University of Cincinnati

Date: 4/24/2014

I, Rebecca L Currier, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Epidemiology (Environmental Health).

It is entitled:
Impact of FUT2 Genotype on National Pediatric Population Burden of Norovirus-Associated Acute Gastroenteritis

Student’s name: Rebecca L Currier

This work and its defense approved by:

Committee chair: Ardythe Morrow, Ph.D.

Committee member: Sean Moore, M.S., M.D.

Committee member: Kim Dietrich, Ph.D.

Committee member: Mary Staat, M.D.

Committee member: Changchun Xie, Ph.D.
Impact of $FUT2$ Genotype on National Pediatric Population Burden of Norovirus-Associated Acute Gastroenteritis

A dissertation submitted to the Graduate School of the University of Cincinnati in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in the Division of Epidemiology and Biostatistics of the Department of Environmental Health of the College of Medicine

by Rebecca Lynn Currier B.S. Louisiana Tech University April 24, 2014

Committee Chair: Ardythe Luxion Morrow, Ph.D.
Abstract

BACKGROUND: Norovirus is a leading cause of acute gastroenteritis (AGE). Noroviruses bind histo-blood group antigens (HBGAs) on the gut surface, but only 70-80% of individuals have a functional copy of the FUT2 gene required for gut HBGA expression. We hypothesized that “secretors”, individuals with at least one active FUT2 allele, are more likely to have norovirus infections than “non-secretors”.

METHODS: From 12/2011 to 11/2012, active AGE surveillance was performed through the CDC New Vaccine Surveillance Network, comprised of six geographically representative US pediatric sites. Cases were recruited from emergency departments and inpatient units; healthy controls were recruited at well-child visits. Salivary DNA was collected to determine FUT2 secretor genotype and genetic ancestry. Stool specimens were analyzed for norovirus by realtime RT-PCR and PCR products were sequenced to genotype the norovirus strain.

RESULTS: Norovirus was detected in 302 (21%) of 1465 AGE cases and 52 (6%) of 826 healthy controls. Norovirus AGE cases were 2.78-fold more likely to be secretors versus controls (p<0.001) in a logistic regression model adjusted for race/ethnic background, age, site, and health insurance. Secretors comprised all 155 cases and 21 asymptomatic infections positive for norovirus GII.4 strains. Among norovirus-negative controls, FUT2 secretors comprised 74% of non-Hispanics and 86% of Hispanics (p=0.004).

CONCLUSIONS: FUT2 genotype is associated with norovirus risk and also varies with ethnicity. During 2011-2012 surveillance, the dominant circulating norovirus, GII.4, exclusively infected secretors. These findings are important to the development of norovirus vaccines. HBGAs produced by FUT2 may also provide a therapeutic target.
Acknowledgements

I would like to thank the excellent faculty of the Division of Epidemiology and Biostatistics at the University of Cincinnati. In particular I’d like to thank Dr. Ardythe Morrow, my advisor and mentor, who has gone far beyond the call of duty on too many occasions to count. Thank you to Dr. Tom DeWitt for helping me to find my home in epidemiology and for introducing me to Dr. Morrow. Many thanks to the members of my qualifying committee: Drs. Ardythe Morrow, Mary Staat, and Paul Succop. Thank you in particular to my dissertation committee: Drs. Ardythe Morrow, Kim Dietrich, Sean Moore, Mary Staat, and Changchun Xie.

The saliva collection kits and human resources support to add DNA collection to national AGE surveillance from 2011-2012 came from Dr. Morrow’s program project (NIH P01 HD 13021). My support as a graduate student has come primarily from the Molecular Epidemiology in Children’s Environmental Health (NIH T32 ES 010957) training program. Special thanks to Drs. Ranjan Deka and Kim Dietrich for their support of my fellowship. This research was also supported by the University of Cincinnati Graduate Student Governance Association (GSGA) Research Fellowship, as well as three additional research/travel awards from the GSGA.

I am indebted to the Medical Scientist Training Program (NIH T32 GM 063483) for the incredible dual training opportunity in medicine and epidemiology. This program has been led by Drs. Leslie Myatt, Patrick Tso, and Neeru Hershey in my six years at UC. Dr. George Deepe and Ms. Laurie Mayleben have been invaluable in my transitions between medical and graduate school. Dr. Samir Shah has given excellent advice as my Senior Physician Scientist mentor.

National AGE surveillance was funded by the CDC/New Vaccine Surveillance Network (NVSN). This research embodied the concept of “team science.” I have been privileged in the last three years to be a part of the NVSN team working with literally hundreds of scientists,
statisticians, data managers, and clinical personnel. Special thanks to Dr. Daniel Payne (team leader), Mary Wikso (data support), Dr. Jan Vinjé (head of calicivirus lab) and Dr. Mary Staat (principal investigator, Cincinnati). This work would not have been possible without the principal investigators at each site and the “boots on the ground” clinical research personnel who rolled up their sleeves to take saliva samples from thousands of children. I wish I could acknowledge each of you personally.

This work was also supported by a large number of laboratory staff. Thank you in particular to Myra Johnson Willwerth, laboratory technician extraordinaire, whose organization, personal dedication, and impressive sense of the “big picture” smoothed out so many bumps in the road. Special recognition to Sarah Hannah Shirley at CDC for completing thousands of stool sample analyses under immense time pressure and a government shutdown.

Fellow graduate students Diana Taft, Olivia Ballard, and Marguerite Schneider provided laboratory support, advice, proofreading, and have been truly wonderful comrades-in-arms. Additional laboratory/administrative support came from Barbara Davidson, Donna Wuest, Aimee LeFevres, Brett Schibler, and Tyler Woods. Matt Leonard provided helpful database management advice. Drs. Sara Lazaros-Palacios and Leah Kottyan were exceptionally generous with their time advising me on the intricacies of microarray analysis.

Thank you for the support from my incredible family and friends. My parents (Paula and John Currier) and sister (Sarah Currier) have been constant cheerleaders, proofreaders, and general providers of unconditional love and bad jokes.

Finally, thank you to my wonderful grandparents Velma Sibille, Winfred Sibille, and John Currier, who never stopped asking “So, when are you going to graduate?” I love you.
# Table of Contents

Abstract .............................................................................................................................. ii

Acknowledgements ........................................................................................................... iv

Table of Tables .................................................................................................................. viii

Table of Figures ................................................................................................................ ix

Section 1: Introduction, Specific Aims, and Significance .................................................... 1

Section 2: Research Manuscript ........................................................................................ 6

Abstract ............................................................................................................................. 6

Introduction ....................................................................................................................... 7

Methods ............................................................................................................................. 8

Results ............................................................................................................................... 11

Discussion .......................................................................................................................... 13

Section 3: Genotype Analysis Methods .............................................................................. 17

Microarray data cleaning ................................................................................................. 17

TaqMan Analysis .............................................................................................................. 18

Section 4: Ancestry Classification ...................................................................................... 19

Self-identified ancestry ...................................................................................................... 19

Available genetic ancestry information ........................................................................... 21

Principal Components Analysis (PCA) ............................................................................ 21

STRUCTURE analysis ...................................................................................................... 22
Benefits to subjects................................................................. 131

Importance of knowledge to be gained......................................... 131

Table of Tables

Table 1. Summary of secretor/norovirus literature published in previous decade .......... 43
Table 2. Demographics of norovirus AGE cases and healthy control study completers.. 44
Table 3. Relationship of secretor status to norovirus infection ................................ 45
Table 4. Variation in ImmunoChip quality control protocols in published literature ..... 46
Table 5. Self-identified race/ethnicity of subjects with stool and saliva data.............. 47
Table 6. Secretor distribution in controls using four self-identified categories............. 48
Table 7. Morrow data sensitivity to outlier definition ........................................... 49
Table 8. Comparison of self-identified ancestry to STRUCTURE identified ancestry... 50
Table 9. STRUCTURE handling of admixed subjects ........................................... 51
Table 10. Secretor status of potential control groups .......................................... 52
Table 11. Study non-completion of eligible children during surveillance hours........... 53
Table 12. Demographic differences in eligible non-completers vs. study completers .... 54
Table 13. Association of secretor status with potential confounders.......................... 55
Table 14. Reasons for non-completion of eligible children during surveillance hours, by study site ............................................................................................................. 56
# Table of Figures

- Figure 1. Flow diagram of subject participation .......................................................... 57
- Figure 2. FUT2 secretors in norovirus-free controls and norovirus AGE cases, by norovirus genotype .......................................................... 58
- Figure 3. PCA of Morrow data set using AIMs ................................................................. 60
- Figure 4. Triangle plot of STRUCTURE results ............................................................... 61
- Figure 5. Distribution of self-identified ancestry in AGE cases and controls ............. 62
- Figure 6. Distribution of final genetic ancestry classifications in AGE cases and controls .......................................................... 63
- Figure 7. Differences in sample collection by study site .............................................. 64
Section 1: Introduction, Specific Aims, and Significance

Each year an estimated 5.5 million Americans suffer from norovirus (NoV)-induced acute gastroenteritis (AGE) (1). Since the introduction of the rotavirus vaccine, NoV has become the leading cause of childhood diarrhea, responsible for 20% of medical visits for AGE (2), and 58% of domestically-acquired foodborne illnesses (1). A highly infectious virus, NoV is known in popular culture as the cause of AGE outbreaks on cruise ships (3). In 2013, NoV AGE increased 52% due to the emergence of the novel GII.4 Sydney strain (4). Its high transmissibility makes NoV an important public health concern, particularly in closed settings such as long-term care facilities, military barracks, and day care centers. Children under five years are particularly vulnerable, with an incidence four times higher than the general population (5).

For decades it was noted that some individuals are unusually resistant to symptomatic NoV infection (6). In vitro studies then revealed that NoVs bind to human histo-blood group antigens (HBGAs) in a strain-specific manner (7). The expression of HBGAs at the surface of the gut is largely controlled by the FUT2 (secretor) gene. This gene is inactive in 25% of individuals, known as “non-secretors.” Most non-secretors in the U.S. have a homozygous nonsense mutation at the 428 position in FUT2 (8), hypothesized to give innate protection against several common strains of NoV. A high-dose challenge study using a strain of the NoV genogroup GII.4 was recently published by investigators at Cincinnati Children’s. Of the 40 adults enrolled in the challenge study, 70% of secretors and 6% of non-secretors developed AGE after direct challenge (p<0.01) (9). Studies of NoV outbreak investigations show that some NoV strains attack secretors nearly exclusively (10, 11), but others display no such specificity (12).
The current literature supports the importance of secretor genetics as a key determinant of risk in challenge studies and certain NoV outbreaks (see Table 1). However, population-based studies are lacking to define the impact of secretor genetics on the population burden of NoV AGE. Furthermore, most outbreak studies do not address the relationship of secretor genetics to asymptomatic shedding of NoV, a major cause of outbreak spread (13).

NoV vaccines are currently under development (14, 15). As NoV becomes the focus of efforts to reduce childhood AGE, it is important to determine the extent to which non-secretor individuals are innately resistant to this disease, which has public health implications for vaccination and outbreak control strategies (6).

The purpose this dissertation research is to evaluate the impact of FUT2 secretor genetics on the national pediatric population’s burden of symptomatic and asymptomatic NoV infection. This research was designed to test the hypothesis that secretor genetics determines the pediatric population’s risk of NoV-positive AGE and the shedding of NoV in asymptomatic children. This dissertation addresses two specific aims:

- **Aim 1.** Determine the association between FUT2 (secretor) genotype and risk of pediatric NoV AGE detected using a national surveillance network.
- **Aim 2.** Compare asymptomatic shedding of NoV GII.4 in secretors versus non-secretors.

Acute gastroenteritis (AGE) causes significant global morbidity and mortality: 800,000 pediatric deaths occur each year due to AGE (16), with at least 200,000 deaths caused by NoV (15). The U.S. experiences 11,255 deaths per year directly attributable to gastroenteritis, with NoV as a leading infectious cause (17). NoV is the leading cause of AGE outbreaks worldwide and the lead cause of pediatric AGE in the U.S (2, 18). In the U.S., NoV infections result in 400,000 emergency department (ED) visits and 1.7 million office visits, resulting in $284 million
in direct health care costs (5). Children under five have a particularly high incidence of NoV AGE, accounting for nearly 1 million health care visits per year. By age five, 1 in 6 U.S. children have had an outpatient visit, 1 in 14 have been seen in the ED, and 1 in 278 have been hospitalized due to NoV AGE (2). Due to this substantial morbidity, NoV is considered a prime vaccine candidate, and several vaccines are in development (14, 15).

Aim 1 provides critical information on how the secretor-pathogen relationship impacts the U.S. pediatric population’s susceptibility to current circulating NoV strains. If HBGAs are shown to be important in population NoV infections, possibilities arise for research into medical food therapeutics that provide a soluble receptor decoy for the HBGA receptors that the FUT2 enzyme produces on mucosal surfaces. These receptor decoys could protect from infection in much the same way the non-nutritive carbohydrates in human milk are hypothesized to protect from some pathogens (19).

The relationship of secretor status to NoV susceptibility also directly impacts the design of vaccine trials, and is significant for future vaccination policy. The Cincinnati Children’s Hospital Medical Center’s Vaccine Treatment and Evaluation Unit (CCHMC VTEU) recently demonstrated that non-secretors are highly resistant to the GII.4 strain used in our current NoV challenge studies (9). These results have led to the decision to exclude non-secretors from further dose-response challenge studies except as negative controls (personal communication with Dr. Robert Frenck, PI). The results of this research have the potential to directly impact the choice of study population in the design of future vaccine trials and public health policy regarding vaccine deployment.

Differences in population baseline susceptibility are particularly important since not all populations have the same distribution of secretors and non-secretors. Numerous FUT2
polymorphisms have been described which inactivate or decrease the activity of the secretor gene. The presence of these polymorphisms produces the non-secretor phenotype in 20-30% of individuals in most ancestral populations.(8) A notable exception are Hispanic-Americans of Meso-American descent, in whom the vast majority are secretors (20), (personal communication, Guillermo Ruiz-Palacios). In both the Caucasian and African American populations, the non-secretor phenotype nearly exclusively produced by a G428A nonsense mutation resulting in a premature stop codon and a therefore truncated, nonfunctional protein.(8)

NoV is highly infectious, with an infectious dose of only 10 viral particles (21). Its most common mode of transmission is person-to-person contact, with asymptomatic shedding believed to be particularly important in outbreak spread (22). Asymptomatic food handlers in particular are a source of outbreaks with as many as 12% testing positive for NoV (23). Currently, little is known about the relationship between secretor status and asymptomatic shedding. A single study has addressed this question in a Nicaraguan population. The study population had a low (<5%) occurrence of the non-secretor genotype and no detection of GII.4 (24), limiting its generalizability to NoV circulation in the United States, which has nearly 25% non-secretor genotype and predominant circulation of NoV GII.4. Aim 2 evaluates the secretor/carrier relationship in a national sample of healthy control children. This research will provide critical information on the extent to which non-secretors, resistant to symptomatic infection by some NoV strains, are still capable of spreading the virus to susceptible individuals.

In summary, previous work has shown that secretor status predisposes to susceptibility to symptomatic AGE caused by many, but not all, NoV strains. The relevance of this association to the overall population burden of AGE has not been tested due to the logistical limitations; the overwhelming majority of published secretor-NoV clinical studies are retrospective studies of
specific outbreaks. To test this association, a large-scale effort has been needed to identify AGE cases across the U.S.

Such a large-scale effort exists in the form of the CDC’s New Vaccine Surveillance Network (NVSN). Partnership with the NVSN for this study leverages the resources of an established national research group with systems in place to collect detailed demographic and clinical information on pediatric AGE cases and perform pathogen testing. The addition of salivary DNA collection to NVSN’s protocols during the 2011-2012 AGE surveillance system made this dissertation possible by adding genomic data to the NVSN’s description of pediatric AGE.
Section 2: Research Manuscript

At the time of submission of this dissertation, the manuscript was in preparation.

Title: Innate susceptibility to norovirus infections influenced by FUT2 genotype in a United States pediatric population

Abstract

Background: Norovirus is a leading cause of acute gastroenteritis (AGE). Noroviruses bind to histo-blood group antigens (HBGAs) on the gut surface, but only 70-80% of individuals have a functional copy of the FUT2 (“secretor”) gene required for gut HBGA expression; these individuals are known as “secretors.” Susceptibility to some noroviruses depends on FUT2 secretor status, but the population-wide impact of this association is not established. Methods: From 12/2011 to 11/2012, active AGE surveillance was performed through the CDC New Vaccine Surveillance Network, comprised of six geographically representative US pediatric sites. Cases were recruited from emergency departments and inpatient units; healthy controls were recruited at well-child visits. Salivary DNA was collected to determine secretor status and genetic ancestry. Stool specimens were tested for norovirus by realtime RT-PCR; positive samples were sequenced to determine norovirus genotype. Results: Norovirus was detected in 302 (21%) of 1465 AGE cases and 52 (6%) of 826 healthy controls. Norovirus AGE cases were 2.8-fold more likely than norovirus-negative controls to be secretors (p<0.001) in a logistic regression model adjusted for race/ethnicity, age, site, and health insurance. Secretors comprised all 155 cases and 21 asymptomatic GII.4 norovirus infections. Hispanics of Meso-American ancestry were more likely than others to be secretors (in controls, 96% versus 74%, p<0.001).
Conclusions: FUT2 secretor status is associated with norovirus infection and varies by ethnicity. GII.4 norovirus exclusively infected secretors. These findings are important to the design and interpretation of norovirus vaccine trials and the design of agents that may block norovirus-HBGA binding.

Introduction

Each year 19-21 million Americans suffer from acute gastroenteritis (AGE) due to norovirus infection.(25) The illness is characterized by sudden onset of severe vomiting and dehydrating diarrhea lasting 1-3 days.(26) Its high transmissibility makes norovirus an important public health concern, particularly in closed settings such as long-term care facilities, schools and child care centers, and military barracks.(26, 27) Medically-attended norovirus AGE is most common in children under five, and in areas with limited healthcare resources the illness causes 200,000 pediatric deaths each year.(15, 25, 28) Norovirus is now the leading cause of childhood diarrhea in populations with high rotavirus vaccination coverage.(28)

Human susceptibility to norovirus infection is thought to be at least partially dependent upon the FUT2 gene.(9-12, 20, 29-34) FUT2 controls the secretion of ABO histo-blood group antigens (HBGAs) at the gut surface. ABO carbohydrates have been identified as binding ligands and presumed receptors for caliciviruses, the virus family which includes norovirus.(30) However, in vitro and in silico studies have shown that noroviruses bind to HBGAs in a strain-specific manner.(31, 32)

At least one functional FUT2 allele is present in 70-80% of individuals, who are referred to as FUT2 “secretors.”(8) FUT2 is inactivated in the remaining 20-30% (“non-secretors”) by homozygous 428G>A nonsense mutations in European and African populations, or less common
variants such as the 385A>T missense mutation found in Asian populations.(8) Hispanic individuals of Meso-American descent rarely carry these inactivating mutations.(8, 20)

In challenge studies, FUT2 non-secretors have demonstrated resistance to specific noroviruses, developing neither symptoms nor antibody response to norovirus genotypes GI.1 or GII.4.(9, 33, 35) In epidemiologic studies, FUT2 secretors have demonstrated significantly greater susceptibility to symptomatic infection by some norovirus genotypes (GI.1, GII.3, GII.4)(10, 11, 34, 35) but not others (GI.3, GII.3).(12, 34)

No studies have been conducted in the US to define the impact of secretor genetics on the overall risk of sporadic norovirus AGE. Furthermore, secretor genetics in relation to asymptomatic shedding of norovirus has not been studied in populations in which the non-secretor genotype is common. Finally, while in vitro work suggests that A and B blood group antigens may modify risk of norovirus infection in secretor individuals,(31) this question has not been well addressed in epidemiologic studies.

With norovirus vaccines currently in development(36), understanding population patterns of susceptibility to this pathogen is crucial. Our study utilizes a large, prospective, geographically diverse AGE surveillance network to identify sporadic pediatric norovirus infections. We use these surveillance data to determine the relationship of FUT2 secretor status to symptomatic and asymptomatic norovirus infection, modification of that relationship by non-O blood group types, and to describe FUT2 secretor status in the U.S. by race and ethnicity.

Methods

Active surveillance

The New Vaccine Surveillance Network (NVSN) performs active surveillance of pediatric diseases, coordinated by the Centers for Disease Control and Prevention (CDC). For
this study, enrollment occurred during the 12-month period of December 2011 through November 2012 at the University of Rochester Medical Center (Monroe County, New York), Vanderbilt University Medical Center (Davidson County, Tennessee), Cincinnati Children’s Hospital Medical Center (Hamilton County, Ohio), Texas Children’s Hospital (Harris County, Texas), Children’s Mercy Hospital and Clinics (Kansas City, Missouri), and Seattle Children’s Hospital (King County, Washington). Approval was obtained from the institutional review boards at each site and the CDC.

Surveillance methods have been described previously. (28, 37, 38) Briefly, prospective active surveillance was conducted in the emergency departments and inpatient units of NVSN institutions. Eligible AGE cases were between 14 days and 5 years of age with AGE symptoms of less than ten days duration. AGE symptoms were defined as diarrhea ≥3 episodes or vomiting ≥1 episodes within 24 hours. Children were excluded if they had a noninfectious cause of diarrhea or vomiting, immunodeficiency, or had transferred from another hospital.

