruiz-palacios GM, Quigley M, et al. Seroepidemiology and Risk Factors for Sporadic Norovirus/ Mexico strain. *J Infect Dis*. 2004;(189):2027-2036.
24. Smit T, Steele A, Peenze I. Study of Norwalk virus and Mexico virus infections at Ga-Rankuwa Hospital, Ga-Rankuwa, South Africa. *J Clin …*. 1997;35(9):2381-2385. http://jcm.asm.org/content/35/9/2381.short. Accessed November 23, 2014.
25. Hens N, Goeyvaerts N, Aerts M, Shkedy Z, Van Damme P, Beutels P. Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infect Dis*. 2009;9:5. doi:10.1186/1471-2334-9-5.
26. Lindesmith LC, Donaldson E, Leon J, et al. Heterotypic humoral and cellular immune responses following Norwalk virus infection. *J Virol*. 2010;84(4):1800-1815. doi:10.1128/JVI.02179-09.
27. Moe CL, Sair A, Lindesmith L, Estes MK, Jaykus L-A. Diagnosis of norwalk virus infection by indirect enzyme immunoassay detection of salivary antibodies to recombinant norwalk virus antigen. *Clin Diagn Lab Immunol*. 2004;11(6):1028-1034. doi:10.1128/CDLI.11.6.1028-1034.2004.

28. Verhoef L, Vennema H, van Pelt W, et al. Use of norovirus genotype profiles to differentiate origins of foodborne outbreaks. *Emerg Infect Dis*. 2010;16(4):617-624. doi:10.3201/eid1604.090723.

29. Hansman GS, Natori K, Shirato-Horikoshi H, et al. Genetic and antigenic diversity among noroviruses. *J Gen Virol*. 2006;87(Pt 4):909-919. doi:10.1099/vir.0.81532-0.

30. Vyse a J, Gay NJ, Hesketh LM, Pebody R, Morgan-Capner P, Miller E. Interpreting serological surveys using mixture models: the seroepidemiology of measles, mumps and rubella in England and Wales at the beginning of the 21st century. *Epidemiol Infect*. 2006;134(6):1303-1312. doi:10.1017/S0950268806006340.

31. Steens A, Waaijenborg S, Teunis PFM, et al. Age-dependent patterns of infection and severity explaining the low impact of 2009 influenza A (H1N1): evidence from serial serologic surveys in the Netherlands. *Am J Epidemiol*. 2011;174(11):1307-1315. doi:10.1093/aje/kwr245.

32. Leon JS, Kingsley DH, Montes JS, et al. Randomized, double-blinded clinical trial for human norovirus inactivation in oysters by high hydrostatic pressure processing. *Appl Environ Microbiol*. 2011;77(15):5476-5482. doi:10.1128/AEM.02801-10.

33. Seitz SR, Leon JS, Schwab KJ, et al. Norovirus infectivity in humans and persistence in water. *Appl Environ Microbiol*. 2011;77(19):6884-6888. doi:10.1128/AEM.05806-11.

34. Liu P, Yuen Y, Hsiao H-M, Jaykus L-A, Moe C. Effectiveness of liquid soap and hand sanitizer against Norwalk virus on contaminated hands. *Appl Environ Microbiol*. 2010;76(2):394-399. doi:10.1128/AEM.01729-09.

35. Herwaldt B, Lew J, Moe C. Characterization of a variant strain of Norwalk virus from a food-borne outbreak of gastroenteritis on a cruise ship in Hawaii. *J Clin …*.

1994;32(4):861-866. http://jcm.asm.org/content/32/4/861.short. Accessed November 23, 2014.

36. Moe C, Gentsch J, Ando T. Application of PCR to detect Norwalk virus in fecal specimens from outbreaks of gastroenteritis. *J Clin …*. 1994;32(3):642-648. http://jcm.asm.org/content/32/3/642.short. Accessed November 23, 2014.