Healthy controls between 14 days and 5 years of age were enrolled during scheduled well-child visits at affiliated clinics. They were excluded if they had symptoms of AGE within 14 days of enrollment, immunodeficiency, or symptoms of acute respiratory infection within 3 days of enrollment. The enrollment of healthy controls was frequency-matched to the enrollment of AGE cases at each site by age and calendar month.

Written, informed consent was obtained in English at all sites, or in Spanish at the Kansas City, Seattle or Houston sites. Eligible children were not enrolled if the consenting adult did not speak an available language. Demographic and clinical data were systematically collected from all subjects. Whole saliva was collected from subjects at enrollment and stored in DNA stabilization buffer (Oragene·Discover for assisted collection, DNA Genotek). Whole stool
samples were obtained from both cases and controls, with 98% of cases providing stool within 7 days of enrollment. Stool samples were tested for GI and GII norovirus by real-time RT-PCR assay(39) that included an MS2 coliphage extraction control followed by amplification of the positive samples by conventional RT-PCR and sequencing. Noroviruses were genotyped by comparing the sequences to a CaliciNet database of norovirus prototype strains.(39)

**Human genotyping and ancestry determination**

Human DNA from saliva was genotyped using both RT-PCR and the ImmunoChip (Illumina Infinium), which has been described previously.(40) The ImmunoChip was selected for its coverage of the *FUT2* and *ABO* locus as well as its inclusion of genetic ancestral informative markers (AIMs).

Samples were prepared for ImmunoChip analysis using Illumina protocols. Only samples and single nucleotide polymorphisms (SNPs) with call rates above 95% were retained. Samples were checked for agreement with self-identified gender and autosomal heterozygosity <0.4.

O/non-O blood type was determined using SNP rs505922, which acts as a surrogate for the deletion site which creates the typical O phenotype.(41)

*FUT2* SNPs rs601338 (428G>A nonsense mutation) and rs1800030 (rare 849G>A nonsense mutation) were directly genotyped from the ImmunoChip. The additional non-secretor mutations 385 T>G and 571C>T were imputed from ImmunoChip data using the program IMPUTE2. This program calculates the genotype of unobserved SNPs with >0.9 probability by comparing highly-associated observed SNPs to a database of haplotypes constructed from the 1,000 Genomes Project.(42)

For quality control, 10% of all samples as well as samples which failed ImmunoChip analysis were analyzed at the *FUT2* 428 position using RT-PCR (TaqMan, Life Technologies).
Samples from subjects of self-identified Asian ancestry were further analyzed at position 385 by TaqMan assay.

Genetic ancestry was determined using 1723 AIMs on the ImmunoChip. Principal Components Analysis (PCA) was used to identify and remove genetic outliers of greater than six standard deviations. Genetic clusters were then created using the well-described program STRUCTURE. Briefly, STRUCTURE uses Markov Chain Monte Carlo methods to create genetically similar clusters by maximizing within-group Hardy-Weinberg equilibrium. Program parameters included three assumed genetic sources with admixture among groups. The resulting clusters were labeled according to the self-identified race/ethnicity most prevalent in each group.

Statistical analysis

Demographic and clinical data for both cases and controls were compared with secretor genotype by chi-square or Fisher’s exact test. The Cochrane-Mantel-Haenszel test was used to control for race, ethnicity, and genetic ancestry as appropriate. Logistic regression was used to generate odds ratios adjusted for covariates. The regression model was stratified by genetic ancestry. Data were analyzed by RLC in collaboration with ALM.

Results

Enrolled norovirus cases and controls

During the year of prospective surveillance, 2954 children with AGE were identified, of whom 2092 (71%) were enrolled in the study (Figure 1). There were 1122 children enrolled as healthy controls. Among enrolled children, testing was completed on stool and saliva specimens from 1465 (70%) children with AGE and 826 (74%) healthy controls. Sequence-confirmed GI and GII norovirus RNA was detected in stool specimens from 302 (21%) AGE cases and 52 (6%) controls (Figure 1). Norovirus AGE cases were comparable to controls for demographic
variables such as sex and insurance status but differed modestly from controls by age, study site, and self-identified race/ethnicity (Table 1).

Genetic ancestry grouping

A total of 2241 (98%) subjects had Immunochip data for detailed analysis of genetic ancestry. PCA using AIMs identified 21 genetic outliers (all self-identified Asian / Pacific Islander). The remaining 2220 subjects were assigned by STRUCTURE into three genetically similar clusters (White, Black, and Hispanic) which had 95% agreement with self-identified race/ethnicity. Of the 582 (26%) subjects who self-identified multiple race/ethnicities, 62% reported backgrounds that were Hispanic and White. Genetic ancestry was considered to agree with self-identified race/ethnic background if any self-identified category was matched.

Secretor genotype and norovirus infection

*FUT2* non-secretor polymorphisms at positions 385, 571, and 849 were not observed in this study population. The *FUT2* 428 non-secretor polymorphism was observed and genotype at this position was therefore used to determine secretor status. Of the demographic variables shown in, race, ethnicity, genetic ancestry, and study site were significantly associated with *FUT2* secretor status; after controlling for AGE case status, study site was not significant. The varying secretor status distribution among controls by race, ethnicity, and genetic ancestry reflected the expected population distribution (Table 3).

Norovirus AGE cases were significantly more likely to be secretors than norovirus-negative healthy controls (p<0.001, odds ratio 2.78 95% CI: 1.77-4.37) in a logistic regression model that included genetic ancestry, age, study site, and health insurance. Results were similar regardless of self-identified or genetically-determined ancestral background (Table 3), when norovirus-positive (asymptomatic) children were included or excluded from the controls, and
when a model of children <12 months included breastfeeding status. O/non-O blood type did not differ significantly in norovirus AGE cases versus norovirus-negative controls (p = 0.48).

Among the children studied, 14 different norovirus genotypes were detected. Only one co-infection with multiple noroviruses (GI.6 and GII.1) was detected. The most commonly detected genotype, norovirus GII.4, was detected in 155 (51%) norovirus AGE cases including 28 (9%) GII.4 Den Haag 2006b, 71 (24%) GII.4 New Orleans 2009, and 56 (19%) GII.4 Sydney 2012. All (100%) GII.4-positive samples from AGE cases were collected from children who were FUT2 secretors (Figure 2). The second most common norovirus genotype, GII.6, also exhibited a significant secretor specificity (p=0.01). The remaining norovirus genotypes did not display secretor specificity (Figure 2).

Norovirus asymptomatic carriers were also more likely to be secretors when compared to norovirus-negative healthy controls (odds ratio 5.64 95% CI: 1.16-7.69, p=0.02) after adjusting for genetic ancestry (Table 3). All of the 21 GII.4 and 3 GII.6 carriers (46% of asymptomatic carriers) were secretors.

Discussion

This large, multi-site study found that a point nonsense mutation which occurs in nearly one-quarter of US children dramatically impacts risk of norovirus infection. FUT2 secretors were at significantly greater risk for both symptomatic and asymptomatic norovirus infections. This finding should be considered in designing Phase III clinical trials, as innate susceptibility to norovirus infections varies by this genetic factor. Challenge studies using noroviruses known to have a secretor preference are already incorporating the secretor status of test subjects into the study design.(44)
The significant association between FUT2 secretor status and risk of norovirus infection reported in this study is driven by the secretor specificity of norovirus genotypes GII.4 and GII.6. These noroviruses comprised 67% of infections in symptomatic children and 46% of infections in asymptomatic children. More than half of norovirus AGE cases were infected with norovirus GII.4 alone; this virus is the predominant genotype in the US and elsewhere in the world. Compared to other norovirus genotypes, GII.4 causes more severe symptoms, has poorer outcomes, and is more likely to be transmitted to others. Our finding that all GII.4 norovirus symptomatic and asymptomatic children were genetic secretors suggests that the non-secretor genetic characteristic offers near-total protection from GII.4 noroviruses.

Furthermore, we found that susceptibility to norovirus due to secretor genotype varies by ethnic group. Among norovirus-free controls, 74% of self-identified non-Hispanics and 86% of self-identified Hispanics are FUT2 secretors. Genetically-identified Hispanics of Meso-American descent are at highest risk, as 96% are FUT2 secretors. This difference in susceptibility to GII.4 and GII.6 norovirus should be taken into consideration in epidemiologic surveillance and in norovirus vaccine effectiveness studies.

The mechanism of FUT2 secretor susceptibility to norovirus infection is the affinity of noroviruses for HBGAs present on cells in the gut. FUT2 enzyme causes secretion of H-antigen at the gut surface. Additional glycosylation of the H-antigen is possible, such as the addition of A or B antigens by ABO. We saw an increased association of norovirus with secretors but within that group did not see an association with O vs. non-O blood type.

This mechanism also suggests a potential therapy for both protection from and treatment of norovirus. It has been proposed that compounds which block norovirus binding could be used as novel therapeutics. This strategy is already at play in nature. Human milk from FUT2
secretors has been shown to have norovirus-binding elements, and secretor infants fed with human milk higher in these elements have lower risk of AGE.\(^{(48, 49)}\)

The results of this study could potentially influence decisions to prioritize vaccination of individuals who lack innate protection. However, we note that non-secretors are susceptible to many noroviruses and thus remain candidates for vaccination. Furthermore, an individual’s secretor status is rarely known, and targeted immunization is challenging to enact.

Our study has several limitations. We studied the circulating noroviruses only for 2011-2012 within the U.S., and noroviruses are known to shift over time and geography. However, our surveillance shows that 21% of children under five years old with AGE tested positive for norovirus, similar to 2008-2010 surveillance.\(^{(28)}\) Norovirus continues to be the leading cause of epidemic and sporadic AGE across all age groups, and our surveillance of sporadic AGE found a distribution of norovirus genotypes similar to that found during U.S. surveillance of AGE outbreaks over the same period.\(^{(27)}\) GII.4 variants that infect secretor individuals, including GII.4 Den Haag, GII.4 New Orleans and GII.4 Sydney, have remained the predominant cause of norovirus AGE in the U.S.\(^{(25, 28)}\) Furthermore, in our study, we found only the 428 inactivating polymorphism, though we examined four known \(FUT2\) inactivating polymorphisms. The lack of other polymorphisms may be explained by having only 21 children of self-identified Asian ancestry in this study. In Asians, \(FUT2\) polymorphisms other than 428G>A often define non-secretors or low-secretors. Though we cannot fully extrapolate our results to regions such as East Asia where other \(FUT2\) inactivating mutations are found, we expect results to be similar for any inactivating mutation.\(^{(8)}\) Additionally, genotype may not fully capture expression of gut HBGAs. Some genetic secretors have low expression of H-antigen and a phenotype similar to
non-secretors.(50) Future studies should examine whether low-H expression in genetic secretors might also confer protection against norovirus.

In conclusion, our study indicates that FUT2 genotype is a major factor driving pediatric norovirus infection risk, particularly with the GII.4 viruses that are predominant in the U.S. Our findings of strain- and ethnicity- dependent susceptibility are relevant to understanding variation in norovirus risk within and between populations and may help guide the design and interpretation of norovirus vaccine trials.
Section 3: Genotype Analysis Methods

Microarray data cleaning

A total of 3888 samples were run by the Morrow lab from 2012-2013, including 2657 samples from the NVSN project. These 3888 samples included replicates and DNA from other studies in order to have enough sample size to determine clusters and genetic ancestry. All Morrow lab samples were from studies conducted on the U.S. pediatric population and are therefore not expected to differ substantially in ancestral background from NVSN samples.

Raw data was available for 3851 samples, with the remaining 37 samples having been damaged during the chipping process and rendered unreadable. Raw intensity data from these samples was analyzed using GenomeStudio version 2011.1 (Illumina, Inc.). A reference clustering algorithm was applied in order to determine preliminary call rates for both individual samples and individual SNPs. This clustering algorithm was created by our collaborators by manually clustering all ImmunoChip SNPs on a sample set of over 6000 subjects. After this preliminary algorithm was applied, samples and SNPs with call rates below 80% were removed.

A review of the literature using the ImmunoChip reveals that this custom-designed chip has poorer SNP performance than standard commercially available microarrays. Researchers have wide variation in both initial and final cutoffs chosen for quality control. Therefore the most permissive cutoff described in the literature, 80%, was selected as the initial cutoff (Table 4).

SNPs were then removed if they came from the Y chromosome or had less than ten samples in any cluster. SNPs with call rates below 95% were manually reclustering to improve call rates. The literature describes final call rate cutoffs between 90-99% (Table 4), so 95% was selected as a moderate value. SNPs and samples with call rates below 95% were excluded.
Gender was automatically calculated for samples by examining heterozygosity at the X chromosome. This was compared against recorded gender. Three discrepancies were identified. Upon further examination, two subjects self-identified as male called as female were shown to have XXY karyotype. The remaining discrepancy was considered a sample misplate and was removed from the final dataset.

The cleaned Morrow data set consisted of 3497 samples, including 3314 (95%) unique samples and 183 (5%) quality control replicates. The FUT2 SNP of primary interest, rs601338, was successfully called for 2238 NVSN subjects.

TaqMan Analysis

Some samples collected did not have enough DNA to run ImmunoChip analysis, or failed ImmunoChip analysis. When possible, a TaqMan assay (LifeTechnologies) for rs601338 was run with the remaining DNA from this sample in order to obtain FUT2 genotype. An additional 53 subjects were able to have FUT2 genotype determined using this method.

For subjects who self-identified as Asian, an additional TaqMan assay was run to identify the most common inactivating mutation in that population, rs1047781, which creates a missense mutation at position 385. The TaqMan assay was inconclusive, and it is unclear whether this is because of a reagent failure or because no non-secretors were present. As referenced in the research manuscript, this polymorphism as well as other more rare non-secretor mutations were imputed from the ImmunoChip data. Non-secretors were not identified by this method.
**Section 4: Ancestry Classification**

In any genetic study, it is of particularly important that cases and controls come from a similar genetic background, or that differences are appropriately adjusted for. The potential for bias exists if both the allelic frequency of the gene under study varies by ancestry and if the disease prevalence varies by genetic ancestry (52). This research is at risk for such a bias based upon the known differences in \textit{FUT2} genotype in MesoAmericans and Asians vs. Caucasians and African Americans as well as the observed differences in self-identified race/ethnicity among cases and controls. It was therefore necessary to identify the ancestral background of study participants.

The purpose of the analysis described below was to create a classification scheme which optimally both reflected subjects’ self-identified ancestral backgrounds and also created genetically similar groups for appropriate comparison.

**Self-identified ancestry**

Classifying individuals into groups of genetically similar ancestry based on self-identification has three major challenges in the context of \textit{FUT2} analysis: (1) the arbitrariness of reducing multiple race/ethnicity designations to one category (2) the known genetic diversity among those of self-identified Hispanic ethnicity and (3) the lack of information to clearly identify individuals of MesoAmerican descent, who are expected to have different allele frequencies at \textit{FUT2} 428 than the rest of the U.S. population. However, this method has the advantage of standardized data collection done on each study participant.

A preliminary analysis was conducted based upon parent-identified ethnicity and race of study participants. Standard CDC/NIH nomenclature was used. Parents of participants were asked to identify one option for ethnicity from Hispanic, Non-Hispanic, Unknown, or Refused.
Parents were asked to identify race and allowed to select multiple options from White, Black/African American, American Indian / Alaskan Native, Asian, Pacific Islander, Other, Unknown, None. The self-identified race and ethnicity distribution according to CDC/NIH nomenclature is shown in Table 5.

Because race and ethnicity were collected as separate variables and multiple races could be selected, hundreds of potential “genetic ancestral backgrounds” could be identified from this dataset. In order to simplify the analysis, four categories were chosen based upon known $FUT2$ genotype variation: MesoAmerican (includes both Hispanic and American Indian), White, Black, and Other (not identifying with any previous groups). While the $FUT2$ 428 allele frequency is not expected to vary in Whites vs. Blacks, the groups were not combined for historical reasons.

Even when simplifying to these four categories, 552 (19%) of individuals could fit into multiple groups. A breakdown of the secretor distribution in controls using all possible combinations of these categories is shown in Table 6. 87-91% of control individuals who self-identified as MesoAmerican were secretors, compared to 71-76% of individuals who self-identified as Black, White, or Black/White. These data suggest that, in a final categorization scheme, self-identified ethnicity should be given precedence over race to create genetically similar groups for appropriate analysis of $FUT2$.

For this categorization scheme, all subjects who self-identified as Hispanic or American Indian were classified as MesoAmerican. All subjects who self-identified as Black (but not as Hispanic, American Indian, or White) were classified as Black. All subjects who self-identified as White (but not as Hispanic, American Indian or Black) were classified as White. Subjects who self-identified as both Black and White (but not as Hispanic or American Indian) were randomly assigned to White or Black. All remaining subjects were assigned to Other.
Importantly, this categorization yielded expected differences in \textit{FUT2} 428 allele frequencies in the controls, shown in Table 2. These data suggest that ancestral grouping based on self-identified race/ethnicity, though crude, might be an adequate method for analyzing the impacts of this polymorphism. However, the presence of 13\% non-secretors in the “MesoAmerican” controls suggests that a more fine-grained analysis would improve the appropriateness of ancestry assignments.

\textit{Available genetic ancestry information}

For 50 subjects, insufficient ImmunoChip information was available to more finely determine genetic ancestry, and the self-identified ancestry was used. The remaining 2241 subjects had additional information available via the ImmunoChip-measured Ancestry Informative Markers (AIMs).

AIMs are those SNPs known to show large allele frequency differences among ancestral populations (52). They are frequently used to classify subjects by genetic ancestry. AIMs on the ImmunoChip were visually inspected within GenomeStudio to confirm appropriate SNP clustering and were manually adjusted if necessary. Only autosomal SNPs with call rates above 95\% were retained for final analysis. 1723 SNPs were retained after applying these quality control measures.

\textit{Principal Components Analysis (PCA)}

Principal Components Analysis (PCA) was used to visualize the basic population structure of the data set and to identify genetic outliers. This was conducted using the PLINK program EIGENSTRAT (53). Initial parameters were set to dimensions (k) = 10, removal iterations (m) = 5, dimensions to remove outliers (t) = 3, and standard deviations for outlier removal (s) = 6.
All outliers were identified within 2 iterations and 2 outlier removal dimensions. A sensitivity analysis for dimensions was run by varying k from 1 to 10. No change was seen after the third dimension. A sensitivity analysis for number of standard deviations was run by varying from s = 3 to 7 as shown in Table 7. From s = 4 to 6, the same 20 individuals who self-identified as ‘Other’ were identified as outliers. Of these 20, 17 self-identified as Asian, 2 self-identified as Pacific Islander, and 1 self-identified as Asian/Pacific Islander. All were FUT2 secretors at the 428 site.

Because this group is not expected to have variation at the FUT2 428 position, their inclusion in this study does not contribute to the analysis. We thus chose s = 6 as the parameter which maximized the number of included samples while clearly identifying as outliers those subjects whose inclusion was inappropriate. Final parameters used were k = 3, m = 4, t = 3, and s = 6. Results are represented in Figure 3.

Figure 3 generally indicates three populations (bottom left, bottom right, and top right) with a good deal of admixture between the bottom right and the other two groups. This pattern is what would be expected from the American population: a heterogenous Hispanic group (top right) admixed with Caucasians (bottom right), as well as admixture among Caucasians and African Americans (bottom left).

STRUCTURE analysis

The 21 subjects identified as outliers from PCA were classified as “Other.” For the remaining subjects, final assignment to genetically similar groups was done using the well-described STRUCTURE method (43). STRUCTURE uses Markov Chain Monte Carlo methods to probabilistically assign samples to clusters of genetic similarity by maximizing within-group
Hardy-Weinberg equilibrium. Program parameters were: 3 assumed genetic sources, admixed populations, 10,000 burn-in period with 10,000 repetitions.

Figure 4 visualizes the STRUCTURE results. The three populations identified by STRUCTURE were named according to the self-identified race/ethnicity which had the greatest membership in each group. (Table 3 contains a breakdown of agreement for each group).

STRUCTURE gives probability assignments for each group, allowing for identification of admixed individuals. An individual was assigned to a certain “pure” population if he/she was given ≥ 80% probability of belonging to that group by STRUCTURE. If the sum of probabilities for two groups was needed to reach 80%, the individual was considered to belong to both groups. If no combination of two groups surpassed 80%, the individual was considered to be of White, Black, and MesoAmerican ancestry.

To evaluate the appropriateness of this classification method, we look at both agreement with self-identified ancestry and the distribution of FUT2 genotypes. Table 8 demonstrates that STRUCTURE agrees with self-identified race/ethnicity 97% of the time.

Final Ancestry Assignments

Analyzing the data using the eight ancestry categories identified using STRUCTURE and PCA found a significant relationship of secretor status in NoV AGE cases vs. controls after adjusting for genetic ancestry (p < 0.001 by Cochran Mantel Hantzel test, adjusted exact odds ratio 2.71 95% CI 1.79-4.39). However, it was desirable to collapse the ancestral groups into the fewer categories shown in Table 3 to simplify data presentation. Collapsing the groups yields nearly identical results (p < 0.001, odds ratio 2.81 95% CI: 1.81-4.57). This simplification was therefore deemed to be acceptable.
An advantage to STRUCTURE is that it gives probability assignments to each cluster. It thus makes the reassignment of admixed individuals to a single cluster more objective. Table 9 shows these final assignments, which differ substantially from how subjects would have been randomly assigned or assigned to one group by default as in the self-identified ancestry classification system. Figure 4 allows visualization of how these probabilities are distributed.