37. Maple P a C, Gray J, Breuer J, Kafatos G, Parker S, Brown D. Performance of a time-resolved fluorescence immunoassay for measuring varicella-zoster virus immunoglobulin G levels in adults and comparison with commercial enzyme immunoassays and Merck glycoprotein enzyme immunoassay. *Clin Vaccine Immunol*. 2006;13(2):214-218. doi:10.1128/CVI.13.2.214-218.2006.

38. Debbink K, Costantini V, Swanstrom J, et al. Human norovirus detection and production, quantification, and storage of virus-like particles. *Curr Protoc Microbiol*. 2014;31:1-59. doi:10.1002/9780471729259.mc15k01s31.Human.

39. Blutt SE, Conner ME. The gastrointestinal frontier: IgA and viruses. *Front Immunol*. 2013;4(November):402. doi:10.3389/fimmu.2013.00402.

40. Franck KT, Fonager J, Ersbøll AK, Böttiger B. Norovirus epidemiology in community and health care settings and association with patient age, Denmark. *Emerg Infect Dis*. 2014;20(7):1123-1131. doi:10.3201/eid2007.130781.

41. Martino B Di, Profio F Di, Ceci C, et al. Seroprevalence of Norovirus Genogroup IV Antibodies among. *Emerg Infect Dis*. 2014;20(11):1828-1832.

42. Son H, Jeong HS, Cho M, et al. Seroepidemiology of predominant norovirus strains circulating in Korea by using recombinant virus-like particle antigens. *Foodborne Pathog Dis*. 2013;10(5):461-466. doi:10.1089/fpd.2012.1300.

43. Tan M, Jiang X. Vaccine against norovirus. *Hum Vaccin Immunother*. 2014;10(6):1449-1456. doi:10.4161/hv.28626.

44. Hoa Tran TN, 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*. 2013;56(3):185-193. doi:10.1016/j.jcv.2012.11.011.

45. Yu Y, Yan S, Li B, Pan Y, Wang Y. Genetic diversity and distribution of human norovirus in China (1999-2011). *Biomed Res Int*. 2014;2014:196169. doi:10.1155/2014/196169.

46. Kapikian A, Wyatt R, Dolin R. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *J Virol*. 1972;10(5):1075-1081. http://jvi.asm.org/content/10/5/1075.short. Accessed December 9, 2014.

47. Dolin R, Blacklow NR, DuPont H, et al. Biological Properties of Norwalk Agent of Acute Infectious Nonbacterial Gastroenteritis. *Exp Biol Med*. 1972;140(2):578-583. doi:10.3181/00379727-140-36508.

48. Kapikian A, Greenberg H, Cline W, et al. Prevalence of antibody to the Norwalk agent by a newly developed immune adherence hemagglutination assay. *J Med Virol*. 1978;2(4):281-294.

49. Prasad B, Hardy M, Dokland T, Bella J, Rossmann M, Estes M. X-ray crystallographic structure of the Norwalk virus capsid. *Science (80-)*. 1999;286(5438):287-290.

50. Graham DY, Jiang X, Tanaka T, Opekun AR, Madore H, Estes MK. Norwalk virus infection of volunteers: new insights based on improved assays. *J Infect Dis*. 1994;170(1):34-43.

51. Teunis PF, Moe C, Liu P, et al. Norwalk virus: how infectious is it? *J Med Virol*. 2008;80(8):1468-1476.

52. Lin CM, Wu FM, Kim HK, Doyle MP, Michael BS, Williams LK. A comparison of hand washing techniques to remove Escherichia coli and caliciviruses under natural or artificial fingernails. *J Food Prot*. 2003;66(12):2296-2301.

53. Kampf G, Grotheer D, Steinmann J. Efficacy of three ethanol-based hand rubs against feline calcivirus, a surrogate virus for norovirus. *J Hosp Infect*. 2005;60:144-149.