This final classification scheme allows for 97% agreement with self-identified ancestry (Table 8) while reflecting the expected population FUT2 428 secretor status in controls. Table 2 illustrates the change in FUT2 secretor status distribution among controls in self-identified race/ethnicity vs. genetic ancestry.

The most notable change in this new classification system is the decrease in overall percentage of individuals classified as MesoAmerican. This was largely achieved by reclassifying these subjects as White. These changes are demonstrated in Figure 5 and Figure 6.

Of note, in the final analysis, self-identification as American Indian was not significantly related to membership in the “MesoAmerican” genetic ancestral group. Therefore, to simplify presentation to a general audience, the grouping of individuals of American Indian and Hispanic descent into a “MesoAmerican” group was not mentioned in the research manuscript. The cluster was renamed “Hispanic” for the manuscript.
Section 5: Control group selection

Two major questions arise when determining if the control group for this study is an acceptable comparator group: was there differential bias in control group vs. case group selection, and does the control group appropriately reflect the general population? These questions are particularly important given the known differences in race/ethnicity between cases and controls, and the association of race/ethnicity with the outcome variable secretor status.

Section 4: Ancestry Classification outlines in detail how the distribution of secretor status in the control group reflects the expected population distributions by genetic ancestry. Because different populations are expected to have different secretor status distributions, and because of the known differential recruitment of these populations, the final statistics relating secretor status to NoV AGE contain adjustments for genetic ancestry. This was done using the Cochran-Mantel-Haenszel test to compare stratified categorical data.

With this adjustment for genetic background, the control group is expected to be an appropriate reflection of the U.S. population. However, it is true that the control group was recruited in a different manner than the cases. AGE cases had reported to emergency departments or been referred to the hospital for symptoms of AGE. Healthy controls were at well-child visits at affiliated clinics. It is not unreasonable to assume that these two populations have different health-seeking behaviors, or that the control group may be inappropriately enriched children healthier than the general population.

To account for any hidden confounding that these differences might create between cases and controls, a second control group was created from the NoV negative AGE cases. These children would be expected to generally have the same health-seeking behaviors and health status as children with NoV AGE. It has also been suggested that rotavirus may also have a
secretor preference. (54) The NVSN data confirm this association (data not shown), so rotavirus positive AGE cases were also removed from the new “NoV-AGE cases” control group.

As an additional comparison, healthy controls without the asymptomatic carriers removed were also evaluated as a control group.

Regardless of control group used, the association between NoV AGE and secretor status was highly significant ($p < 0.001$) after adjustments for genetic ancestry, self-identified race, and self-identified ethnicity (Table 10).
Section 6: Potential Bias and Confounding

Selection bias in recruitment/retention

Not all children eligible for this study during surveillance hours enrolled in the study, and not all enrolled subjects submitted both stool and saliva specimens. Reasons for study non-completion of eligible children are shown in Table 11. The most common cause of study non-completion by eligible children was the lack of complete specimen collection (39%).

Basic information from the medical chart was collected on all children who met inclusion criteria during surveillance hours regardless of final study participation. These variables were analyzed to determine sources of selection bias. Table 12 outlines differences between eligible children on surveillance days who completed the study vs. those who did not.

It is important to note that the comparisons in Table 12 differ from the demographic comparisons made in Table 2. The variables for race and ethnicity in Table 12 were determined from the medical chart, which did not always have this information documented, as opposed to patient self-identification in Table 2. This categorization method was necessary in order to make a direct comparison of enrolled and non-enrolled children since non-enrolled children did not have detailed self-identified race/ethnicity information available.

Also important to note that there are 2345 study completers vs. 2291 subjects in the final data set used in Section 2: Research Manuscript. 54 subjects completed the study but did not have their samples analyzed due to assay failure or errors in sample handling (). These failures are separate from recruitment/retention and are thus not addressed here.

Table 12 highlights significant differences between enrolled study completers and eligible children who did not complete the study in the following variables: study site, health insurance status, age group, screening race, and screening ethnicity.
Selection bias can decrease the generalizability of results (external validity). A limitation of this study is that it may not have captured a complete picture of U.S. pediatric AGE. However, this bias threatens the internal validity of study conclusions only if the differentially selected variables are confounders: that is, they have association with both the predictor variable (secretor status) and the outcome variable (NoV AGE case status).

**Potential confounding variables**

Importantly, there was no significant difference in study completion in children eligible for the AGE case group vs. the healthy control group. While the presence of NoV AGE in non-enrolled children cannot be assessed, some idea of rates of NoV in non-participating children can be assessed by looking at NoV in enrolled children who collected stool but not saliva. Stool samples from AGE cases in Cincinnati, Nashville, and Rochester were tested for NoV regardless of whether saliva was collected. NoV was detected in 17% of study completers vs. 15% of subjects who did not collect saliva. This difference was not significant (p = 0.20, Pearson’s chi-squared test).

Table 13 looks at the association of potential confounding variables and secretor status. In addition to ancestry/race/ethnicity, study site and age group are potential confounders.

Genetic background (which encompasses race and ethnicity) is known to be associated with secretor status. It is difficult to assess variation in recruitment/retention of subjects’ by race/ethnicity since this information was not always included in the medical chart. The impact of variability in race/ethnicity in this study was accounted for statistically by stratification and adjustment. Also, the fact that the secretor status distribution in the control group reflected expected population distribution further alleviates concerns that this variation impacts the validity of study conclusions. More details can be found in
Section 5: Control group selection

Study site variation

Much of the remaining variation can be explained by differences in basic demographics and enrollment practices at each study site. Statistically, the association of secretor status with age group in controls becomes non-significant after adjusting for study site.

The reasons for non-completion were different by site (Table 14) and largely expected because each site samples from a different population. For example, Seattle has the highest % of non-completion caused by language barriers (22%) while Houston has the lowest (0%). Seattle is the most racially diverse recruitment site and is expected to attract more patients who do not speak English. Houston has a patient population of both English speakers and Spanish speakers, and the Spanish-speaking population is significant enough that Houston has Spanish-speaking enrollers and study material.

Failure to collect all specimens, the most common reason for non-completion, varied greatly by site (Table 14). These differences are likely caused by different infrastructure at each site available to perform the saliva collection and follow up on stool collection. One site, Kansas City, was able to collect saliva from every enrolled participants. If only subjects from Kansas City are analyzed, thus completely removing the potential impact of different saliva collection practices, this study’s findings of a significant association between NoV AGE and secretor status remain.
Bibliography


17. Hall AJ, Curns AT, McDonald LC, Parashar UD, Lopman BA. The roles of Clostridium difficile and norovirus among gastroenteritis-associated deaths in the United States,


HIV-1 infection. AIDS. 2006;20(5):685-9. Epub 2006/03/04. doi: 10.1097/01.aids.0000216368.23325.bc


PONE-D-12-28276 [pii]. PubMed PMID: 23967090; PubMed Central PMCID: PMC3742662.


ng.2667 [pii]. PubMed PMID: 23749187; PubMed Central PMCID: PMC3757343.


Tables and Figures
<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>Study type</th>
<th>Secretor method</th>
<th>NoV</th>
<th>Findings</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>77</td>
<td>Challenge</td>
<td>Phenotype, 428 genotype</td>
<td>8FIIa (GI.1)</td>
<td>Only secretors became ill.</td>
<td>(55)</td>
</tr>
<tr>
<td>Sweden</td>
<td>53</td>
<td>Outbreaks (nosocomial and sporadic)</td>
<td>428 genotype and saliva binding assay</td>
<td>GII (Lordsdale-like)</td>
<td>Secretors sig more likely to be ill.</td>
<td>(56)</td>
</tr>
<tr>
<td>Sweden (blood donors)</td>
<td>105</td>
<td>Retrospective cohort</td>
<td>FUT2/FUT3 genotype</td>
<td>GII.4 (antibody search)</td>
<td>Non-secretors sig more likely to be NoV IgG-</td>
<td>(57)</td>
</tr>
<tr>
<td>Denmark</td>
<td>61</td>
<td>Outbreaks</td>
<td>Genotyped at 428 and 571</td>
<td>GII.4</td>
<td>Secretors sig associated with outbreak.</td>
<td>(58)</td>
</tr>
<tr>
<td>China</td>
<td>146</td>
<td>Outbreaks</td>
<td>Phenotype</td>
<td>GII-3, GII-4</td>
<td>Secretors sig associated with outbreaks.</td>
<td>(11)</td>
</tr>
<tr>
<td>Spain (nursing home)</td>
<td>116</td>
<td>Nosocomial outbreak</td>
<td>Saliva binding assay to NoV virus-like protein</td>
<td>GII.4</td>
<td>Secretors sig more likely to become ill.</td>
<td>(10)</td>
</tr>
<tr>
<td>Sweden</td>
<td>33</td>
<td>Outbreak</td>
<td>Genotyped at 428</td>
<td>Gl.3</td>
<td>No association found.</td>
<td>(12)</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>28</td>
<td>Citywide surveillance</td>
<td>Genotype at 428 and 571</td>
<td>multiple</td>
<td>Only secretors NoV+.</td>
<td>(20)</td>
</tr>
<tr>
<td>Nicaragua (children)</td>
<td>19</td>
<td>Cross section</td>
<td>Phenotype</td>
<td>multiple</td>
<td>No association between secretor status and asx carriage.</td>
<td>(24)</td>
</tr>
<tr>
<td>U.S.</td>
<td>40</td>
<td>Challenge</td>
<td>Phenotype</td>
<td>GII.4</td>
<td>Secretors sig more likely to become ill.</td>
<td>(9)</td>
</tr>
<tr>
<td>Brazil (children)</td>
<td>12</td>
<td>Case/ control</td>
<td>Phenotype</td>
<td>multiple</td>
<td>No association found.</td>
<td>(59)</td>
</tr>
<tr>
<td>West Africa (children)</td>
<td>37</td>
<td>single hospital surveillance</td>
<td>Phenotype</td>
<td>GII.4 and others</td>
<td>Secretors sig more likely to have NoV AGE.</td>
<td>(60)</td>
</tr>
<tr>
<td>Vietnam</td>
<td>50</td>
<td>single hospital surveillance</td>
<td>Phenotype</td>
<td>GII.3, GII.4</td>
<td>Secretors sig more likely to have GII.4 but not GII.3.</td>
<td>(54)</td>
</tr>
</tbody>
</table>
Table 2. Demographics of norovirus AGE cases and healthy control study completers

<table>
<thead>
<tr>
<th></th>
<th>Norovirus AGE cases, n (% of cases)</th>
<th>All healthy controls, n (% of controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>302 (100)</td>
<td>826 (100)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>171 (57)</td>
<td>428 (52)</td>
</tr>
<tr>
<td>Female</td>
<td>131 (43)</td>
<td>398 (48)</td>
</tr>
<tr>
<td><strong>Study site</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nashville</td>
<td>66 (22)</td>
<td>157 (19)</td>
</tr>
<tr>
<td>Rochester</td>
<td>8 (3)</td>
<td>58 (7)</td>
</tr>
<tr>
<td>Cincinnati</td>
<td>47 (16)</td>
<td>208 (25)</td>
</tr>
<tr>
<td>Seattle</td>
<td>23 (8)</td>
<td>54 (7)</td>
</tr>
<tr>
<td>Houston</td>
<td>43 (14)</td>
<td>107 (13)</td>
</tr>
<tr>
<td>Kansas City</td>
<td>115 (38)</td>
<td>242 (29)</td>
</tr>
<tr>
<td><strong>Health Insurance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Insurance</td>
<td>231 (76)</td>
<td>627 (76)</td>
</tr>
<tr>
<td>Private Insurance</td>
<td>59 (20)</td>
<td>180 (22)</td>
</tr>
<tr>
<td>None/Unknown</td>
<td>12 (4)</td>
<td>19 (2)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-6 mo</td>
<td>34 (11)</td>
<td>218 (26)</td>
</tr>
<tr>
<td>6-12 mo</td>
<td>81 (27)</td>
<td>165 (20)</td>
</tr>
<tr>
<td>12-24mo</td>
<td>94 (31)</td>
<td>225 (27)</td>
</tr>
<tr>
<td>24-36 mo</td>
<td>51 (17)</td>
<td>95 (12)</td>
</tr>
<tr>
<td>36-60 mo</td>
<td>42 (14)</td>
<td>123 (15)</td>
</tr>
<tr>
<td><strong>Self-identified race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>145 (48)</td>
<td>324 (39)</td>
</tr>
<tr>
<td>Black</td>
<td>99 (33)</td>
<td>376 (46)</td>
</tr>
<tr>
<td>Multiple/other</td>
<td>58 (19)</td>
<td>126 (15)</td>
</tr>
<tr>
<td><strong>Self-identified ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>210 (70)</td>
<td>681 (82)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>92 (30)</td>
<td>145 (18)</td>
</tr>
<tr>
<td><strong>Genetic ancestry</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>135 (45)</td>
<td>335 (41)</td>
</tr>
<tr>
<td>Black</td>
<td>101 (33)</td>
<td>392 (47)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>53 (18)</td>
<td>74 (9)</td>
</tr>
<tr>
<td>Unable to classify</td>
<td>13 (4)</td>
<td>25 (3)</td>
</tr>
</tbody>
</table>

Norovirus AGE cases vs. controls, *p<0.05, **p<0.01, ***p<0.001
<table>
<thead>
<tr>
<th></th>
<th>Secretors in norovirus AGE cases / n (%)</th>
<th>Secretors in asymptomatic norovirus carriers / n (%)</th>
<th>Secretors in norovirus-negative controls / n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>276/302 (91)</td>
<td>47/52 (90)</td>
<td>591/774 (76)</td>
</tr>
<tr>
<td><strong>Self-identified race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>136/145 (94)</td>
<td>17/18 (94)</td>
<td>240/306 (78)</td>
</tr>
<tr>
<td>Black</td>
<td>85/99 (86)</td>
<td>23/25 (92)</td>
<td>253/351 (72)</td>
</tr>
<tr>
<td>Multiple/Other</td>
<td>55/58 (95)</td>
<td>7/9 (78)</td>
<td>98/117 (84)</td>
</tr>
<tr>
<td><strong>Self-identified ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>185/210 (88)</td>
<td>37/42 (88)</td>
<td>475/639 (74)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>91/92 (99)</td>
<td>10/10 (100)</td>
<td>116/135 (86)</td>
</tr>
<tr>
<td><strong>Genetic ancestry</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>123/135 (91)</td>
<td>17/19 (89)</td>
<td>235/316 (74)</td>
</tr>
<tr>
<td>Black</td>
<td>87/101 (86)</td>
<td>24/27 (89)</td>
<td>267/365 (73)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>53/53 (100)</td>
<td>5/5 (100)</td>
<td>66/69 (96)</td>
</tr>
<tr>
<td>Unable to classify</td>
<td>13/13 (100)</td>
<td>1/1 (100)</td>
<td>23/24 (96)</td>
</tr>
</tbody>
</table>

*p<0.05 vs norovirus-negative controls, **p<0.01 vs norovirus-negative controls, ***p<0.001 vs norovirus-negative controls

^92% of individuals in White genetic ancestry group self-identified as White
98% of individuals in Black genetic ancestry group self-identified as Black
97% of individuals in Hispanic genetic ancestry group self-identified as Hispanic
Table 4. Variation in ImmunoChip quality control protocols in published literature

<table>
<thead>
<tr>
<th>Reference</th>
<th>Initial sample call rate</th>
<th>Final sample call rate</th>
<th>Initial SNP call rate</th>
<th>Final SNP call rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(61)</td>
<td></td>
<td></td>
<td></td>
<td>95%</td>
</tr>
<tr>
<td>(51)</td>
<td></td>
<td>80%</td>
<td></td>
<td>99%</td>
</tr>
<tr>
<td>(62)</td>
<td>95%</td>
<td>98%</td>
<td>95%</td>
<td>98%</td>
</tr>
<tr>
<td>(63)</td>
<td>98%</td>
<td></td>
<td></td>
<td>98%</td>
</tr>
<tr>
<td>(64)</td>
<td>90%</td>
<td></td>
<td></td>
<td>98%</td>
</tr>
<tr>
<td>(65)</td>
<td>95%</td>
<td></td>
<td></td>
<td>95%</td>
</tr>
</tbody>
</table>
Table 5. Self-identified race/ethnicity of subjects with stool and saliva data

<table>
<thead>
<tr>
<th>Ethnic Category</th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic or Latino</td>
<td>408</td>
<td>145</td>
<td>553</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>1056</td>
<td>679</td>
<td>1735</td>
</tr>
<tr>
<td>Unknown (individuals not reporting ethnicity)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Ethnic Category: Total of All Subjects</td>
<td>1565</td>
<td>826</td>
<td>2291</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Racial Categories</th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian/Alaska Native</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Asian</td>
<td>16</td>
<td>17</td>
<td>33</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Black or African American</td>
<td>559</td>
<td>376</td>
<td>935</td>
</tr>
<tr>
<td>White</td>
<td>629</td>
<td>324</td>
<td>953</td>
</tr>
<tr>
<td>More Than One Race</td>
<td>131</td>
<td>179</td>
<td>205</td>
</tr>
<tr>
<td>Unknown or Not Reported</td>
<td>124</td>
<td>28</td>
<td>152</td>
</tr>
<tr>
<td>Racial Categories: Total of All Subjects</td>
<td>1465</td>
<td>826</td>
<td>2291</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Report on Hispanics/Latinos</th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian or Alaska Native</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Black or African American</td>
<td>26</td>
<td>10</td>
<td>36</td>
</tr>
<tr>
<td>White</td>
<td>252</td>
<td>95</td>
<td>365</td>
</tr>
<tr>
<td>More Than One Race</td>
<td>16</td>
<td>14</td>
<td>32</td>
</tr>
<tr>
<td>Unknown or Not Reported</td>
<td>108</td>
<td>25</td>
<td>134</td>
</tr>
<tr>
<td>Racial Categories: Total of Hispanics or Latinos</td>
<td>408</td>
<td>145</td>
<td>553</td>
</tr>
<tr>
<td></td>
<td>Secretors</td>
<td>Nonsecretors</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>270 (74)</td>
<td>97 (26)</td>
<td></td>
</tr>
<tr>
<td>Black/White</td>
<td>30 (71)</td>
<td>12 (29)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>181 (76)</td>
<td>57 (24)</td>
<td></td>
</tr>
<tr>
<td>White/MesoAmerican</td>
<td>90 (87)</td>
<td>13 (13)</td>
<td></td>
</tr>
<tr>
<td>MesoAmerican</td>
<td>24 (89)</td>
<td>3 (11)</td>
<td></td>
</tr>
<tr>
<td>MesoAmerican/Black</td>
<td>13 (81)</td>
<td>3 (19)</td>
<td></td>
</tr>
<tr>
<td>MesoAmerican/Black/White</td>
<td>10 (91)</td>
<td>1 (9)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>20 (91)</td>
<td>2 (9)</td>
<td></td>
</tr>
</tbody>
</table>
Table 7. Morrow data sensitivity to outlier definition

<table>
<thead>
<tr>
<th>Standard deviations (s)</th>
<th>NVSN outliers removed</th>
<th># outliers self-identified as 'Other'</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>470</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>57</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>
Table 8. Comparison of self-identified ancestry to STRUCTURE identified ancestry.

<table>
<thead>
<tr>
<th>Self-identified genetic group</th>
<th>STRUCTURE Genetic Group</th>
<th>Black</th>
<th>Black/White</th>
<th>White</th>
<th>White/Meso</th>
<th>Meso</th>
<th>Meso/Black</th>
<th>Meso/Black/White</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>761 (34)</td>
<td>113</td>
<td>2 (0.1)</td>
<td>1 (0.04)</td>
<td>0</td>
<td>7 (0.3)</td>
<td>4 (0.2)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Black/White</td>
<td>6 (0.3)</td>
<td>109</td>
<td>2 (0.1)</td>
<td>0 (0)</td>
<td>0</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>3 (0.1)</td>
<td>13 (0.6)</td>
<td>574 (26)</td>
<td>26 (1)</td>
<td>1 (0.04)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (0.04)</td>
<td></td>
</tr>
<tr>
<td>White/Meso</td>
<td>0 (0)</td>
<td>9 (0.4)</td>
<td>44 (2)</td>
<td>260 (12)</td>
<td>35 (2)</td>
<td>1 (0.04)</td>
<td>3 (0.1)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Meso</td>
<td>3 (0.1)</td>
<td>11 (5)</td>
<td>2 (0.1)</td>
<td>112 (5)</td>
<td>7 (0.3)</td>
<td>4 (0.2)</td>
<td>3 (0.1)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Meso/Black</td>
<td>13 (0.6)</td>
<td>18 (0.8)</td>
<td>0 (0)</td>
<td>1 (0.04)</td>
<td>0 (0)</td>
<td>1 (0.04)</td>
<td>12 (0.5)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Meso/Black/White</td>
<td>1 (0.04)</td>
<td>19 (0.8)</td>
<td>2 (0.1)</td>
<td>2 (0.1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (0.04)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4 (0.2)</td>
<td>8 (0.4)</td>
<td>2 (0.1)</td>
<td>20 (0.9)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>20 (0.9)</td>
<td></td>
</tr>
</tbody>
</table>

Perfect match, 77%
Agreement, 20%
Disagreement, 3%
Table 9. STRUCTURE handling of admixed subjects

<table>
<thead>
<tr>
<th>Multi-ancestral STRUCTURE assignment</th>
<th>Final STRUCTURE group</th>
<th>Black</th>
<th>White</th>
<th>Meso</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black / White</td>
<td></td>
<td>174 (58%)</td>
<td>126 (42%)</td>
<td></td>
</tr>
<tr>
<td>White / Meso</td>
<td></td>
<td>157 (37%)</td>
<td>265 (63%)</td>
<td></td>
</tr>
<tr>
<td>Meso / Black</td>
<td></td>
<td>9 (69%)</td>
<td>4 (31%)</td>
<td></td>
</tr>
<tr>
<td>Meso/Black / White</td>
<td></td>
<td>15 (65%)</td>
<td>5 (22%)</td>
<td>3 (13%)</td>
</tr>
</tbody>
</table>
Table 10. Secretor status of potential control groups

<table>
<thead>
<tr>
<th>Secretors in NoV cases / n (% cases)</th>
<th>Secretors in healthy controls / n (% controls)</th>
<th>Secretors in NoV- controls / n (% controls)</th>
<th>Secretors in NoV-, rota-cases / n (% controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>276/302 (91)**</td>
<td>638/826 (77)</td>
<td>591/774 (76)</td>
</tr>
</tbody>
</table>

**Self-ID race**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>136/145 (94)</td>
<td>257/324 (79)</td>
<td>240/306 (78)</td>
</tr>
<tr>
<td>Black</td>
<td>85/99 (86)</td>
<td>276/376 (73)</td>
<td>235/351 (72)</td>
</tr>
<tr>
<td>Multiple/Other</td>
<td>55/58 (95)</td>
<td>105/126 (83)</td>
<td>98/117 (84)</td>
</tr>
</tbody>
</table>

**Self-ID ethnicity**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Hispanic</td>
<td>185/210 (88)</td>
<td>512/681 (75)</td>
<td>475/639 (74)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>91/92 (99)</td>
<td>126/145 (87)</td>
<td>116/135 (86)</td>
</tr>
</tbody>
</table>

**Genetic ancestry**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>123/135 (91)**</td>
<td>252/335 (75)</td>
<td>235/316 (74)</td>
</tr>
<tr>
<td>Black</td>
<td>87/101 (86)*</td>
<td>291/392 (74)</td>
<td>267/365 (73)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>53/53 (100)</td>
<td>71/74 (96)</td>
<td>66/69 (96)</td>
</tr>
<tr>
<td>Asian</td>
<td>3/3 (100)</td>
<td>12/12 (100)</td>
<td>12/12 (0)</td>
</tr>
<tr>
<td>Unable to classify#</td>
<td>10/10 (100)</td>
<td>12/13 (92)</td>
<td>11/12 (92)</td>
</tr>
</tbody>
</table>

*p<0.05 vs. NoV AGE cases, adjusted for race, ethnicity, or ancestry as appropriate.

**p<0.05 vs. NoV AGE cases, adjusted for race, ethnicity, or ancestry as appropriate.
Table 11. Study non-completion of eligible children during surveillance hours

<table>
<thead>
<tr>
<th></th>
<th>Cases, n (% of cases)</th>
<th>Controls, n (% of controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refused</td>
<td>409 (28)</td>
<td>316 (40)</td>
</tr>
<tr>
<td>MD refused</td>
<td>33 (2)</td>
<td>7 (1)</td>
</tr>
<tr>
<td>Language barrier</td>
<td>262 (18)</td>
<td>76 (10)</td>
</tr>
<tr>
<td>Missed</td>
<td>152 (11)</td>
<td>97 (12)</td>
</tr>
<tr>
<td>No stool</td>
<td>288 (20)</td>
<td>242 (31)</td>
</tr>
<tr>
<td>No saliva</td>
<td>228 (16)</td>
<td>38 (5)</td>
</tr>
<tr>
<td>No specimens</td>
<td>67 (5)</td>
<td>7 (1)</td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (0.3)</td>
<td>3 (0.3)</td>
</tr>
<tr>
<td></td>
<td>Eligible non-completers, n (% category)</td>
<td>Study completers, n (% category)</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2229 (49)</td>
<td>2345 (51)</td>
</tr>
<tr>
<td><strong>Case status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE case</td>
<td>1444 (49)</td>
<td>1510 (51)</td>
</tr>
<tr>
<td>Healthy control</td>
<td>785 (49)</td>
<td>835 (52)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1169 (48)</td>
<td>1251 (52)</td>
</tr>
<tr>
<td>Female</td>
<td>1060 (49)</td>
<td>1094 (51)</td>
</tr>
<tr>
<td><strong>Study site</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nashville</td>
<td>302 (39)</td>
<td>478 (61)</td>
</tr>
<tr>
<td>Rochester</td>
<td>188 (62)</td>
<td>114 (38)</td>
</tr>
<tr>
<td>Cincinnati</td>
<td>425 (44)</td>
<td>543 (56)</td>
</tr>
<tr>
<td>Seattle</td>
<td>281 (65)</td>
<td>153 (35)</td>
</tr>
<tr>
<td>Houston</td>
<td>388 (57)</td>
<td>299 (44)</td>
</tr>
<tr>
<td>Kansas City</td>
<td>645 (46)</td>
<td>758 (54)</td>
</tr>
<tr>
<td><strong>Health Insurance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public insurance</td>
<td>677 (27)</td>
<td>1823 (73)</td>
</tr>
<tr>
<td>Private Insurance</td>
<td>163 (27)</td>
<td>443 (73)</td>
</tr>
<tr>
<td>None/Unknown</td>
<td>1389 (95)</td>
<td>78 (5)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-6 mo</td>
<td>344 (39)</td>
<td>533 (61)</td>
</tr>
<tr>
<td>6-12 mo</td>
<td>416 (46)</td>
<td>490 (54)</td>
</tr>
<tr>
<td>12-24 mo</td>
<td>603 (49)</td>
<td>621 (51)</td>
</tr>
<tr>
<td>24-36 mo</td>
<td>352 (51)</td>
<td>335 (49)</td>
</tr>
<tr>
<td>36-60 mo</td>
<td>514 (58)</td>
<td>366 (42)</td>
</tr>
<tr>
<td><strong>Screening Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>695 (45)</td>
<td>849 (55)</td>
</tr>
<tr>
<td>Black</td>
<td>828 (46)</td>
<td>959 (54)</td>
</tr>
<tr>
<td>Multiple/Other</td>
<td>706 (57)</td>
<td>537 (43)</td>
</tr>
<tr>
<td><strong>Screening ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>1816 (48)</td>
<td>1980 (52)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>413 (53)</td>
<td>365 (47)</td>
</tr>
</tbody>
</table>

*p<0.05, **p<0.001
### Table 13. Association of secretor status with potential confounders

<table>
<thead>
<tr>
<th></th>
<th>Secretors in all controls / n (%)</th>
<th>Secretors in all subjects / n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>638/826 (77)</td>
<td>1842/2291 (80)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Male</td>
<td>333/428 (78)</td>
<td>983/1222 (80)</td>
</tr>
<tr>
<td>Female</td>
<td>305/398 (77)</td>
<td>859/1069 (80)</td>
</tr>
<tr>
<td><strong>Study site</strong></td>
<td>NS</td>
<td>**</td>
</tr>
<tr>
<td>Nashville</td>
<td>124/157 (79)</td>
<td>392/474 (83)</td>
</tr>
<tr>
<td>Rochester</td>
<td>39/58 (67)</td>
<td>94/112 (75)</td>
</tr>
<tr>
<td>Cincinnati</td>
<td>155/208 (75)</td>
<td>405/532 (76)</td>
</tr>
<tr>
<td>Seattle</td>
<td>44/54 (82)</td>
<td>126/149 (85)</td>
</tr>
<tr>
<td>Houston</td>
<td>89/107 (83)</td>
<td>243/274 (89)</td>
</tr>
<tr>
<td>Kansas City</td>
<td>187/242 (77)</td>
<td>582/750 (79)</td>
</tr>
<tr>
<td><strong>Health Insurance</strong></td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Public insurance</td>
<td>481/627 (77)</td>
<td>1437/1782 (81)</td>
</tr>
<tr>
<td>Private Insurance</td>
<td>140/180 (78)</td>
<td>341/431 (79)</td>
</tr>
<tr>
<td>None/Unknown</td>
<td>17/19 (90)</td>
<td>64/78 (82)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>*</td>
<td>NS</td>
</tr>
<tr>
<td>0-6 mo</td>
<td>161/218 (74)</td>
<td>411/525 (78)</td>
</tr>
<tr>
<td>6-12 mo</td>
<td>118/165 (72)</td>
<td>375/475 (79)</td>
</tr>
<tr>
<td>12-24mo</td>
<td>184/225 (82)</td>
<td>491/605 (81)</td>
</tr>
<tr>
<td>24-36 mo</td>
<td>81/95 (85)</td>
<td>278/329 (85)</td>
</tr>
<tr>
<td>36-60 mo</td>
<td>94/123 (76)</td>
<td>287/357 (80)</td>
</tr>
<tr>
<td><strong>Self-ID race</strong></td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>White</td>
<td>257/324 (79)</td>
<td>797/953 (84)</td>
</tr>
<tr>
<td>Black</td>
<td>276/376 (73)</td>
<td>700/935 (75)</td>
</tr>
<tr>
<td>Multiple/Other</td>
<td>105/126 (83)</td>
<td>345/403 (86)</td>
</tr>
<tr>
<td><strong>Self-ID ethnicity</strong></td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>512/681 (75)</td>
<td>1335/1738 (77)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>126/145 (87)</td>
<td>507/553 (92)</td>
</tr>
<tr>
<td><strong>Genetic ancestry</strong></td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>White</td>
<td>252/335 (75)</td>
<td>728/916 (80)</td>
</tr>
<tr>
<td>Black</td>
<td>291/392 (74)</td>
<td>739/989 (75)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>71/74 (96)</td>
<td>309/315 (98)</td>
</tr>
<tr>
<td>Asian</td>
<td>12/12 (100)</td>
<td>21/21 (100)</td>
</tr>
<tr>
<td>Unable to Determine</td>
<td>12/13 (92)</td>
<td>45/50 (90)</td>
</tr>
</tbody>
</table>

*p<0.05, **p<0.001
Table 14. Reasons for non-completion of eligible children during surveillance hours, by study site

<table>
<thead>
<tr>
<th></th>
<th>Study site</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nashville</td>
<td>Rochester</td>
<td>Cincinnati</td>
<td>Seattle</td>
<td>Houston</td>
</tr>
<tr>
<td>Refused</td>
<td>113 (37)</td>
<td>71 (38)</td>
<td>144 (34)</td>
<td>84 (30)</td>
<td>174 (45)</td>
<td>139 (22)</td>
</tr>
<tr>
<td>MD refused</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>28 (10)</td>
<td>1 (0.3)</td>
<td>11 (2)</td>
</tr>
<tr>
<td>Language barrier</td>
<td>30 (10)</td>
<td>13 (7)</td>
<td>12 (3)</td>
<td>62 (22)</td>
<td>0 (0)</td>
<td>221 (34)</td>
</tr>
<tr>
<td>Missed</td>
<td>25 (8)</td>
<td>30 (16)</td>
<td>26 (6)</td>
<td>27 (10)</td>
<td>2 (0.5)</td>
<td>139 (22)</td>
</tr>
<tr>
<td>No stool</td>
<td>93 (31)</td>
<td>31 (17)</td>
<td>205 (48)</td>
<td>18 (6)</td>
<td>49 (13)</td>
<td>134 (21)</td>
</tr>
<tr>
<td>No saliva</td>
<td>31 (10)</td>
<td>34 (18)</td>
<td>27 (6)</td>
<td>53 (19)</td>
<td>121 (21)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No specimens</td>
<td>8 (3)</td>
<td>9 (5)</td>
<td>10 (2)</td>
<td>9 (3)</td>
<td>38 (10)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Unknown</td>
<td>2 (0.7)</td>
<td>0 (0)</td>
<td>1 (0.2)</td>
<td>0 (0)</td>
<td>3 (0.8)</td>
<td>1 (0.2)</td>
</tr>
</tbody>
</table>
Figure 1. Flow diagram of subject participation

Flow diagram of subject participation in the study, including eligibility assessment, enrollment, specimen collection, testing for norovirus, and final determination of norovirus status. Panel A (top): The process of AGE case inclusion. Panel B (bottom): The process of control inclusion. In each panel, the left-hand column of boxes indicate subjects included at each study step; the right-hand column of boxes detail the subjects excluded from each study step and the primary reasons for exclusion.
Figure 2. FUT2 secretors in norovirus-free controls and norovirus AGE cases, by norovirus genotype

Stacked bar chart representing the percentage of FUT2 “secretors” or “non-secretors” in norovirus AGE cases compared to the norovirus-free healthy controls. The first eight individual bars from the left represent the secretor status distribution of all 302 norovirus AGE cases infected with a specific norovirus genogroup, genotype or strain. The most commonly detected norovirus genotype, GII.4, was detected in 155 AGE cases. The GII.4 strains identified - New Orleans, Sydney, and Den Haag - were only detected in FUT2 secretors (p<0.001, compared to the frequency of secretors in controls). The second most common norovirus genotype, GII.6, also exhibited a secretor specificity (p=0.01). Noroviruses GII.1, GII.7, GI, and others, occurred less frequently and did not display secretor specificity. The “Other GII” category represents four less commonly detected norovirus GII genotypes (GII.2, GII.3, GII.8, and GII.13) and one case co-infected with GII.1 and GI.6. FUT2 secretors represented 76% of the norovirus-free healthy
controls (bar, far right), consistent with the expected population frequency. Not depicted are the 52 asymptomatic norovirus carriers who were enrolled as healthy controls. The distribution of strains, and secretor specificity of strains carried asymptptomatically was consistent with the pattern found in norovirus AGE (see text).
Figure 3. PCA of Morrow data set using AIMs
Figure 4. Triangle plot of STRUCTURE results
Figure 5. Distribution of self-identified ancestry in AGE cases and controls
Figure 6. Distribution of final genetic ancestry classifications in AGE cases and controls
Figure 7. Differences in sample collection by study site
Appendix A: Annotated Data Collection form for AGE Cases

NEW VACCINE SURVEILLANCE NETWORK - GASTROENTERITIS SURVEILLANCE YEAR 7 ANNOTATED CRF

Screening Log
Caseid (8 characters; first digit must be an E or S [E=enrolled, S=screened only]; digits 2-4 can be a letter or number; digits 5-8 must be numeric [0-9])
Subject ID number ________ (if enrolled)
Screening ID number ___________ (if not enrolled)

Study site (1 digit numeric; must be 1, 2 or 3; list box)
Site
1- Vanderbilt
2- Rochester
3- Cincinnati
4- Seattle
5- Houston
6- Kansas City

Provider (1 digit numeric; must be 1, 2 or 3; list box)
Provider Type
1- Inpatient
2- ED

Hospital (1 digit numeric; must be 1, 2, 3, 4 or 5; list box)
Hospital
1- CCHMC
2- RGH
4- SMH
5- VCH
6- SCH
7- TCH
8- CMH

abiniscr (up to 3 characters)
Abstractor Initials ________

abiniscr2_9 (up to 3 characters)
2nd abstractor Initials ________

deiniscr (up to 3 characters)
Data entry Initials ________

Scrdate (date field, mm/dd/yyyy, range from 11/01/2011 to 10/31/2012.)
Screen Date

admitdate (date field, mm/dd/yyyy, range from 10/22/2011 to 10/31/2012 check: admitdate should be < or = to scrdate. alert if not)
Admission / Visit Date

admittime (24 hr clock)
Admission / Visit Time

dob (date field, mm/dd/yyyy)
Date of birth

Agecalc (Calculated age, Age_calc = visitdate-Birthdate;
If visitdate-birthdate<14 days or >=11 years Flag “outside eligible age range”)
Calculated Age

Agedays (1 or 2 digit numeric, value range >14 days and <4015 days)
Age in days □□

Agemonths (1 or 2 digit numeric, value range >0 and <132 months)
Age in months □□

Ageyears (1 or 2 digit numeric, value range >0 years and <11 years)
Age in years □□

Insurance (1 digit numeric; must be 1, 2, 3, 4 or 8; list box)
1 Public
2 Private
3 Both
4 None, self pay
8 Unknown

Sex (1 digit numeric; must be 1, 2 or 8; list box)
1 Male
2 Female
8 Unknown

Race, Race2_8 (1 digit numeric; must be 1, 2, 3, 4, 5, 6, 7 or 8; list box)
Race
1 -- White
2 -- Black/ African American
3 -- American Indian/ Alaska Native
4 – Asian
5 -- Native Hawaiian/ Other Pacific Islander
6 – Hispanic
7-- Other
8 – Unknown

Adxd1 …… Adxd10 (characters 1-8)
Adiag1 …… Adiag10 (enable only if adxd1-adxd10=8. characters; field length ≤100 characters)

- adxd1 or adiag1
- adxd2 or diad2
- adxd3 or adiag3
- adxd4 or adiag4
- adxd5 or adiag5
- adxd6 or adiag6
- adxd7 or adiag7
- adxd8 or adiag8
- adxd9 or adiag9
- adxd10 or adiag10

Exclusion Criteria:
transfer (1 digit numeric; must be 0 or 1)
Transfer from another hospital after admission > 48 hours
0 --No
1 – Yes
8--Unknown

noninfect (1 digit numeric; must be 0 or 1)
Has non-infectious or other identifiable cause of symptom
0 --No
1 -- Yes
8--Unknown

immcomp (1 digit numeric; must be 0 or 1)
Child is immunocompromised
0 --No
1 – Yes
8--Unknown

prevenroll (1 digit numeric; must be 0 or 1)
Previously enrolled for same episode
0 --No
1 -- Yes
8--Unknown

elig (Auto calculated)
If Exclusion criteria 1-4 = No or unknown, then Eligible_Auto=Yes; Else, Eligible_Auto=No
If Eligible_Auto=No, Flag “Record should not be entered”
Eligible Auto_Calculated
Consent (1 digit numeric; must be 1, 2, 3, 4, 5, 6 or 7; mandatory; list box)
1 – Yes x
2 – Refused
3 – Discharged
4 – No Parent/ Legal Guardian
5 – MD Refused
6 – Missed Due to Time
7 – Parent/ Legal Guardian does not speak English

Consentini_9
Who consented subject __ __

Consentini2_9
Who consented subject (2nd person) __ __ __

(Enroll; 1 digit numeric; must be 0 or 1; Condition: requires Eligible_Auto= 1, and consent =1)

interview
0—No (If no, go to whynotenrolled)
1-- Yes

Whynotenrolled (digit numeric; must be 1, 2, 3; mandatory if interview =0)
Reason not enrolled
1 – Discharged
2 – Missed
3 – Other ________________

68
Parent Interview

intvwdate (date field: mm/dd/yyyy; 11/01/2011 to 10/31/2012 check: intvwdate should be > or =
scrdate. alert if not.) IF provider=2 and this field is empty, value is propagated with scrdate in
screening section.
Date of Interview □□/□□/□□□□

intvwini (up to 3 characters)
Interviewer Initials □□□

intvwini2_9 (up to 3 characters)
2nd interviewer initials □□□

deinicsr (up to 3 characters)
Data entry Initials □□□

dedatecsr (date field: mm/dd/yyyy; must be after Oct 31, 2010)
Data entry date □□/□□/□□□□

relation (up to 1 digit numeric; values= 0-9)
1. What is your relationship to the child?
   1--Mother
   2--Stepmother
   3--Father
   4--Stepfather
   5--Grandmother
   6--Grandfather
   7--Aunt
   8--Uncle
   9--Other (If other, go to 1a)
   0--Refused/No Response

   relothsp (Character ≤ 50 spaces)
1a. If other, please specify_________________

daysill (numeric up to 2 digits each; 88 = unknown, 99 = refused/ no response)
2. How long has your child been sick with this diarrhea and/or vomiting illness,
   including today?
□□Days

diarrh (1 digit numeric; must be 0, 1, 8 or 9)
3. Has your child had diarrhea during this illness?
   0-- No
   1--Yes (If yes, go 3a, 3b, 3c)
   8--Unknown
   9-- Refused
diarrhstart (numeric up to 2 digits; 88 = unknown, 99 = refused/ no response)
3a. How many days ago did it begin, including today?
□□Days

diarrhdays (numeric up to 2 digits; 88 = unknown, 99 = refused/ no response)
3b. How many days did your child have diarrhea, including today?
□□Days

diarrhepiso (numeric up to 2 digits; 88 = unknown, 99 = refused/ no response)
3c. What was the greatest number of episodes in 24 hours?
□□Episodes

vomit (1 digit numeric; must be 0, 1, 8 or 9)
4. Has your child had vomiting during this illness?
0-- No
1--Yes (If yes, go 4a, 4b, 4c)
8--Unknown
9-- Refused

vomitstart (numeric up to 2 digits each; 88 = unknown, 99 = refused/ no response)
4a. How many days ago did it begin, including today?
□□Days

vomitdays (numeric up to 2 digits each; 88 = unknown, 99 = refused/ no response)
4b. How many days did your child have vomiting, including today?
□□Days

vomitepiso (numeric up to 2 digits each; 88 = unknown, 99 = refused/ no response)
4c. What was the greatest number of episodes in 24 hours?
□□Episodes

fever (1 digit numeric; must be 0, 1, 8, or 9)
5. Has your child had a fever during this illness?
0-- No
1--Yes (If yes, go 5a, 5b, 5c, 5d)
8--Unknown
9-- Refused

feverstart (numeric up to 2 digits; 88 = unknown, 99 = refused/ no response)
5a. How many days ago did it begin, including today?
□□Days

feverdays (numeric up to 2 digits; 88 = unknown, 99 = refused/ no response)
5b. How many days did your child have a fever, including today?
□□Days
5c. What was the highest temperature you measured?