54. Eleraky NZ, Potgieter LN, Kennedy MA. Virucidal efficacy of four new disinfectants. *J Am Anim Hosp Assoc*. 2002;38(3):231-234.

55. Gulati BR, Allwood PB, Hedberg CW, Goyal SM. Efficacy of commonly used disinfectants for the inactivation of calicivirus on strawberry, lettuce, and a food-contact surface. *J Food Prot*. 2001;64(9):1430-1434.

56. Arias A, Emmott E, Vashist S, Goodfellow I. Europe PMC Funders Group Progress towards the prevention and treatment of norovirus infections. *Futur Microbiol*. 2013;8(11):1475-1487. doi:10.2217/fmb.13.109.Progress.

57. Atmar RL, Estes MK. Diagnosis of noncultivatable gastroenteritis viruses, the human caliciviruses. *Clin Microbiol Rev*. 2001;14(1):15-37. doi:10.1128/CMR.14.1.15-37.2001.

58. Desai R, Hembree CD, Handel A, et al. Severe outcomes are associated with genogroup 2 genotype 4 norovirus outbreaks: a systematic literature review. *Clin Infect Dis*. 2012;55(2):189-193. doi:10.1093/cid/cis372.

59. Domingo E, Quer EJ, Novella S, John J. Basic concepts in RNA virus evoution. *FASEB*. 1996;10:859-864.

60. Duizer E, Bijkerk P, Rockx B, De Groot A, Twisk F, Koopmans M. Inactivation of caliciviruses. *Appl Environ Microbiol*. 2004;70(8):4538-4543. doi:10.1128/AEM.70.8.4538-4543.2004.

61. Hutson AM, Atmar RL, Graham DY, Estes MK. Norwalk Virus Infection and Disease Is Associated with ABO Histo – Blood Group Type. *J Infect Dis*. 2002;185:1335-1337.

62. Leshem E, Barclay L, Wikswo M, et al. Genotype GI.6 norovirus, United States, 2010-2012. *Emerg Infect Dis*. 2013;19(8):1317-1320. doi:10.3201/eid1908.130445.

63. Lopman BA, Reacher MH, Vipond IB, Sarangi J, Brown DWG. Clinical Manifestation of Norovirus Gastroenteritis in Health Care Settings. *Clin Infect Dis*. 2004;39(July):318-324.

64. Lysén M, Thorhagen M, Brytting M, Hjertqvist M, Andersson Y, Hedlund K-O. Genetic diversity among food-borne and waterborne norovirus strains causing outbreaks in Sweden. *J Clin Microbiol*. 2009;47(8):2411-2418. doi:10.1128/JCM.02168-08.

65. Macinga DR, Sattar S a, Jaykus L-A, Arbogast JW. Improved inactivation of nonenveloped enteric viruses and their surrogates by a novel alcohol-based hand sanitizer. *Appl Environ Microbiol*. 2008;74(16):5047-5052. doi:10.1128/AEM.00487-08.

66. Martella V, Campolo M, Lorusso E, et al. Norovirus in Captive Lion Cub. *Emerg Infect Dis*. 2007;13(7):1071-1073.

67. Matthews JE, Dickey BW, Miller RD, et al. The epidemiology of published norovirus outbreaks: a systematic review of risk factors associated with attack rate and genogroup. *Epidemiol Infect*. 2013;140(7):1161-1172. doi:10.1017/S0950268812000234.The.

68. Donnell GMC, Russell A. Antiseptics and Disinfectants : Activity , Action , and Resistance. *Clin Microbiol Rev*. 1999;12(1):147-179.

69. Mesquita JR, Barclay L, Nascimento MSJ, Vinjé J. Novel norovirus in dogs with diarrhea. *Emerg Infect Dis*. 2010;16(6):980-982. doi:10.3201/eid1606.091861.

70. Siebenga JJ, Vennema H, Renckens B, et al. Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. *J Virol*. 2007;81(18):9932-9941. doi:10.1128/JVI.00674-07.