□□□°F (range ≥98, ≤105)

measureby (1 digit numeric; must be 1, 2, 3, 4, 8 or 9)

5d. How was the highest temperature measured?

1-- Rectal
2-- Axillary
3-- Oral
4-- Other (If other go to 5e)
8-- Unknown
9-- Refused

5e. If other, please specify __________________________

6. In the past 7 days, did your child touch or have contact with any of the following animals? Please select all that apply.

Boxes with True/False where False=0 and True=1; at least 1 value must be entered; If ContactNone, ContactUnk, or ContactRef = “True”, then deactivate remaining boxes

□ ContactNone – No animals
□ ContactUnk – Unknown if contact occurred
□ ContactRef – Refused/No Response
□ ContactBird – Bird
□ ContactCat – Cat/kitten
□ ContactDog – Dog/puppy
□ ContactChick – Chicken/chick
□ ContactPig – Pig
□ ContactHorse – Horse
□ ContactGoat – Goat or sheep
□ ContactCow – Cow
□ ContactReptile – Reptile
□ ContactAmph – Amphibian
□ ContactOther (if other, go to 6a) – Other type of animal

6a. If other, please specify:

animalfair (numeric, must be 0, 1, 8 or 9)

7. In the past 7 days, did your child visit a fair where there were animals, a farm with animals, or a petting zoo?

0— No
1— Yes
8—Unknown
9—Refused

foreigntravel (numeric, must be 0, 1, 8, or 9)
8. In the past 7 days, did your child travel outside of the United States?
0—No
1—Yes (if yes, go to 8a)
8—Unknown
9—Refused

tavelcountry1-5 (drop-down lists of countries)
8a. If yes, list the country or countries:
   travelcountry1 ______________________________
   travelcountry2 ______________________________
   travelcountry3 ______________________________
   travelcountry4 ______________________________
   travelcountry5 ______________________________

foreignvisitor (numeric, must be 0, 1, 8, or 9)
9. In the past 7 days, did your child have contact with anyone who recently visited or returned from a country outside of the United States?
0—No
1—Yes (if yes, go to 9a)
8—Unknown
9—Refused

visitorcountry1-5 (drop-down lists of countries)
9a. If yes, list the country or countries:
   visitorcountry1 ______________________________
   visitorcountry2 ______________________________
   visitorcountry3 ______________________________
   visitorcountry4 ______________________________
   visitorcountry5 ______________________________

behave (1 digit numeric; must be 1, 2, 3, 4, 5, 8 or 9)
10a. How is your child behaving currently? (Choose the most severe behavior, with seizure being the worst and normal being the least)
1—Normal
2—Less playful
3—Fussy/ Irritable
4—Lethargic/ Listless
5—Seizure
8—Unknown
9—Refused

eyes (1 digit numeric, must be 1, 2, 8, or 9)
10b. Do you think the child’s eyes are normal or sunken?
1—Normal
2—Sunken
8—Unknown
9—Refused

takefluids (1 digit numeric, must be 1, 2, 8, or 9)
10c. Is your child not interested in drinking, unable to drink, or refuses to feed?
1—Takes fluids normally
2—Not able to drink/refuses to feed
8—Unknown
9—Refused

skintest (1 digit numeric, must be 1, 2, 3, 8, or 9)
10d. Pinch test – lightly pull on the child’s skin to test for dehydration.
1—Normal – skin retracts immediately
2—Slowly – fold is visible for less than 2 seconds
3—Very slowly – the fold is visible for more than 2 seconds
8—Unknown
9—Refused

medconsult (1 digit numeric; must be 0, 1, 8 or 9)
11. Was a doctor, nurse, or other medical person consulted by telephone before this visit?
0—No
1—Yes
8—Unknown
9—Refused

oralrehydra (1 digit numeric; must be 0, 1, 8 or 9)
12. Did your child receive any oral rehydration fluids, such as Pedalyte, during this illness (before coming for this visit)?
0—No
1—Yes
8—Unknown
9—Refused

13. Has your child received any of the following therapies for this illness before this visit?

antibiotic (1 digit numeric; must be 0, 1, 8 or 9)
13a. Antibiotic
0—No
1—Yes
8—Unknown
9—Refused

irtherapy (1 digit numeric; must be 0, 1, 8 or 9)
13b. Intravenous Rehydration Therapy
0-- No
1--Yes
8--Unknown
9--Refused

14. Has your child received any of the following therapies for this illness during this visit?

antibioticdur (1 digit numeric; must be 0, 1, 8 or 9)
14a. Antibiotic
0-- No
1--Yes
8--Unknown
9--Refused

irtherapydur (1 digit numeric; must be 0, 1, 8 or 9)
14b. Intravenous Rehydration Therapy
0-- No
1--Yes
8--Unknown
9--Refused

15. Did anyone in the household miss any time from a job or business due to child’s illness?

0-- No
1—Yes (go to 15a)
8--Unknown
9 --Refused

timemissed (1 digit numeric; must be 0, 1, 8, or 9)
15a. If yes, how many days have been missed so far (total)?
0—Less than 1 day
1—1-2 days
2—3-4 days
3—5-6 days
4—7-8 days
5—greater than 8 days
8--Unknown
9 --Refused

wages (5 digit numeric; 8888=unknown, 9999=refused/no response)
15b. If yes, what is your estimate of the amount of wages lost? 0=none, 8888=unknown, 9999=refused/no response

school (1 digit numeric; must be 0, 1, 8, or 9)
16. Did the enrolled child miss any days of preschool or school due to illness?
0-- No
1—Yes (go to 16a and 16b)
8 --Unknown
9 --Refused

schooldays (2 digit numeric; 88=unknown, 99=refused/no response)
16b. If yes, how many days? 0=none, 77=less than 1 day, 88=unknown, 99=refused/no response

visitcost (4 digit numeric, value 0-9999; 8888=unknown, 9999=refused)
17. How much did you have to pay out of pocket for your child’s current visit to the hospital or ED? 0 = None, 8888 = Unknown, 9999 = Refused

18. Before arrival at hospital or ED, which costs did you incur related to this illness?
(Boxes with True/False where False=0 and True=1; at least 1 value must be entered; If Cost1None, Cost1Unk, or Cost1Ref = “True”, then deactivate remaining boxes)

☐ Cost1None – None
☐ Cost1Unk – Unknown
☐ Cost1Ref – Refused/No Response
☐ Cost1OTC – Over the counter medications
☐ Cost1Prescrip – Prescription medications
☐ Cost1Home – Home remedies
☐ Cost1Transp – Transportation
☐ Cost1Clinic – Clinic or healthcare facility visit
☐ Cost1Other – Other costs (if other, go to 18a)
Cost1OtherSp (Character ≤50 spaces)
18a. If other, specify:

18b. Estimated costs. 8888=Unknown, 9999=Refused
(4 digit numeric; 8888=unknown, 9999=refused/no response)
Est1OTC - Over the counter medications, estimated cost ____________
Est1Prescrip – Prescription medications, estimated cost ____________
Est1Home – Home remedies, estimated cost ____________
Est1Transp – Transportation, estimated cost ____________
Est1Clinic – Clinic or healthcare facility visit, estimated cost ____________
Est1Other – Other costs, estimated cost ____________

indirtotal (4 digit numeric; 8888=unknown, 9999=refused/no response)
18c. Please estimate the total costs (8888=unknown, 9999=refused/no response):

hispcr (1 digit numeric; must be 0, 1, 8 or 9)
19. Is your child Hispanic or Latino?
0-- No
1--Yes
8 --Unknown
9 --Refused
20. Is your child: (Check all that apply)
1 --White
2 --Black or African American
3 --American Indian/Alaska Native
4 --Asian
5 -- Native Hawaiian/Other Pacific Islander
6-- Other
7-- None
8 -- Unknown
9 --Refused/no response

Racecr1 thru racecr6 (1 digit numeric)

bearly (1 digit numeric; must be 0, 1, 8 or 9)

21. Was your child born more than one month early (more than 4 weeks early or less than 36 weeks gestation)?
0--No
1--Yes (if yes, go 21a)
8--Unknown
9-- Refused

bwks (numeric up to 2 digits; >4 and <18; 88 = unknown, 99 = refused/ no response)

21a. If yes, how many weeks early was your child born?
□□Weeks

Bwtlbs, bwtozs (2 numeric digits each, 88 = unknown, 99 = No response/refused)

22. What was your child’s birth weight?
bwtlbs  □□lbs.
bwtoz  □□ozs.

OR
Bwgms (numeric up to 4 digits, 8888 = unknown, 9999 = No response/refused)
bwgms □□□□gms.

breastf (1 digit numeric; must be 0, 1, 8 or 9)

23a. Did the mother ever breastfeed or pump breast milk to feed this child?
0--No (If no, go to 23b)
1--Yes (If yes, go to 23c, 23d)
8--Unknown
9—Refused
23b. If no, how old was the child the first time he or she ate any kind of food?
firstfoodwks (2 digit numeric; 66=less than 1 week, 77=has not eaten any foods, 88=unknown, 99=refused/no response)
  □□ Weeks
  OR
firstfoodmonths (2 digit numeric)
  □□ Months

23c. If yes, how many weeks or months did/has the mother breastfed or pumped milk to feed the child?
Breastfhowlongwks (2 digit numeric; 66=less than 1 week, 88=unknown, 99=refused/no response)
  □□ Weeks
  OR
Breastfhowlongmonths (2 digit numeric)
  □□ Months

breastfstill (1 digit numeric; must be 0, 1, 8 or 9)
23d. If yes, is the mother still currently breastfeeding or feeding pumped milk to this child?
  0—No (go to 23e)
  1—Yes
  8--Unknown
  9-- Refused

23e. If no, how old was the child the first time he or she drank liquids other than breast milk?
firstdrinkwks (2 digit numeric; 66=less than 1 week, 77=has not had other liquids, 88=unknown, 99=refused/no response)
  □□ Weeks
  OR
Firstdrinkmonths (2 digit numeric)
  □□ Months

daycare (1 digit numeric; must be 0, 1, 8 or 9)
24. Does your child attend any type of day care or preschool for more than 4 hours/week?
  0--No
  1—Yes (If yes, go to 23a)
  8--Unknown
  9-- Refused

dcnum (1 digit numeric; must be 1, 2, 3, 8 or 9)
24a. If yes, with how many unrelated children?
  1--Fewer than 6
  2 --6-12
  3 --More than 12
  8 --Unknown
  9—Refused
diapers (1 digit numeric; must be 0, 1, 8 or 9)
25. Does your child currently wear diapers?
   0-- No
   1--Yes
   8 --Unknown
   9 --Refused

diapershouse (2 digit numeric; value 0-99, 88=unknown, 99=refused/no response)
26. How many other children in the household currently wear diapers?

AGEoutside (1 digit numeric; must be 0, 1, 8 or 9)
27. Did your child come into contact with anyone outside the household with diarrhea or vomiting in the past week?
   0-- No
   1--Yes
   8 --Unknown
   9 --Refused

AGEinside (1 digit numeric; must be 0, 1, 8 or 9)
28. Did your child come into contact with anyone inside the household with diarrhea or vomiting in the past week?
   0-- No
   1--Yes
   8 --Unknown
   9 --Refused

HHincome (1 digit numeric, must be 1-9)
29. What was your household income last year (before taxes)?
   1 – Less than or equal to $25,000
   2 -- $25,001-$50,000
   3 -- $50,001- $75,000
   4 -- $75,001-$100,000
   5 -- $100,001-$125,000
   6 -- $125,001 - $150,000
   7 -- Over $150,000
   8 -- Unknown
   9 --Refused/no response

totpl (2 digit numeric; 88=unknown, 99=refused/no response)
30a. Including the index child, how many people live in your household? ______

30b. Please list all household members and whether they have also experienced diarrhea or vomiting in the previous 7 days:
<table>
<thead>
<tr>
<th>Household member’s relation to child</th>
<th>Age (years)</th>
<th>Sex</th>
<th>Did member have diarrhea or vomiting in past 7 days?</th>
<th>If yes, what medical care was sought for this illness?</th>
<th>Date symptoms began</th>
<th>Date symptoms ended</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. HHrelation1</td>
<td>HHage1</td>
<td>HHsex1</td>
<td>HHill1</td>
<td>HHcare1</td>
<td>HHstart1</td>
<td>HHstop1</td>
</tr>
<tr>
<td>2. HHrelation2</td>
<td>HHage2</td>
<td>HHsex2</td>
<td>HHill2</td>
<td>HHcare2</td>
<td>HHstart2</td>
<td>HHstop2</td>
</tr>
<tr>
<td>3. HHrelation3</td>
<td>HHage3</td>
<td>HHsex3</td>
<td>HHill3</td>
<td>HHcare3</td>
<td>HHstart3</td>
<td>HHstop3</td>
</tr>
<tr>
<td>4. HHrelation4</td>
<td>HHage4</td>
<td>HHsex4</td>
<td>HHill4</td>
<td>HHcare4</td>
<td>HHstart4</td>
<td>HHstop4</td>
</tr>
<tr>
<td>5. HHrelation5</td>
<td>HHage5</td>
<td>HHsex5</td>
<td>HHill5</td>
<td>HHcare5</td>
<td>HHstart5</td>
<td>HHstop5</td>
</tr>
<tr>
<td>6. HHrelation6</td>
<td>HHage6</td>
<td>HHsex6</td>
<td>HHill6</td>
<td>HHcare6</td>
<td>HHstart6</td>
<td>HHstop6</td>
</tr>
<tr>
<td>7. HHrelation7</td>
<td>HHage7</td>
<td>HHsex7</td>
<td>HHill7</td>
<td>HHcare7</td>
<td>HHstart7</td>
<td>HHstop7</td>
</tr>
<tr>
<td>8. HHrelation8</td>
<td>HHage8</td>
<td>HHsex8</td>
<td>HHill8</td>
<td>HHcare8</td>
<td>HHstart8</td>
<td>HHstop8</td>
</tr>
<tr>
<td>9. HHrelation9</td>
<td>HHage9</td>
<td>HHsex9</td>
<td>HHill9</td>
<td>HHcare9</td>
<td>HHstart9</td>
<td>HHstop9</td>
</tr>
<tr>
<td>10. HHrelation10</td>
<td>HHage10</td>
<td>HHsex10</td>
<td>HHill10</td>
<td>HHcare10</td>
<td>HHstart10</td>
<td>HHstop10</td>
</tr>
<tr>
<td>11. HHrelation11</td>
<td>HHage11</td>
<td>HHsex11</td>
<td>HHill11</td>
<td>HHcare11</td>
<td>HHstart11</td>
<td>HHstop11</td>
</tr>
<tr>
<td>12. HHrelation12</td>
<td>HHage12</td>
<td>HHsex12</td>
<td>HHill12</td>
<td>HHcare12</td>
<td>HHstart12</td>
<td>HHstop12</td>
</tr>
<tr>
<td>13. HHrelation13</td>
<td>HHage13</td>
<td>HHsex13</td>
<td>HHill13</td>
<td>HHcare13</td>
<td>HHstart13</td>
<td>HHstop13</td>
</tr>
<tr>
<td>14. HHrelation14</td>
<td>HHage14</td>
<td>HHsex14</td>
<td>HHill14</td>
<td>HHcare14</td>
<td>HHstart14</td>
<td>HHstop14</td>
</tr>
<tr>
<td>15. HHrelation15</td>
<td>HHage15</td>
<td>HHsex15</td>
<td>HHill15</td>
<td>HHcare15</td>
<td>HHstart15</td>
<td>HHstop15</td>
</tr>
</tbody>
</table>
HHrelation1-15: (Character ≤50 spaces)
Household member’s relation to child

HHage1-15: (3 digit numeric; 777= less than 1 year, 888=unknown, 999=refused/no response)
Age (years)

HHsex1-15: (1 digit numeric; value must be 1, 2, 8, or 9)
Sex
1—Male
2—Female
8—Unknown
9—Refused

HHill1-15: (1 digit numeric; value must be 0, 1, 8, or 9)
Did member have diarrhea or vomiting in past 7 days?
0—No
1—Yes
8—Unknown
9—Refused

HHcare1-15: (1 digit numeric; value must be 0, 1, 8, or 9)
If yes, what medical care was sought for this illness?
0—None
1—Hospitalized
2—ED visit
3—Outpatient visit
4—Telephone call with medical staff
8—Unknown
9—Refused

HHstart1-15: (date field, mm/dd/yyyy, range from 10/01/2011 to 10/31/2012 check: HHstart1-15 should be ≤ to scrdate. alert if not; 88/88/8888= unknown; 99/99/9999=refused)
Date symptoms began

HHstop1-15: (date field, mm/dd/yyyy, range from 10/24/2011 to 10/31/2012 check: HHstop1-15 should be ≤ to scrdate. alert if not; 77/77/7777 =ongoing; 88/88/8888= unknown; 99/99/9999=refused)
Date symptoms ended

momage_7 (2 digit numeric; 88=unkown, 99=refused/no response)
31. How old is the child’s biological mother? □□Years

mdegree_7 (numeric, 1 digit, legitimate values 0-5, 8,9)
32. What is the highest degree or diploma the biological mother has completed?
0 None
1 GED
2 HS Diploma
5 2-Year college degree (associate’s or technical degree)
3 4-Year college degree (bachelor’s degree)
4 Graduate degree (master’s, doctorate, medical, etc.)
8 Unsure/Unknown
9 Refused/No Response

Sickloc (numeric, 1 digit, legitimate values 1-9)
33. When your child is sick, where do you usually take your child?
1--Doctor’s Office
2--Public Health Clinic or Community Health Center
3--Hospital based Practice
4--Hospital Emergency Room
5--Urgent Care Center
6--Some Other Kind of Place
7--No Usual Place
8--Unknown
9—Refused

rotavac (1 digit numeric; must be 0, 1, 8 or 9)
34a. Has your child received a rotavirus vaccine?
0--No
1--Yes (If yes, go to 34b)
8--Unknown
9--Refused

rotavacnum (1 digit numeric; must be 0 to <=5, 8=Not Sure/Unsure 9=Refused)
34b. Number of doses ever received:
□ Doses

tcons (1 numeric digit; must be 0 or 1)
35. Consent for pathogen testing of stool specimen?
0 --No
1 --Yes

salivacons (1 numeric digit; must be 0 or 1)
36. Consent for pathogen testing of saliva specimen?
0 --No
1 --Yes
9 –Not applicable (child ≥5 years old)

followcons (1 numeric digit; must be 0 or 1)
37. Consent for follow-up of household AGE?
0 --No
1 –Yes
othercons (1 numeric digit; must be 0 or 1)
38. Consent for testing additional specimens collected by the hospital?
   0 --No
   1 --Yes

Stoolvisit (1 numeric digit, must be 0 or 1)
39. Was a stool specimen collected during visit?
   0—No
   1—Yes (If yes, go to 39a)

abinisp_9 (up to 3 characters)
If yes, by whom _ _ _

abinisp2_9 (up to 3 characters)
Secondary specimen collection initials _ _ _

salivavisit (1 numeric digit, must be 0 or 1)
40. Was a saliva specimen collected during visit?
   0—No
   1—Yes (If yes, go to 40a)
   9 —Not applicable (child ≥5 years old)

salivaini (up to 3 characters)
If yes, by whom _ _ _

salivaini2 (up to 3 characters)
Secondary specimen collection initials _ _ _
MEDICAL RECORD REVIEW

ALL PATIENTS

abinich (up to 3 characters)
Data abstractor initials □□□

abinich2_9 (up to 3 characters)
2nd abstractor initials □□□

deinich (up to 3 characters)
Data entry Initials □□□

dobch (date field, mm/dd/yyyy; 88/88/8888 = unknown check this against dob in screening log, if there is data in both places. have an alert appear at dobch if the two do not match.
1. DOB: □□/□□/□□□□

Sexch (1 digit numeric; mandatory; must be 1, 2 or 8)
2. Gender:
1--Male
2--Female
8 –Unknown

insurch (1 digit numeric; must be 0, 1 or 8)
3. Does the child have health insurance?
0 --No
1—Yes (If yes, go to number 3a)
8 --Unknown

insplanch (1 digit numeric; must be 0, 1, 2 or 8)
3a. If yes is the insurance plan private (commercial) or public (Medicaid)?
0 -- Public
1 -- Private
2 -- Both
8 -- Unknown if public or private

SUBJECTS ENROLLED IN THE ED ONLY
(Enable the following questions only when provider=2)

outvisitdat (date field, mm/dd/yyyy, range 11/01/2011 to 10/31/2012 check: should be = to scrdate, the date in screening section
IF this field is empty, and provider=2, propagate with scrdate when available.)
1. Visit Date □□/□□/□□□□

ddxdc1 ……… dxdC10
(characters, format= □□□□□□ 999.99=Unknown/Missing DC Summary)
dct1 ……….. dct10 (characters, format= □□□□□□)
ddiagC1_7…….ddiagC10_7 (number, format = □ allow values 1-9)
ddiagC1 …… ddiagC10 Enable only when ddiagC1_7-ddiagC10_7 = 9. (characters; field length ≤100 characters)

2. Discharge Diagnosis (ICD-9 code, CPT Code or descriptive text if codes are not available)
   1: ddxxC1 □□□□□ or dcept1 □□□□□ or ddiagC1_7 or ddiagC1____
   2: ddxxC2 □□□□□ or dcept2 □□□□□ or ddiagC2_7 or ddiagC2____
   3: ddxxC3 □□□□□ or dcept3 □□□□□ or ddiagC3_7 or ddiagC3____
   4: ddxxC4 □□□□□ or dcept4 □□□□□ or ddiagC4_7 or ddiagC4____
   5: ddxxC5 □□□□□ or dcept5 □□□□□ or ddiagC5_7 or ddiagC5____
   6: ddxxC6 □□□□□ or dcept6 □□□□□ or ddiagC6_7 or ddiagC6____
   7: ddxxC7 □□□□□ or dcept7 □□□□□ or ddiagC7_7 or ddiagC7____
   8: ddxxC8 □□□□□ or dcept8 □□□□□ or ddiagC8_7 or ddiagC8____
   9: ddxxC9 □□□□□ or dcept9 □□□□□ or ddiagC9_7 or ddiagC9____
  10: ddxxC10 □□□□□ or dcept10 □□□□□ or ddiagC10_7 or ddiagC10____

outadm (1 digit numeric; must be 0, 1 or 8)
3. Was the child admitted to a surveillance hospital ≤6 days of initial ED visit for AGE?
   0 --No
   1 --Yes (If yes, go to number 3a, 3b)
   8 --Unknown

outhosp (1 digit numeric; must be 1, 2, 3, 4 or 5)
3a. If yes, hospital to which child was admitted
   1- CCHMC
   2- RGH
   4- SMH
   5- VCH
   6- SCH
   7- TCH
   8- CMH

outadmdt (date field: mm/dd/yyyy; 11/01/2011 to 11/06/2012)
3b. Admission Date

surday (1 digit numeric; must be 0, 1 or 8))
3c. Was child admitted on a surveillance day?
   0 --No
   1 --Yes (If yes, go to number 3d)
   8 --Unknown

inelig (1 digit numeric; must be 0, 1 or 8)
3d. Other than being enrolled as an outpatient, is child eligible for inpatient surveillance study?
   0 --No (If no, go to 3e)
   1 --Yes (If yes, go to number 3f, 3g)
   8 --Unknown

   Inelig (1 digit numeric; must be 2 or 3)
3e. If No, indicate appropriate exclusion criteria:
1 --Transfer from another hospital after admission of > 48 hours
2 – Non-infectious or other identifiable cause of symptoms
3 --Immunocompromised

inpatid (8 characters; must start with E)
3f. If yes, inpatient subject ID (only if assigned) ____________

3g. Admission Diagnosis
adxdi ……. adxdC10 (characters, 1-8)
diagC1 ……. diagC10 (characters; field length ≤100 characters)
Admission Diagnosis (descriptive if diagnosis is not in list)
adxdi or diagC1 _____________________________
adxdi2 or diagdC2 _____________________________
adxdi3 or diagC3 _____________________________
adxdi4 or diagC4 _____________________________
adxdi5 or diagC5 _____________________________
adxdi6 or diagC6 _____________________________
adxdi7 or diagdC7 _____________________________
adxdi8 or diagC8 _____________________________
adxdi9 or diagC9 _____________________________
adxdi0 or diagC10 _____________________________

Inpatient or Outpatient Admitted to a Hospital within 6 Days
Enable this section below when provider = 1 OR when outadm = 1
1. Inpatient visit dates

admitch (date field, mm/dd/yyyy, range from 11/01/2011 to 11/06/2012) If empty and outadm=1, propagate with outadmdt if available)
a. Admission Date: □□/□□/□□□□

ddtch (date field, mm/dd/yyyy, range from 11/01/2011 to 11/30/2012) check: should be > or = to admitch
b. Discharge Date: □□/□□/□□□□

billed23 (1 digit numeric)
2. Was this a 23 hours stay?

0-- No
1--Yes
8—Unknown

hicu (1 digit numeric)
3. Was this an ICU stay?

0-- No
1--Yes (If yes, go to 3a)
8--Unknown

hicudays (numeric up to 3 digits)
3a. Days in ICU □□□

houtcome (1 digit numeric)
4. Outcome:

0 -- Died
1 – Survived
2 -- Transferred (If transferred, go to 4a, 4b)
8 -- Unknown

htranshos (Character, up to 100)
4a. If transferred, name of receiving hospital________________

htranout
4b. If transferred, outcome:

0 -- Died
1 -- Survived
8 -- Unknown

ddxIO1 ........ ddxIO10 (characters, format= □□□□□□□□ 999.99=Unknown/Missing DC Summary)
ddxIOC1_7….ddxIOC10_7 (numeric, 1 digit, 1-9)
ddiagIO1 ........ ddiagIO10 :enable only when ddxIOC1_7-ddxIOC10 = 9 (characters; field length ≤100 characters)

5. Discharge Diagnosis (ICD9 coded or descriptive text if ICD9 code is not available)

ddxIO1 □□□□□□□□ or ddxIOC1_7 or ddiagIO1________________
ddxIO2 □□□□□□□□ or ddxIOC2_7 or ddiagIO2________________
ddxIO3 □□□□□□□□ or ddxIOC3_7 or ddiagIO3________________
ddxIO4 □□□□□□□□ or ddxIOC4_7 or ddiagIO4________________
ddxIO5 □□□□□□□□ or ddxIOC5_7 or ddiagIO5________________
ddxIO6 □□□□□□□□ or ddxIOC6_7 or ddiagIO6________________
ddxIO7 □□□□□□□□ or ddxIOC7_7 or ddiagIO7________________
ddxIO8 □□□□□□□□ or ddxIOC8_7 or ddiagIO8________________
ddxIO9 □□□□□□□□ or ddxIOC9_7 or ddiagIO9________________
ddxIO10 □□□□□□□□ or ddxIOC10_7 or ddiagIO10________________
RESEARCH LAB RESULTS

deinisp_9
Data entry initials □□□

specimencol (1 numeric digit; must be 0, 1,)
1. Was a stool specimen collected?
0--No
1--Yes

coldat (date field, mm/dd/yyyy; legitimate values: 11/01/2011 to 11/10/2012: coldat should be > or = scrdate. alert if not.
1a. Date specimen collected: □□/□□/□□□□

specimentest (1 numeric digit)
1b. Specimen tested?
0 –No (If no, go to 23c)
1—Yes (If yes go 23e, 23f)

treason (1 numeric digit)
1c. Reason test not done
1-- Inadequate specimen, unable to test
2-- Specimen lost in transit
3--Other (If other go to 23d)

othreason (character field, length ≤ 50 characters)
1d. If other specify ________________

tdat (Date field; mm/dd/yyyy, legitimate values 11/01/2011 to 11/30/2012 check: tdat should be > or = to coldat. alert if not)
1e. Date specimen tested □□/□□/□□□□

tresult (1 numeric digit; must be 0, 1, or 2 )
1f. Rotavirus results
0 -- Negative
1 – Positive
2 – Inconclusive

visualpos (1 numeric digit; must be 0, 1, or 2; activate when tresult=0 or 2)
1g. If EIA result obtained by using a plate reader was negative or inconclusive, was the specimen visually positive?
0 -- No
1 – Yes
2 – Inconclusive

stoolblood (1 numeric digit; must be 0, 1, or 8 )
1h. Did stool contain visible blood?
0 -- No
1 – Yes
8 – Unknown

salivacol (1 numeric digit; must be 0, 1,)

2. Was a saliva specimen collected?
0--No
1--Yes (go to 2a)
9 – Not applicable (child ≥5 years old)

salcoldata (date field, mm/dd/yyyy; legitimate values: 11/01/2011 to 11/10/2012: coldata should be ≥ or = scrdate. alert if not.)

2a. Date specimen collected: □□/□□/□□□□

salivatest (1 numeric digit)

2b. Specimen tested?
0 – No (If no, go to 2c)
1—Yes (If yes go 2e, 2f)

saltreason (1 numeric digit)

2c. Reason test not done
1-- Inadequate specimen, unable to test
2-- Specimen lost in transit
3--Other (If other go to 2d)

saltothereason (character field, length ≤ 50 characters)
2d. If other specify ____________

saltdate (Date field; mm/dd/yyyy, legitimate values 11/01/2011 to 11/30/2012 check: tdate should be ≥ or = to scrdate. alert if not)

2e. Date specimen tested □□/□□/□□□□

saltresult (1 numeric digit; must be 0, 1, or 2 )
2f. Secretor status result
0 – FUT2 absent
1 – FUT2 present
2 – Inconclusive

3. Were any blood product specimens collected during this child’s normal course of treatment?

Whole blood? bloodcol (1 numeric digit; must be 0, 1, 8)
0--No
1--Yes
8 – Unknown
bloodcoldat (date field, mm/dd/yyyy; range 11/01/2011 to 11/10/2012):
Date collected: □□/□□/□□□□

Serum? serumcol (1 numeric digit; must be 0, 1, 8)
0--No
1--Yes
8 –Unknown

sercoldat (date field, mm/dd/yyyy; range 11/01/2011 to 11/10/2012):
Date collected: □□/□□/□□□□

Plasma? plasmacol (1 numeric digit; must be 0, 1, 8)
0--No
1--Yes
8 –Unknown

plasmacoldat (date field, mm/dd/yyyy; range 11/01/2011 to 11/10/2012):
Date collected: □□/□□/□□□□

Other type of blood product? Othercol (1 numeric digit; must be 0, 1, 8)
0--No
1--Yes
8 –Unknown

Othercoldat (date field, mm/dd/yyyy; range 11/01/2011 to 11/10/2012):
Date collected: □□/□□/□□□□
NEW VACCINE SURVEILLANCE NETWORK - GASROENTERITIS SURVEILLANCE

ROTAVIRUS VACCINATION VERIFICATION

Abinivv_9

Abstractor initials □□□

abinivv2_9

2nd abstractor initials □□□

deinivv_9

Data entry initials □□□

Provider Information (required)

rvacver_7 (range=0,1,2,7,8,
1. Number of rotavirus vaccinations recorded.
 0 □  1 □  2 □  3 □  7 □ N/A (child < 6 weeks) 8 □ Unknown/Unable to review chart

2. Rotavirus Dose 1
rpver1_7 (range=1-4) rvac1dat_7 (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)
 1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
 2 □ None in record
 3 □ Patient not in practice
 4 □ Unable to review chart

rvac1type_9

Vaccine type (range=1, 2, 8)
 1 □ RotaTeq (Merck)
 2 □ Rotarix (GSK)
 8 □ Unknown

3. Rotavirus Dose 2
rpver2_7(range=1-4) rvac2dat_7 (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)
 1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
 2 □ None in record
 3 □ Patient not in practice
 4 □ Unable to review chart

rvac2type_9

Vaccine type (range=1, 2, 8)
 1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

4. Rotavirus Dose 3
rpver3_7 (range=1-4) vac3dat_7 (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)

1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart

rvac3type_9
Vaccine type (range=1, 2, 8)
1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

Registry Information (optional)

rvacver_7r (range=0,1,2,7,8)
1. Number of rotavirus vaccinations recorded.
0 □ 1 □ 2 □ 3 □ 7 □ N/A (child < 6 weeks) 8 □ Unknown/Unable to review chart

2. Rotavirus Dose 1
rpver1_7r (range=1-4) rvac1dat_7r (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)

1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart

rvac1type_9r
Vaccine type (range=1, 2, 8)
1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

3. Rotavirus Dose 2
rpver2_7r (range=1-4) rvac2dat_7r (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)

1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart

rvac2type_9r
Vaccine type (range=1, 2, 8)

1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

4. Rotavirus Dose

Vaccine Date (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)

1 □ Vaccination verified
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart
FOLLOW-UP QUESTIONNAIRE

followup (1 digit numeric; must be 0 or 1)
Enable the rest of the following questions only if followup=1
Did the household complete the follow-up questionnaire?
0—No
1—Yes

fupdate (date field, mm/dd/yyyy, range from 11/01/2011 to 11/30/2012; fupdate should occur after scrdate)
Date follow-up was conducted

fupini (up to 3 characters)
Interviewer initials □□□

fupini2(up to 3 characters)
2nd interviewer initials □□□

fupdeini (up to 3 characters)
Data entry Initials □□□

stillsick (1 digit numeric; must be 0, 1, 8 or 9)
1. Is the enrolled child still experiencing any diarrhea or vomiting?
0—No (If no, go to 1a)
1—Yes
8—Unknown
9—Refused

stopdate (date field, mm/dd/yyyy, range from 11/01/2011 to 11/30/2012. 77/77/7777=ongoing, 88/88/8888=unknown)
1a. Date symptoms stopped

2. Have you spent any additional money to treat the index child’s stomach illness??
(Boxes with True/False where False=0 and True=1; at least 1 value must be entered; If Cost1None, Cost1Unk, or Cost1Ref = “True”, then deactivate remaining boxes)
□ Cost2None – None
□ Cost2Unk – Unknown
□ Cost2Ref – Refused/No Response
□ Cost2OTC – Over the counter medications
□ Cost2Prescrip – Prescription medications
□ Cost2Home – Home remedies
□ Cost2Transp – Transportation
□ Cost2Clinic – Clinic or healthcare facility visit
□ Cost2Other – Other costs (if other, go to 2a)

Cost2OtherSp (Character ≤50 spaces)
2a. If other, specify:

2b. Estimated costs.  8888=Unknown, 9999=Refused
(4 digit numeric; 8888=unknown, 9999=refused/no response)

**Est2OTC** – Over the counter medications, estimated cost ______________

**Est2Prescrip** – Prescription medications, estimated cost ______________

**Est2Home** – Home remedies, estimated cost ______________

**Est2Transp** – Transportation, estimated cost ______________

**Est2Clinic** – Clinic or healthcare facility visit, estimated cost ______________

**Est2Other** – Other costs, estimated cost ______________

*indirtotalfup (4 digit numeric; 8888=unknown, 9999=refused/no response)*

2c. Please estimate the total costs (8888=unknown, 9999=refused/no response):

newvisit (1 digit numeric, must be 0, 1, 8, or 9)

3. Did the enrolled child visit the hospital or ED again for the same illness?

0—No

1—Yes (if yes, go to 3a)

8 --Unknown

9 —Refused

*newvisitcost (4 digit numeric; 8888=unknown, 9999=refused/no response)*

3a. Out of pocket cost for the additional visit(s):

*fuptimemissed (1 digit numeric; must be 0, 1, 8, or 9)*

4. Since the index child left the hospital or ED, did anyone in the household miss any time from a job or business due to illness in the household?

0—No

1—Yes (go to 4a and 4b)

8 --Unknown

9 --Refused

*fuptimetotal (1 digit numeric, must be 0-5, 8, or 9)*

4a. If yes, how many days have been missed so far (total)?

0—Less than 1 day

1—1-2 days

2—3-4 days

3—5-6 days

4—7-8 days

5—greater than 8 days

8 --Unknown

9 —Refused

*fupwages (5 digit numeric; 8888=unknown, 9999=refused/no response)*

4b. If yes, what is your estimate of the amount of wages lost? 0=none, 8888=unknown, 9999=refused/no response
fupschool (1 digit numeric; must be 0, 1, 8, or 9)
5. Since the index child left the hospital or ED, did the enrolled child miss any days of preschool or school due to illness?
0—No
1—Yes (go to 5a)
8—Unknown
9—Refused

fupschooldays (2 digit numeric; 88=unknown, 99=refused/no response)
5a. If yes, how many days? 0=none, 77=less than 1 day, 88=unknown, 99=refused/no response

6. Please list all household members other than the index child and whether they have experienced diarrhea or vomiting since the index child’s date of enrollment:

<table>
<thead>
<tr>
<th>Household member’s relation to child</th>
<th>Age (years)</th>
<th>Sex</th>
<th>Did member have diarrhea or vomiting since child’s enrollment date</th>
<th>If yes, what medical care was sought for this illness?</th>
<th>Date symptoms began</th>
<th>Date symptoms ended</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. FUPrelation1</td>
<td>FUPage1</td>
<td>FUPsex1</td>
<td>FUPill1</td>
<td>FUPcare1</td>
<td>FUPstart1</td>
<td>FUPstop1</td>
</tr>
<tr>
<td>2. FUPrelation2</td>
<td>FUPage2</td>
<td>FUPsex2</td>
<td>FUPill2</td>
<td>FUPcare2</td>
<td>FUPstart2</td>
<td>FUPstop2</td>
</tr>
<tr>
<td>3. FUPrelation3</td>
<td>FUPage3</td>
<td>FUPsex3</td>
<td>FUPill3</td>
<td>FUPcare3</td>
<td>FUPstart3</td>
<td>FUPstop3</td>
</tr>
<tr>
<td>4. FUPrelation4</td>
<td>FUPage4</td>
<td>FUPsex4</td>
<td>FUPill4</td>
<td>FUPcare4</td>
<td>FUPstart4</td>
<td>FUPstop4</td>
</tr>
<tr>
<td>5. FUPrelation5</td>
<td>FUPage5</td>
<td>FUPsex5</td>
<td>FUPill5</td>
<td>FUPcare5</td>
<td>FUPstart5</td>
<td>FUPstop5</td>
</tr>
<tr>
<td>6. FUPrelation6</td>
<td>FUPage6</td>
<td>FUPsex6</td>
<td>FUPill6</td>
<td>FUPcare6</td>
<td>FUPstart6</td>
<td>FUPstop6</td>
</tr>
<tr>
<td>7. FUPrelation7</td>
<td>FUPage7</td>
<td>FUPsex7</td>
<td>FUPill7</td>
<td>FUPcare7</td>
<td>FUPstart7</td>
<td>FUPstop7</td>
</tr>
<tr>
<td></td>
<td>FUPrelation</td>
<td>FUPage</td>
<td>FUPsex</td>
<td>FUPill</td>
<td>FUPcare</td>
<td>FUPstart</td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>8</td>
<td>FUPrelation8</td>
<td>FUPPage8</td>
<td>FUPsex8</td>
<td>FUPill8</td>
<td>FUPcare8</td>
<td>FUPstart8</td>
</tr>
<tr>
<td>9</td>
<td>FUPrelation9</td>
<td>FUPPage9</td>
<td>FUPsex9</td>
<td>FUPill9</td>
<td>FUPcare9</td>
<td>FUPstart9</td>
</tr>
<tr>
<td>10</td>
<td>FUPrelation10</td>
<td>FUPPage10</td>
<td>FUPsex10</td>
<td>FUPill10</td>
<td>FUPcare10</td>
<td>FUPstart10</td>
</tr>
<tr>
<td>11</td>
<td>FUPrelation11</td>
<td>FUPPage11</td>
<td>FUPsex11</td>
<td>FUPill11</td>
<td>FUPcare11</td>
<td>FUPstart11</td>
</tr>
<tr>
<td>12</td>
<td>FUPrelation12</td>
<td>FUPPage12</td>
<td>FUPsex12</td>
<td>FUPill12</td>
<td>FUPcare12</td>
<td>FUPstart12</td>
</tr>
<tr>
<td>13</td>
<td>FUPrelation13</td>
<td>FUPPage13</td>
<td>FUPsex13</td>
<td>FUPill13</td>
<td>FUPcare13</td>
<td>FUPstart13</td>
</tr>
<tr>
<td>14</td>
<td>FUPrelation14</td>
<td>FUPPage14</td>
<td>FUPsex14</td>
<td>FUPill14</td>
<td>FUPcare14</td>
<td>FUPstart14</td>
</tr>
<tr>
<td>15</td>
<td>FUPrelation15</td>
<td>FUPPage15</td>
<td>FUPsex15</td>
<td>FUPill15</td>
<td>FUPcare15</td>
<td>FUPstart15</td>
</tr>
</tbody>
</table>

fuprelation1-15: (Character ≤50 spaces)
Household member’s relation to child

fupage1-15: (3 digit numeric; 777=less than 1 year, 888=unknown, 999=refused/no response)
Age (years)

fupsex1-15: (1 digit numeric; value must be 1, 2, 8, or 9)
Sex
1—Male
2—Female
8—Unknown
9—Refused

fupill1-15: (1 digit numeric; value must be 0, 1, 8, or 9)
Did member have diarrhea or vomiting since child’s enrollment date?
0—No
1—Yes
8—Unknown
9—Refused

fupcare1-15: (1 digit numeric; value must be 0, 1, 8, or 9)
If yes, what medical care was sought for this illness?

0—None
1—Hospitalized
2—ED visit
3—Outpatient visit
4—Telephone call with medical staff
8—Unknown
9—Refused

fupstart1-15: (date field, mm/dd/yyyy, range from 10/01/2011 to 11/30/2012 check: fupstart1-15 should be ≥ to scrdate. alert if not; 88/88/8888= unknown; 99/99/9999=refused)
Date symptoms began

fupstop1-15: (date field, mm/dd/yyyy, range from 10/24/2011 to 11/30/2012 check: fupstop1-15 should be ≥ to scrdate. alert if not; 77/77/7777=ongoing, 88/88/8888= unknown; 99/99/9999=refused)
Date symptoms ended

HHvisitcost (4 digit numeric; 8888=unknown, 9999=refused/no response)

7. If any household members other than the index child visited the ED or hospital due to a recent diarrhea or vomiting illness, what was the out-of-pocket cost of that visit?