71. Smiley JR, Chang KO, Hayes J, et al. Characterization of an Enteropathogenic Bovine Calicivirus Representing a Potentially New Calicivirus Genus. *J Virol*. 2002;76(20):10089-10098. doi:10.1128/JVI.76.20.10089.

72. Poel WHM Van Der, Vinjé J, Heide R Van Der. Norwalk-Like Calicivirus Genes in Farm Animals. *Emerg Infect Dis*. 2000;6(1):36-41.

73. Tan M, Jiang X. The p domain of norovirus capsid protein forms a subviral particle that binds to histo-blood group antigen receptors. *J Virol*. 2005;79(22):14017-14030. doi:10.1128/JVI.79.22.14017-14030.2005.

74. Wobus CE, Karst SM, Thackray LB, et al. Replication of Norovirus in cell culture reveals a tropism for dendritic cells and macrophages. *PLoS Biol*. 2004;2(12):e432. doi:10.1371/journal.pbio.0020432.

75. Hasing ME, Lee BE, Preiksaitis JK, 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. *J Clin Microbiol*. 2013;51(7):2204-2211. doi:10.1128/JCM.00663-13.

76. Payne DC, Vinjé J, Szilagyi PG, et al. Norovirus and medically attended gastroenteritis in U.S. children. *N Engl J Med*. 2013;368(12):1121-1130. doi:10.1056/NEJMsa1206589.

77. Mead PS, Slutsker L, Dietz V, et al. Food-Related Illness and Death in the United States. *Emerg Infect Dis*. 1999;5(5):607-625.

78. Yen C, Hall a. J. Editorial Commentary: Challenges to Estimating Norovirus Disease Burden. *J Pediatric Infect Dis Soc*. 2013;2(1):61-62. doi:10.1093/jpids/pis134.

79. Hall AJ, Lopman B a, Payne DC, et al. Norovirus disease in the United States. *Emerg Infect Dis*. 2013;19(8):1198-1205. doi:10.3201/eid1908.130465.

80. Lindesmith LC, Beltramello M, Donaldson EF, et al. Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. *PLoS Pathog*. 2012;8(5):e1002705. doi:10.1371/journal.ppat.1002705.

81. Zheng D-P, Zhang LB, Fang ZY, et al. Distribuion of wild type 1 poliovirus genotypes in China. *J Infect Dis*. 1993;(186):1361-1367.

82. Black R, Greenberg H, Kapikian A, Brown K, Becker S. Acquisition of serum antibody to Norwalk Virus and rotavirus and relation to diarrhea in a longitudinal study of young children in rural Bangladesh. *1J Infect Dis*. 982;145(4):483-489.

83. Patel MM, Widdowson M-A, Glass RI, Akazawa K, Vinjé J, Parashar UD. Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerg Infect Dis*. 2008;14(8):1224-1231. doi:10.3201/eid1408.071114.

84. Jiang XI, Wang MIN, Graham DY, Estes MK. Expression , Self-Assembly , and Antigenicity of the Norwalk Virus Capsid Protein. *J Virol*. 1992;66(11):6527-6532.

85. Hall AJ, Wikswo ME, Manikonda K, Roberts VA, Yoder JS, Gould LH. Acute Gastroenteritis Surveillance through the National Outbreak Reporting System, United States. *Emerg Infect Dis*. 2013;19(8):1305-1309.

86. Bartsch SM, Lopman BA, Hall AJ, Parashar D, Lee BY. The Potential Economic Value of a Human Norovirus Vaccine for the United States. *Vaccine*. 2013;30(49):7097-7104. doi:10.1016/j.vaccine.2012.09.040.The.

87. Saito M, Goel-Apaza S, Espetia S, et al. Multiple norovirus infections in a birth cohort in a Peruvian Periurban community. *Clin Infect Dis*. 2014;58(4):483-491. doi:10.1093/cid/cit763.