8. Since the index child left the hospital, which costs have you incurred due to vomiting or diarrhea illness in other members of the household?
   (Boxes with True/False where False=0 and True=1; at least 1 value must be entered; If CostHHNone, CostHHUnk, or CostHHRef = “True”, then deactivate remaining boxes)
   □ CostHHNone – None
   □ CostHHUnk – Unknown
   □ CostHHRef – Refused/No Response
   □ CostHHOTC - Over the counter medications
   □ CostHHPrescrip – Prescription medications
   □ CostHHHome – Home remedies
   □ CostHHTransp – Transportation
   □ CostHHClinic – Clinic or healthcare facility visit
   □ CostHHOther – Other costs (if other, go to 8a)

CostHHOtherSp (Character ≤50 spaces)
8a. If other, specify:

8b. Estimated costs. 8888=Unknown, 9999=Refused
   (4 digit numeric; 8888=unknown, 9999=refused/no response)
   EstHHOTC - Over the counter medications, estimated cost ______________
   EstHHPrescrip – Prescription medications, estimated cost ______________
   EstHHHome – Home remedies, estimated cost ______________
   EstHHTransp – Transportation, estimated cost ______________
   EstHHClinic – Clinic or healthcare facility visit, estimated cost ______________
EstHHOther – Other costs, estimated cost ______________

indirtotalHH (4 digit numeric; 8888=unknown, 9999=refused/no response)

8c. Please estimate the total costs (8888=unknown, 9999=refused/no response):

timestamp
Date and time of last edit
Appendix B: Annotated Data Collection form for Healthy Controls
NEW VACCINE SURVEILLANCE NETWORK - HEALTHY CONTROL SURVEILLANCE YEAR 7 ANNOTATED CRF

Screening Log
Caseid (8 characters; must start with E or S [E=enrolled, S=screened only]; digits 2-4 can be a letter or number; digits 5-8 must be numeric [0-9])

Subject ID number (if enrolled)
Screening ID number (if not enrolled)

Studysite (1 digit numeric; must be 1, 2 or 3; list box)
Site
1- Vanderbilt
2- Rochester
3- Cincinnati
4- Seattle
5- Houston
6- Kansas City

Provider (1 digit numeric; must be 4; can be automatically selected)
Provider Type
4- Healthy Control

Hospital (1 digit numeric; must be 1, 2, 3, 4 or 5; list box)
Hospital
1- CCHMC
2- RGH
4- SMH
5- VCH
6- SCH
7- TCH
8- CMH

abiniscr (up to 3 characters)
Abstractor Initials

abiniscr2_9 (up to 3 characters)
2nd abstractor Initials

deiniscr (up to 3 characters)
Data entry Initials

Scrdate (date field, mm/dd/yyyy, range from 11/01/2011 to 10/31/2012.)
Screen Date
admitdate (date field, mm/dd/yyyy, range from 10/22/2011 to 10/31/2012 check: admitdate should be < or = to scrdate. alert if not)
Admission / Visit Date

admittime (24 hr clock)
Admission / Visit Time

dob (date field, mm/dd/yyyy)
Date of birth

Agecalc (Calculated age, Age_calc = visitdate-Birthdate; If visitdate-birthdate<14 days or >=11 years Flag “outside eligible age range”)
Calculated Age

Agedays (1 or 2 digit numeric, value range >14 days and <4015 days)
Age in days □□

Agemonths (1 or 2 digit numeric, value range >0 and <132 months)
Age in months □□

Ageyears (1 or 2 digit numeric, value range >0 years and <11 years)
Age in years □□

Insurance (1 digit numeric; must be 1, 2, 3, 4 or 8; list box)
1 Public
2 Private
3 Both
4 None, self pay
8 Unknown

Sex (1 digit numeric; must be 1, 2 or 8; list box)
1 Male
2 Female
8 Unknown

Race, Race2_8 (1 digit numeric; must be 1, 2, 3, 4, 5, 6, 7 or 8; list box)
Race
1 -- White
2 -- Black/ African American
3 -- American Indian/ Alaska Native
4 – Asian
5 -- Native Hawaiian/ Other Pacific Islander
6 – Hispanic
7-- Other
8 – Unknown
Exclusion Criteria:

prevage (1 digit numeric; must be 0 or 1)
Child has had any diarrhea (loose stools), or vomiting on the day of the visit or in the 14 days preceding the visit
0 -- No
1 -- Yes
8 -- Unknown

prevari (1 digit numeric; must be 0 or 1)
Child has had any acute respiratory infection symptoms on the day of the visit or in the 3 days preceding the visit.
0 -- No
1 -- Yes
8 -- Unknown

immcomp (1 digit numeric; must be 0 or 1)
Child is immunocompromised
0 -- No
1 -- Yes
8 -- Unknown

elig (Auto calculated)
If Exclusion criteria 1-4 = No or unknown, then Eligible_Auto=Yes; Else, Eligible_Auto=No
If Eligible_Auto=No, Flag “Record should not be entered”
Eligible Auto_Calculated

Consent (1 digit numeric; must be 1, 2, 3, 4, 5, 6 or 7; mandatory; list box)
1 -- Yes
2 -- Refused
3 -- Discharged
4 -- No Parent/ Legal Guardian
5 -- MD Refused
6 -- Missed Due to Time
7 -- Parent/ Legal Guardian does not speak English

Consentini_9
Who consented subject ___

Consentini2_9
Who consented subject (2nd person) ___

(Enroll; 1 digit numeric; must be 0 or 1; Condition: requires Eligible_Auto= 1, and consent =1)

interview
0— No (If no, go to whynotenrolled)
1-- Yes

Why not enrolled (digit numeric; must be 1, 2, 3; mandatory if interview =0)
Reason not enrolled
1 – Discharged)
2 – Missed)
3 – Other______________
Parent Interview

intvwdate (date field: mm/dd/yyyy; 11/01/2011 to 10/31/2012 check: intvwdate should be > or = scrdate. alert if not.) IF this field is empty, value is propagated with scrdate in screening section.

Date of Interview □□/□□/□□□□

intvwini (up to 3 characters)
Interviewer Initials □□□

intvwini2_9 (up to 3 characters)
2nd interviewer initials □□□

deinicsr (up to 3 characters)
Data entry Initials □□□

dedatecsr (date field: mm/dd/yyyy; must be after Oct 31, 2011)
Data entry date □□/□□/□□□□

relation (up to 1 digit numeric; values= 0-9)

1. What is your relationship to the child?

1--Mother
2--Stepmother
3--Father
4--Stepfather
5--Grandmother
6--Grandfather
7--Aunt
8--Uncle
9--Other (If other, go to 1a)
0—Refused

relothsp (Character ≤ 50 spaces)
1a. If other, please specify_________________

dobch (date field, mm/dd/yyyy; 88/88/8888= unknown. check this against dob in screening log, if there is data in both places. have an alert appear at dobch if the two do not match.

2. DOB: □□/□□/□□□□

Sexch (1 digit numeric; mandatory; must be 1, 2 or 8)

3. Gender:
1--Male
2--Female
8 --Unknown

insurch (1 digit numeric; must be 0, 1 or 8)

4. Does the child have health insurance?
0 -- No
1—Yes (If yes, go to number 4a)
8 -- Unknown

insplanchn (1 digit numeric; must be 0, 1, 2 or 8)
4a. If yes is the insurance plan private (commercial) or public (Medicaid)?
0 -- Public
1 -- Private
2 -- Both
8 -- Unknown if public or private

arivis2wk (1 digit numeric; must be 0, 1, 8 or 9)
5. During the past two weeks has your child had symptoms of a respiratory illness such as cough, earache, nasal congestion, runny nose, shortness of breath, rapid or shallow breathing, sore throat, vomiting after cough, or wheezing?
0-- No
1—Yes (go to 5a)
8 -- Unknown
9 -- Refused

arivis2wkdt (88/88/8888=unknown, 99/99/9999=refused)
5a. If yes, what was the date of the most recent symptom?
_____/_____/_______

arivisit (1 digit numeric; must be 0, 1, 8 or 9)
6. Since November 1, 2011, has your child visited an emergency room or been hospitalized because of a fever, sore throat, earache, wheezing, influenza, RSV, or other breathing problem?
0-- No
1—Yes (go to 6a)
8 -- Unknown
9 -- Refused

arivisitdt (88/88/8888=unknown, 99/99/9999=refused)
6a. If yes, date (or month) ______/_____/_________

agevis2wk (1 digit numeric; must be 0, 1, 8 or 9)
7. During the past two weeks has your child had any diarrhea or vomiting?
0-- No
1—Yes (go to 7a)
8 -- Unknown
9 -- Refused

agevis2wkdt (88/88/8888=unknown, 99/99/9999=refused)
7a. If yes, what was the date of the most recent symptom? ___/___/___

agevisit (1 digit numeric; must be 0, 1, 8 or 9)
8. Since November 1, 2011, has your child visited an emergency room or been hospitalized because of diarrhea or vomiting?
   0—No
   1—Yes (go to 8a)
   8—Unknown
   9—Refused

agevisitdt (88/88/8888=unknown, 99/99/9999=refused)

8a. If yes, date (or month) ______/______/________

agevisitany (1 digit numeric; must be 0, 1, 8 or 9)

9. Has your child ever visited an emergency room or been hospitalized because of diarrhea or vomiting?
   0—No
   1—Yes (go to 9a)
   8—Unknown
   9—Refused

agevisitanymonth (1 or 2 digit numeric, value range >0 and <132 months; 888=unknown, 999=refused/no response)

9a. If yes, age of child during most recent visit (in months) ______

10. Has a doctor ever told you that your child has...?

cheme (1 digit numeric; must be 0, 1, 8 or 9)
   Blood Disorder?
   0—No
   1—Yes
   8—Unknown
   9—Refused

cancer (1 digit numeric; must be 0, 1, 8 or 9)
   Cancer?
   0—No
   1—Yes
   8—Unknown
   9—Refused

cdiab (1 digit numeric; must be 0, 1, 8 or 9)
   Diabetes Mellitus?
   0—No
   1—Yes
   8—Unknown
   9—Refused

cgm (1 digit numeric; must be 0, 1, 8 or 9)
<table>
<thead>
<tr>
<th>Question</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic/Metabolic Disorder?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0—No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 --Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9—Refused</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart Disease?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0—No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 --Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9—Refused</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immune Deficiency?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0—No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 --Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9—Refused</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney Disease?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0—No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 --Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9—Refused</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver Disease?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0—No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 --Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9—Refused</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sickle Cell Disease?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0—No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 --Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9—Refused</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerebral Palsy?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0—No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 --Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9—Refused</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Seizure Disorder (not febrile or head injury)?
0-- No
1—Yes
8 --Unknown
9 –Refused

Mental Retardation or Developmental Delay?
0-- No
1—Yes
8 --Unknown
9 –Refused

Other Neurological Disorder
0-- No
1—Yes
8 --Unknown
9 –Refused

If other neurological disorder, please specify______

11. Is your child Hispanic or Latino?
0-- No
1--Yes
8 --Unknown
9 --Refused

12. Is your child: (Check all that apply)
1 --White
2 --Black or African American
3 --American Indian/Alaska Native
4 --Asian
5 -- Native Hawaiian/Other Pacific Islander
6-- Other
7-- None
8 -- Unknown
9 --Refused/no response
     Racecr1 □
     Racecr2 □
bearly (1 digit numeric; must be 0, 1, 8 or 9)
13. Was your child born more than one month early (more than 4 weeks early or less than 36 weeks gestation)?
   0--No
   1--Yes (if yes, go 13a)
   8--Unknown
   9--Refused

bwks (numeric up to 2 digits each; 88 = unknown, 99 = refused/ no response)
13a. If yes, how many weeks early was your child born?
   □□ Weeks

Bwtlbs, bwtozs (2 numeric digits each, 88 = unknown, 99 = No response/refused)
14. What was your child’s birth weight?
   bwtlbs □□lbs.
   bwtoz  □□ozs.

OR
Bwgms (numeric up to 4 digits, 8888 = unknown, 9999 = No response/refused)
   bwgms □□□□gms.

breastf (1 digit numeric; must be 0, 1, 8 or 9)
15a. Did the mother ever breastfeed or pump breast milk to feed this child?
   0--No (If no, go to 15b)
   1--Yes (If yes, go to 15c, 15d)
   8--Unknown
   9--Refused

15b. If no, how old was the child the first time he or she ate any kind of food?
   firstfoodwks (2 digit numeric; 66=less than 1 week, 77=has not eaten any foods, 88=unknown, 99=refused/no response)
   □□ Weeks
   OR
   firstfoodmonths (2 digit numeric)
   □□ Months

15c. If yes, how many weeks or months did/has the mother breastfed or pumped milk to feed the child?
   Breastflowlongwks (2 digit numeric; 66=less than 1 week, 88=unknown, 99=refused/no response)
   □□ Weeks
Breastfeedinghowlong months (2 digit numeric)

☐☐ Months

Breastfeeding still (1 digit numeric; must be 0, 1, 8 or 9)

15d. If yes, is the mother still currently breastfeeding or feeding pumped milk to this child?

0—No (go to 15e)
1—Yes
8—Unknown
9—Refused

15e. If no, how old was the child the first time he or she drank liquids other than breast milk?

Firstdrink wks (2 digit numeric; 66=less than 1 week, 77=has not had other liquids, 88=unknown, 99=refused/no response)

☐☐ Weeks

OR

Firstdrink months (2 digit numeric)

☐☐ Months

daycare (1 digit numeric; must be 0, 1, 8 or 9)

16. Does your child attend any type of day care, preschool, or school for more than 4 hours/week?

0—No
1—Yes
8—Unknown
9—Refused

dcnum (1 digit numeric; must be 1, 2, 3, 8 or 9)

16a. If yes, with how many unrelated children?

1—Fewer than 6
2—6-12
3—More than 12
8—Unknown
9—Refused

AGE outside (1 digit numeric; must be 0, 1, 8 or 9)

17. Did your child come into contact with anyone outside the household with diarrhea or vomiting in the past week?

0—No
1—Yes
8—Unknown
9—Refused

AGE inside (1 digit numeric; must be 0, 1, 8 or 9)

18. Did your child come into contact with anyone inside the household with diarrhea or vomiting in the past week?
0 -- No  
1 -- Yes  
8 -- Unknown  
9 -- Refused  

`totppl` (2 digit numeric; 88=unknown, 99=refused/no response)  

19a. Including the index child, how many people live in your household? _____  

19B. Please list all household members and whether they have experienced diarrhea or vomiting in the previous 7 days: 

<table>
<thead>
<tr>
<th>Household member’s relation to child</th>
<th>Age (years)</th>
<th>Sex</th>
<th>Did member have diarrhea or vomiting in past 7 days?</th>
<th>If yes, what medical care was sought for this illness?</th>
<th>Date symptoms began</th>
<th>Date symptoms ended</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. HHrelation1</td>
<td>HHage1</td>
<td>HHsex1</td>
<td>HHill1</td>
<td>HHcare1</td>
<td>HHstart1</td>
<td>HHstop1</td>
</tr>
<tr>
<td>2. HHrelation2</td>
<td>HHage2</td>
<td>HHsex2</td>
<td>HHill2</td>
<td>HHcare2</td>
<td>HHstart2</td>
<td>HHstop2</td>
</tr>
<tr>
<td>3. HHrelation3</td>
<td>HHage3</td>
<td>HHsex3</td>
<td>HHill3</td>
<td>HHcare3</td>
<td>HHstart3</td>
<td>HHstop3</td>
</tr>
<tr>
<td>4. HHrelation4</td>
<td>HHage4</td>
<td>HHsex4</td>
<td>HHill4</td>
<td>HHcare4</td>
<td>HHstart4</td>
<td>HHstop4</td>
</tr>
<tr>
<td>5. HHrelation5</td>
<td>HHage5</td>
<td>HHsex5</td>
<td>HHill5</td>
<td>HHcare5</td>
<td>HHstart5</td>
<td>HHstop5</td>
</tr>
<tr>
<td>6. HHrelation6</td>
<td>HHage6</td>
<td>HHsex6</td>
<td>HHill6</td>
<td>HHcare6</td>
<td>HHstart6</td>
<td>HHstop6</td>
</tr>
<tr>
<td>7. HHrelation7</td>
<td>HHage7</td>
<td>HHsex7</td>
<td>HHill7</td>
<td>HHcare7</td>
<td>HHstart7</td>
<td>HHstop7</td>
</tr>
<tr>
<td>8. HHrelation8</td>
<td>HHage8</td>
<td>HHsex8</td>
<td>HHill8</td>
<td>HHcare8</td>
<td>HHstart8</td>
<td>HHstop8</td>
</tr>
<tr>
<td>9. HHrelation9</td>
<td>HHage9</td>
<td>HHsex9</td>
<td>HHill9</td>
<td>HHcare9</td>
<td>HHstart9</td>
<td>HHstop9</td>
</tr>
<tr>
<td>10. HHrelation10</td>
<td>HHage10</td>
<td>HHsex10</td>
<td>HHill10</td>
<td>HHcare10</td>
<td>HHstart10</td>
<td>HHstop10</td>
</tr>
<tr>
<td>11. HHrelation11</td>
<td>HHage11</td>
<td>HHsex11</td>
<td>HHill11</td>
<td>HHcare11</td>
<td>HHstart11</td>
<td>HHstop11</td>
</tr>
<tr>
<td>12. HHrelation12</td>
<td>HHage12</td>
<td>HHsex12</td>
<td>HHill12</td>
<td>HHcare12</td>
<td>HHstart12</td>
<td>HHstop12</td>
</tr>
<tr>
<td>HHrelation13</td>
<td>HHage13</td>
<td>HHsex13</td>
<td>HHill13</td>
<td>HHcare13</td>
<td>HHstart13</td>
<td>HHstop13</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>13.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HHrelation1-15: (Character ≤50 spaces)
Household member’s relation to child

HHage1-15: (3 digit numeric; 777= less than 1 year, 888=unknown, 999=refused/no response)
Age (years)

HHsex1-15: (1 digit numeric; value must be 1, 2, 8, or 9)
Sex
1—Male
2—Female
8—Unknown
9—Refused

HHill1-15: (1 digit numeric; value must be 0, 1, 8, or 9)
Did member have diarrhea or vomiting in past 7 days?
0—No
1—Yes
8—Unknown
9—Refused

HHcare1-15: (1 digit numeric; value must be 0, 1, 8, or 9)
If yes, what medical care was sought for this illness?
0—None
1—Hospitalized
2—ED visit
3—Outpatient visit
4—Telephone call with medical staff
8—Unknown
9—Refused

HHstart1-15: (date field, mm/dd/yyyy, range from 10/01/2011 to 10/31/2012 check: admitdate should be ≤ to scrdate. alert if not; 88/88/8888= unknown; 99/99/9999=refused)
Date symptoms began
HHStopt1-15: (date field, mm/dd/yyyy, range from 10/24/2011 to 10/31/2012 check: admitdate should be ≤ to scrdate. alert if not; 77/77/7777=ongoing, 88/88/8888= unknown; 99/99/9999=refused)
Date symptoms ended

momage_7 (2 digit numeric; 88=unkown, 99=refused/no response)
20. How old is the child’s biological mother? □□ Years

mdegree_7 (numeric, 1 digit, legitimate values 0-5, 8,9)
21. What is the highest degree or diploma the biological mother has completed?
   0- None
   1- GED
   2- HS Diploma
   5- 2-Year college degree (associate’s or technical degree)
   3- 4-Year college degree (bachelor’s degree)
   4 -Graduate degree (master’s, doctorate, medical, etc.)
   8 -Unsure/Unknown
   9 -Refused/No Response

Sickloc (numeric, 1 digit, legitimate values 1-9)
22. When your child is sick, where do you usually take your child?
   1--Doctor’s Office
   2--Public Health Clinic or Community Health Center
   3--Hospital based Practice
   4--Hospital Emergency Room
   5--Urgent Care Center
   6--Some Other Kind of Place
   7--No Usual Place
   8--Unknown
   9—Refused

rotavac (1 digit numeric; must be 0, 1, 8 or 9)
23a. Has your child received a rotavirus vaccine?
   0--No
   1--Yes (If yes, go to 31b)
   8--Unknown
   9--Refused

rotavacnum (1 digit numeric; must be 0 to <=5, 8=Not Sure/Unsure 9=Refused)
23b. Number of doses ever received:
   □ Doses

tcons (1 numeric digit; must be 0 or 1)
24. Consent for pathogen testing of stool specimen?
   0 --No
   1 --Yes
25. Consent for pathogen testing of saliva specimen?
0 -- No
1 – Yes
9 – Not applicable (child ≥ 5 years old)