88. Johnson PC, Mathewson JJ, Dupont HL, Greenberg HB. Multiple-Challenge Study of Host Susceptibility to Norwalk Gastroenteritis in US Adults. *J Infect Dis*. 1990;161(1):18-21.

89. Kirby A, Shi J, Montes J, Lichtenstein M, Moe C. Disease course and viral shedding in experimental Norwalk virus and Snow Mountain virus infection. *J Med …*. 2014;2064(86):2055-2064. doi:10.1002/jmv.

90. Parrino T, Schreiber D, Trier J, Kapikan A, Blacklow N.. *N Engl J Med*. 1977;297(2):86-89.

91. Gary GW, Anderson LJ, Keswick BH, et al. Norwalk Virus Antigen and Antibody Response in an Adult Volunteer Study. *J Clin Microbiol*. 1987;25(10):2001-2003.

92. Cubitt WD, Green KY, Payment P. Prevalence of Antibodies to the Hawaii Strain of Human Calicivirus as Measured by a Recombinant Protein Based Immunoassay. *J Med Virol*. 1998;139(October 1997):135-139.

93. Smit TK, Bos P, Peenze I, Jiang X, Estes MK, Steele AD. Seroepidemiological Study of Genogroup I and II Calicivirus Infections in South and Southern Africa. *J Med Virol*. 1999;231(January):227-231.

94. Pelosi E, Lambden PR, Caul EO, et al. The Seroepidemiology of Genogroup 1 and Genogroup 2 Norwalk-Like Viruses in Italy. *J Med Virol*. 1999;58(September 1998):93-99.

95. Dimitrov DH, Dashti SAH, Ball JM, et al. Prevalence of Antibodies to Human Caliciviruses (HuCVs) in Kuwait Established by ELISA Using Baculovirus-Expressed Capsid Antigens Representing Two Genogroups of HuCVs. *J Med Virol*. 1997;51(July 1996):115-118.

96. Jing Y, Qian Y, Huo Y, Wang L, Jiang X. Seroprevalence Against Norwalk-Like Human Caliciviruses in Beijing , China. *J Med Virol*. 2000;60(March 1999):97-101.

97. Parker SP, Cubitt WD, Jiang X. Enzyme immunoassay using baculovirus-expressed human calicivirus (Mexico) for the measurement of IgG responses and determining its seroprevalence in London, UK. *J Med Virol*. 1995;46(3):194-200. doi:10.1002/jmv.1890460305.

98. Hens N, Goeyvaerts N, Aerts M, Shkedy Z, Van Damme P, Beutels P. Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infect Dis*. 2009;9(9):5. doi:10.1186/1471-2334-9-5.

99. Ribes JM, Te CJ, Carmona-vicente N, Ferna M. Norovirus Infections and Seroprevalence of Genotype GII . 4-Specific Antibodies in a Spanish Population. *J Med Virol*. 2015;682(October 2014):675-682. doi:10.1002/jmv.

100. Liu P, Wang X, Lee J-C, et al. Genetic susceptibility to norovirus GII.3 and GII.4 infections in Chinese pediatric diarrheal disease. *Pediatr Infect Dis J*. 2014;33(11):e305-9. doi:10.1097/INF.0000000000000443.

101. Kobayashi S, Fujiwara N, Takeda N, Minagawa H. Seroepidemiological study of norovirus infection in Aichi Prefecture, Japan. *Microbiol Immunol*. 2009;53(6):356-359. doi:10.1111/j.1348-0421.2009.00132.x.

102. Eden J-S, Tanaka MM, Boni MF, Rawlinson WD, White P a. Recombination within the pandemic norovirus GII.4 lineage. *J Virol*. 2013;87(11):6270-6282. doi:10.1128/JVI.03464-12.

103. Noel JS, Fankhauser RL, Ando T, Monroe SS, Glass RI. Identification of a Distinct Common Strain of " Norwalk-like Viruses " Having a Global Distribution . 1997;90(January 1996):1334-1344.