26. Was a stool specimen collected during visit?
0— No
1— Yes

abinisp_9 (up to 3 characters)
Specimen collection initials □□□

abinisp2_9 (up to 3 characters)
2nd Specimen collection initials □□□

27. Was a saliva specimen collected during visit?
0— No
1— Yes
9 – Not applicable (child ≥ 5 years old)

salivaini (up to 3 characters)
Specimen collection initials □□□

salivaini2 (up to 3 characters)
2nd Specimen collection initials □□□
RESEARCH LAB RESULTS

deinisp_9
Data entry initials □□□

specimencol (1 numeric digit; must be 0, 1)
1. Was a stool specimen collected?
0 -- No
1 -- Yes

coldat (date field, mm/dd/yyyy; legitimate values: 11/01/2011 to 11/10/2012: coldat should be > or = scrdate. alert if not.
1a. Date specimen collected: □□/□□/□□□□

specimentest (1 numeric digit)
1b. Specimen tested?
0 -- No (If no, go to 1c)
1 -- Yes (If yes go 1e, 1f)

treason (1 numeric digit)
1c. Reason test not done
1 -- Inadequate specimen, unable to test
2 -- Specimen lost in transit
3 -- Other (If other go to 1d)

othtreason (character field, length ≤ 50 characters)
1d. If other specify ________________

tdat (Date field; mm/dd/yyyy, legitimate values 11/01/2011 to 11/30/2012 check: tdat should be > or = to coldat. alert if not)
1e. Date specimen tested □□/□□/□□□□

tresult (1 numeric digit; must be 0, 1, or 2)
1f. Rotavirus results
0 -- Negative
1 -- Positive
2 -- Inconclusive

visualpos (1 numeric digit; must be 0, 1, or 2; activate when tresult=0 or 2)
1g. If EIA result obtained by using a plate reader was negative or inconclusive, was the specimen visually positive?
0 -- No
1 -- Yes
2 -- Inconclusive

stoolblood (1 numeric digit; must be 0, 1, or 8)
1h. Did stool contain visible blood?
0 -- No
1 – Yes
8 – Unknown

salivacol (1 numeric digit; must be 0, 1, 9)
2. Was a saliva specimen collected?
0--No
1--Yes
9 –Not applicable (child ≥5 years old)

salcoldat (date field, mm/dd/yyyy; legitimate values: 11/01/2011 to 11/10/2012: coldat should be > or = scrdate. alert if not.
2a. Date specimen collected: □□/□□/□□□□

salivatest (1 numeric digit)
2b. Specimen tested?
0 –No (If no, go to 23c)
1—Yes (If yes go 23e, 23f)

saltreason (1 numeric digit)
2c. Reason test not done
1-- Inadequate specimen, unable to test
2-- Specimen lost in transit
3--Other (If other go to 23d)

saltothreason (character field, length ≤ 50 characters)
2d. If other specify ________________

saltdat (Date field; mm/dd/yyyy, legitimate values 11/01/2011 to 11/30/2012 check: tdat should be > or = to coldat. alert if not)
2e. Date specimen tested □□/□□/□□□□

saltresult (1 numeric digit; must be 0, 1, or 2 )
2f. Secretor status result
0 – FUT2 absent
1 – FUT2 present
2 –Inconclusive
NEW VACCINE SURVEILLANCE NETWORK - GASROENTERITIS SURVEILLANCE

ROTAVIRUS VACCINATION VERIFICATION

Abinivv_9

Abstractor initials □□□

abinivv2_9

2nd abstractor initials □□□

deinivv_9

Data entry initials □□□

Provider Information (required)

rvacver_7 (range=0,1,2,7,8,)
1. Number of rotavirus vaccinations recorded.
0 □ 1 □ 2 □ 3 □ 7 □ N/A (child < 6 weeks) 8 □ Unknown/Unable to review chart

2. Rotavirus Dose 1
rpver1_7 (range=1-4) rvac1dat_7 (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)
1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart

rvac1type_9
Vaccine type (range=1, 2, 8)
1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

3. Rotavirus Dose 2
rpver2_7(range=1-4) rvac2dat_7 (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)
1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart

rvac2type_9
Vaccine type (range=1, 2, 8)
1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

4. Rotavirus Dose 3
rpver3_7 (range=1-4) vac3dat_7(date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)
1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart

rvac3type_9
Vaccine type (range=1, 2, 8)
1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

Registry Information (optional)

rvacver_7r (range=0,1,2, 3, 7,8,)
1. Number of rotavirus vaccinations recorded.
0 □ 1 □ 2 □ 3 □ 7 □ N/A (child < 6 weeks) 8 □ Unknown/Unable to review chart

2. Rotavirus Dose 1
rpver1_7r (range=1-4) rvac1dat_7r (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)
1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart

rvac1type_9r
Vaccine type (range=1, 2, 8)
1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

3. Rotavirus Dose 2
rpver2_7r (range=1-4) rvac2dat_7r (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)
1 □ Vaccination verified Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart

rvac2type_9r
Vaccine type (range=1, 2, 8)
1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

4. Rotavirus Dose 3
 rpver3_7r (range=1-4)  vac3dat_7r (date field, mm/dd/yyyy legitimate values 02/01/2006-10/31/2012)
1 □ Vaccination verified  Vaccination Date □□/□□/□□□□ 88/88/8888 = Unknown
2 □ None in record
3 □ Patient not in practice
4 □ Unable to review chart

rvac3type_9r
Vaccine type (range=1, 2, 8)
1 □ RotaTeq (Merck)
2 □ Rotarix (GSK)
8 □ Unknown

timestamp
Date and time of last edit
March 5, 2013

Robert Frenck, M.D.
Institutional Review Board
Cincinnati Children’s Hospital Medical Center
MLC #5020
3333 Burnet Avenue
Cincinnati, Ohio 45229-3039

RE: Mary Allen Staat, “Enhanced active surveillance of pediatric infectious diseases and vaccines.” ePAS Study # 2011-2246

Dear Dr. Frenck:

This is to advise you that the University of Cincinnati (UC) IRB has accepted the Cincinnati Children’s Hospital Medical Center (CCHMC) IRB review of the above referenced protocol and acknowledges that the CCHMC IRB will be the IRB of record for this protocol.

It is understood that the CCHMC IRB will be responsible for review of revisions of the study protocol, for continuing reviews and for reviewing any problems with the research; CCHMC IRB will make accessible to the UC IRB all correspondence related to this protocol and conduct other activities according the IRB Authorization Agreement (Memorandum of Understanding) with the University of Cincinnati.

It is understood that Mary Allen Staat, Principal Investigator, will be responsible for providing any documents requested by the CCHMC IRB. They will also be responsible for informing the UC IRB of any change in the protocol status at CCHMC.

Sincerely,

Michael Linke, Ph.D.
UC IRB Chair

cc: Mary Allen Staat
ePAS Study # 2011-2246
Appendix D: Protection of Human Subjects

Please note that this appendix was constructed from various IRB documents and is not the original work of Rebecca Currier.

Human Subjects Involvement and Characteristics, and Design

Active disease surveillance and vaccine evaluation activities by the New Vaccine Surveillance Network (NVSN) are conducted by pediatric medical centers as part of the Centers for Disease Control and Prevention (CDC) ranked cooperative agreement. Before enrollment began, human subjects protocols were approved by the Institutional Review Boards (IRBs) at each surveillance site and at the CDC. (Cincinnati Children’s IRB protocol # 2011-2246) The proposed work was approved for active surveillance beginning December 1, 2011 and lasting through November 30, 2012. Children were eligible if they were residents of the active surveillance sites’ defined catchment area, are greater than 14 days and less than 5 years of age, and are admitted to participating hospitals and EDs for treatment of AGE of less than or equal to 10 days’ duration. Healthy controls were enrolled from patients attending well-child visits at each active surveillance site year-round (approximately 1000 subjects). Data including demographic information, illness characteristics and socio-economic status were collected from each subject. A saliva sample was collected upon enrollment, and a stool specimen was collected from all subjects at the earliest opportunity following enrollment, no later than 10 days after the onset of acute gastroenteritis (AGE) symptoms. DNA from saliva samples will be extracted, and saliva specimens tested for genetic markers which may predispose to norovirus infection. Stool from each subject was shipped to the Centers for Disease Control and Prevention (CDC) laboratories to conduct pathogen testing.
Necessity to involve children in research

The leading cause of severe AGE among US children under age five is norovirus. There is currently no vaccine against norovirus on the US market, but vaccines and prophylaxes are currently under development. Active surveillance for norovirus is necessary to: 1) better understand and document the disease burden attributable to norovirus among US infants and children; 2) to monitor the distribution of norovirus strains; and, 3) to determine from diverse populations the role of the FUT2 (secretor) genotype as a predictor of norovirus and AGE infections. It is necessary to involve children in this research, as children under five are the group most likely to suffer from medically attended norovirus AGE.

Consent and recruitment

Children were not fully enrolled in the study until written consent in English (or Spanish at some surveillance sites where an approved Spanish–translated consent and interview was made available) was obtained from the child's parent or guardian and assent obtained from the child as required.

Recruitment proceeded through study staff approaching parents of eligible children to invite them to participate in the study. If the parent agreed to participate, parental consent was obtained and a standardized questionnaire (case report form) was administered by study staff. Study staff requested parental permission to review the child’s medical record from the primary care provider. Because this study involved only one study visit, retention efforts focused on appropriate consent from parents for follow-up questioning by telephone. Additionally, a 24 hour
courier service was provided for participants to submit the bulk stool specimen to lessen the burden on subjects to complete the study protocol.

**Involved locations**

**Cincinnati Children’s Hospital Medical Center**

Cincinnati Children’s Hospital Medical Center (CCHMC) is an independent, not-for-profit hospital and research institute affiliated with the University of Cincinnati College of Medicine. CCHMC is a 525-bed hospital located in Hamilton County, Ohio, and is the only children’s hospital within a 50-mile radius. In 2009, the population of Hamilton County was 855,062; approximately 13% were under 18 years of age and 72% were white.

Overall, CCHMC serves >97% of Hamilton County children. In the surrounding 7 counties, only 5 hospitals provide limited care for children. CCHMC is the major pediatric care provider in a service area of approximately 1,847,986 persons in eight adjoining counties, including Hamilton, Butler, Warren, and Clermont Counties in Ohio; Campbell, Kenton, and Boone Counties in Kentucky; and Dearborn County in Indiana, capturing >85% of pediatric hospitalizations from this extended area. The Pediatric Primary Care Center (PPC) is a large outpatient clinic within CCHMC that cares for approximately 17% of the Hamilton County birth cohort.

Cincinnati Children’s Hospital Medical Center has an IRB Federal-Wide Assurance of Compliance Number FWA 00002988 which expires on December 23, 2013. All research at CCHMC is conducted in accordance with the Common Rule, CFR-45.

**Children’s Mercy Hospital**
Children’s Mercy Hospital (CMH), located in Kansas City, Missouri, is a 314-bed comprehensive pediatric medical center and the only free-standing children’s hospital between St. Louis and Denver. CMH provides state-of-the-art care for children from the Kansas City Metropolitan Area, a fifteen-county metropolitan area that is bisected by the border between the states of Missouri and Kansas. As of the 2010 Census, the metropolitan area has a population of 2,035,334. The population of Kansas City, Missouri, is approximately 63% white, 29% black, and 9% Hispanic. In 2010, CMH saw approximately 12,000 inpatient admissions and 66,000 ED visits. The Pediatric Care Clinic (PCC) at CMH is an outpatient clinic serving approximately 46,000 children annually at three locations. Approximately 50% of PCC visits are for well-child care.

Seattle Children’s Hospital

Seattle Children’s Hospital (SCH) is a 305-bed regional pediatric health care center and the only children’s hospital in King County, Washington, with >85% of the non-neonatal pediatric beds in the county. King County is the most populous county in Washington with 1,916,441 residents, and includes Seattle and the Seattle metropolitan area. Fifteen percent of the county population is under 18 years of age. Seventy-five percent are white, 15% are Asian-Pacific Islander, 7% are black, and 8% are Hispanic. Annually, SCH receives referrals from more than 300 hospitals and clinics and saw 14,106 hospital admissions and 38,414 ED visits in 2009. A hospital-associated private practice research network supports office-based research for therapeutic, epidemiological, and vaccine studies.

Texas Children’s Hospital
Texas Children’s Hospital (TCH) is a 582-bed full-care pediatric hospital in Houston, Texas and is the largest children’s hospital in the United States. The catchment area for TCH is Harris County, which includes Houston, and the surrounding counties, Brazoria, Chambers, Fort Bend, Galveston, Liberty, Montgomery, and Waller Counties, with a total population near six million. Harris County has a population of 4.1 million people, approximately 29% of whom are under the age of 18 years. Fifty-seven percent of the population is white, 19% are black, and 41% are Hispanic. Annually, TCH receives over 20,000 admissions and 82,000 ED visits. An electronic review of discharge diagnoses demonstrated that more than 5,000 children were evaluated and diagnosed with AGE in the ED annually during 2005 and 2006.

University of Rochester, New York School of Medicine and Dentistry

Monroe County, New York, is a socially, economically, and racially diverse county that includes the city of Rochester. Monroe County has a total population of 735,343, 27% of whom are under 18 years of age. Approximately 60% of the population is white, 26% are black, and 12% are Hispanic.

Strong Memorial Hospital (SMH) and Rochester General Hospital (RGH) serve >99% of all hospitalizations for Monroe County residents <11 years old. SMH is an 800-bed general teaching hospital that is part of the University of Rochester Medical Center complex. There were 5,445 pediatric admissions and an additional 479 admissions to the Pediatric Intensive Care Unit in 2010. The SMH ED provides pediatric ED care for 54% of all pediatric ED visits among Monroe County children 0-5 years of age and sees >25,000 pediatric visits annually.

RGH is a 547-bed private teaching hospital with a University of Rochester residency program in Pediatrics. There were 742 pediatric admissions in 2010. Six primary care practices,
all with a history of collaboration with the University of Rochester, were used to recruit healthy controls. Together, these six practices serve more than 22,000 patients under age 11.

Vanderbilt University Medical Center

The Monroe Carell Jr. Children’s Hospital at Vanderbilt (VCH) is a nonprofit facility that serves more than 95% of children residing in the Nashville-Davidson County metropolitan area in Tennessee. Davidson County had a total population of 635,710 in 2009. Twenty-two percent of the population is under the age of 18 years. Two-thirds of the population is white, 27% are black, and 9% are Hispanic.

VCH has 238 inpatient beds and a 30-room emergency department. In FY2010, VCH saw 13,811 pediatric admissions and >55,000 pediatric ED visits. The Vanderbilt Pediatric Primary Care Practice, which will be used to enroll healthy controls, provides comprehensive well-child care for approximately 12,000 infants and children in the Nashville-Davidson county metropolitan area, with approximately 34,000 annual visits.

Data protection, entry, and management

Study data collected at sites is submitted to CDC via a specialized web application using FISMA, (Federal Information Security Management Act), compliant 128-bit encryption for all communication. The web application is hosted on Mid-Tier Data Center (MTDC) web servers with a Secure Access Management Services (SAMS) front-end. Together, they provide high security, encryption, authentication, and user management features. User access to the application is limited by the use of unique system user ID and complex password combinations. To obtain a password, users must provide multiple forms of identity, which are verified by
SAMS personnel. All required study data is entered into the application at each site. Once transmitted to CDC, data resides in a password-protected structured query language (SQL) database maintained on the CDC’s secure internal network. The SQL server is authenticated with a password of a minimum of 20 characters with very high complexity. Every change to the data is logged automatically by triggers on the SQL server, independently of who changed the data or what software is used. Data backups are stored on tape and archived in a confidential location. A business continuity plan is on file to mitigate any physical security risks such as theft or fire at MTDC or the SQL server farm. On an annual basis, each MTDC, SAMS, and SQL server is recertified under a security audit called C&A, (Certification and Accreditation) as is required under the FISMA act. NVSN, as an IT project in its entirety, is also recertified for C&A on an annual basis independently from the servers that it relies upon. Through C&A and FISMA, any emerging security threats and vulnerabilities are periodically scanned for, investigated, and mitigated through software, hardware, or policy and procedure changes.

All patient identifiers, including names, medical record numbers, and contact information, are kept at study sites; no identifying information is included on any data sent neither to the CDC nor to sites other than the site from which the data were collected. Original patient records located at the study sites are linked to the study data contained in the SQL database solely through a subject identification number. All federal data handling requirements are practiced.

All records at study sites and at CDC are kept in locked file cabinets in secure facilities, and study personnel agree to the non-release of confidential information or contact data. Limited data sets and results are shared among all study collaborators. Quality assurance protocols are
required for each study site, and site visits for the purpose of quality assessment of participating surveillance sites review these data handling procedures.

All CDC and site staff with access to patient-level are required to sign the “NVSN Data User Agreement for Access and Use of NVSN Patient Level Data.” By signing this agreement, users agree to:

- not transport or utilize data that reside on flash/thumb drives, CDs, floppies, laptops, or other portable devices. The only exception is receipt of official datasets from CDC or other sites by the site’s authorized staff person per protocol and security guidelines.

- not carry patient data on hardcopy listings or forms. The only exception is for authorized staff who enroll or collect data off-site and must transport it to and from their study site.

- not send unencrypted, personally identifiable data via email attachments.

- only access data on secured computers or through secure connections. A computer or connection is considered secure if access is through the study site’s secure data network, which typically is a VPN (virtual private network) at the institution.

- review the data documentation (including annotated data forms) provided by CDC for the analysis datasets prior to using the data to ensure that users are correctly using the variables in the datasets and to contact the PIs or CDC staff with any questions about the correct use of data.

- promptly inform the NVSN Principal Investigator of any deviation from these guidelines.

- upon leaving NVSN, ensure that all study-related materials remain at the NVSN site and inform the PI of the status and location of all materials.
In addition, all SAMS users must agree to the SAMS Rules of Behavior the first time they log in to the system. Users must agree to:

- keep their account private and not share their password with anyone.
- securely store and protect any written copy of their user name and/or password.
- make every effort to prevent others from watching password entry.
- choose passwords that are difficult to guess by avoiding the use of well known personal information.
- log off of the system when finished or whenever leaving their computer unattended.
- not access SAMS or Program applications using an account that belongs to another person.
- not attempt to circumvent any SAMS’ security control mechanism.
- take positive steps to protect the sensitive and/or non-public information in SAMS, the people this information may represent, and the systems designed to protect it.
- report improper or suspicious activities involving SAMS’ information and systems to the SAMS Help Desk.

CDC provides surveillance sites with several methods for feedback including data integrity checks, summary tables, conference calls, and laboratory results. **Team conference calls are held with the sites and the CDC on a biweekly schedule.** Surveillance staff are trained to notify their site PI of any deviations from the protocol. The site PI in turn is required to notify the CDC project officer no later than 24 hrs after discovery. The CDC project officer then notifies their CDC Team Lead, Branch Chief, and Division’s Associate Director for Science of
the protocol deviation. Site PIs are actively involved in reviewing medical charts and in regular contact with site staff to identify any such events.

**Potential Risks**

This study posed no greater than minimal risk to the subjects and the findings had no clinical implications on the subjects. As defined in 45 US Code of Federal Regulations (CFR) 46.102 (i), “Minimal risk means that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.”

Older subjects were asked to provide a saliva specimen by spitting into a supplied container. There are no known risks associated with collecting a saliva sample in this manner. Subjects too young to provide a saliva sample via spitting had their saliva specimen collected using a sponge provided by an Oragene collection kit. The sponge may cause temporary dryness of the mouth that should not last longer than 5-10 minutes. The analysis of DNA from saliva does not include any testing whose results might impact clinical decision making for the individual patient. Only authorized personnel have access to DNA specimens and DNA data results, which are not stored with any patient identifiers.

Subjects were also asked to provide a stool specimen, either by collection of a soiled diaper or by defection into a “stool hat” or other collection apparatus. The only anticipated risk of providing a stool specimen in this manner was contamination of skin with feces from the collection container. This risk is not considered to be significantly greater than during normal toileting activities.
Because norovirus primarily infects children under the age of five years, we have limited this study to children greater than 14 days and less than 5 years of age. Although this age group is considered a vulnerable population as defined by 45 CFR 46, Section D, this study does not involve greater than minimal risk to the subjects. Furthermore, written consent was obtained from all parents or guardians, and assent was obtained from older children enrolled at applicable sites as required by each site IRB.

Participation in this study was entirely voluntary. Subjects could refuse to participate at any point after consenting to be in the study. Refusal to participate in no way affected the subjects' or families' medical care or standing at the health care facility. Specimens from withdrawn subjects were destroyed.

*Protection from risks*

Study staff approached parents of eligible children to invite them to participate in the study. If the parent/guardian agreed to participate after thorough review of the informed consent, written parental consent was obtained. **For all subjects, in-person, written parental permission was maintained and a copy of the signed consent forms was placed in the child’s medical record.** If the child’s parent/guardian was not present to provide written informed consent, the child was not eligible for the study. According to the applicable IRB requirements of each study site, assent was obtained as age-appropriate The informed consent had separate sections where the parent specifically consented to testing the stool for pathogens, testing the saliva for genetic information related to pathogen susceptibility, and recontact to collect additional information. Some sites included an optional future use clause that was part of a separate consent process and not required to participate in the main study.
Benefits to subjects

Sites provided a small reimbursement for subjects ($10-$25 depending upon the site). There were no other direct benefits to the subject. All data entry, chart abstraction, and specimen testing occurred after the patient left the hospital or clinic and therefore could not have affected the patient’s course of treatment, either positively or negatively. This study poses no greater than minimal risk, which is appropriate in proportion to the benefits.

Importance of knowledge to be gained

Norovirus AGE is a problem of significant public health importance due to the high prevalence of NoV both in outbreaks as well as in sporadic cases. NoV is estimated to cause at least 5 million episodes of AGE annually in the United States alone. Children under five have a high incidence of norovirus AGE.

Norovirus does not infect all patients the same way. Observations from both volunteer and outbreak studies show that susceptibility to development norovirus disease varies greatly among individuals: some persons seem resistant, while others suffer prolonged illness. Human ABH histo-blood group antigens may influence susceptibility to NoV with a potential association with FUT2 status. Further study of these associations on nationally circulating strains is needed since different norovirus genotypes may use different receptors. A better understanding of these associations will help to identify those at risk of infection and therefore enable implementation of more targeted control and preventive measures.
This knowledge will impact design of future trials, therapeutics, and public health interventions. The risks to human subjects in obtaining this knowledge is minimal and reasonable.