Distribution Agreement

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

Signature:

_____	April 20, 2015
	_____
	Date

Epidemiologic Profiles: Clinical and Epidemiologic Profiles for Norovirus Outbreaks

By

Joana Yu
Master of Public Health

Epidemiology

_____ [Chair's signature]
Juan Leon, Ph.D., MPH
Committee Chair

_____ [Member's signature]
Aron J. Hall, DVM, MSPH
Committee Member

Abstract

Epidemiologic Profiles: Clinical and Epidemiologic Profiles for Norovirus Outbreaks

By Joana Yu

Background: Noroviruses are the leading cause of acute gastroenteritis outbreaks and foodborne disease outbreaks in the United States. Laboratory testing for norovirus during outbreak investigations has historically been limited by the availability of molecular-based diagnostics and resource constraints at state and local health departments. In the absence of laboratory confirmation, clinical and epidemiologic profiles, such as the Kaplan criteria (vomiting in ≥50% of cases, mean incubation period of 24–48 h, mean duration of illness 12–60 h, and negative bacterial culture) and the ratios of fever-to-vomiting and diarrhea-to-vomiting have proven useful in distinguishing norovirus from bacterial agents.

Methods: Previously proposed clinical and epidemiologic profiles were reevaluated with outbreaks occurring during 2009-2012 and reported through the National Outbreak Reporting System (NORS), specifically those with the following etiologies: laboratory confirmed norovirus (N=2,939), suspected norovirus (N=1,321), laboratory confirmed non-viral (N=1,544), and unknown etiology (N=3,694). Alternative clinical and epidemiologic profiles were developed and evaluated with classification and regression tree (CART) modeling. The performance of previous profiles and the CART predictors was evaluated by Cohen's kappa statistic, as well as the proportion of outbreaks with all criteria reported, sensitivity, specificity and the likelihood ratio.

Results: The Kaplan criteria remained highly specific (100%, 95% CI: 83.2%-100%) with a Cohen's kappa of 0.34 but only 108 (3.7%) confirmed norovirus and 19 (1.2%) confirmed non-viral outbreaks had all information for the criteria reported. With CART modeling, an alternative clinical and epidemiologic profile was developed with the fever-to-vomit ratio <1, the proportion of cases with vomiting ≤0.34, and the proportion of cases with bloody stool <0.12. The CART predictors had a high likelihood ratio of 12.5, a Cohen's kappa of 0.78 (95% CI: 0.75-0.81), and 706 (24.0%) confirmed norovirus and 605 (39.1%) confirmed non-viral outbreaks had information for all criteria reported.

Conclusion: Relative to the Kaplan criteria, the CART predictors were similarly effective in distinguishing norovirus from non-viral outbreaks, but were reported far more frequently in NORS. These predictors provide a useful alternative profile for identifying likely norovirus etiology during outbreak investigations.

Epidemiologic Profiles: Clinical and Epidemiologic Profiles for Norovirus Outbreaks

By

Joana Yu

Bachelor of Science, Biology
Georgia Institute of Technology
2011

Thesis Committee Chair: Juan Leon, Ph.D., MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2015

Acknowledgments

First and foremost, I'd like to thank my field advisor, Aron Hall, for giving me the opportunity to tackle his idea, for allowing me to make mistakes, and for patiently working with me.

Additionally, I'd like to thank my thesis advisor, Juan Leon, for his patience and guidance throughout this process, for always asking the right questions, and helping me articulate ideas.

I'd also like to thank Shacara Johnson for all of her help explaining complicated mathematical modeling into digestible tidbits, sharing her ideas, and always providing helpful advice.

I'd like to thank Weidong Gu and Dana Cole for the opportunity to collaborate and always giving helpful advice.

I'd like to thank Mary Wikswo and Eyal Lesham for their help and support.

Finally, I'd like to thank to thank my friends and family for their enduring support and encouragement.

To Molly, thanks for keeping me company on candy runs, mental health breaks, and unfortunately annoying everyone in their offices with our brainstorming.

To Neetu and Oumar, thanks for always being there to unwind or mutually stress.

To Ryan, thanks for putting up with my crazy hours and data analysis talk all the time.

Table of Contents

Chapter I. Background

Global Impact of Norovirus

Worldwide, diarrheal diseases exact a tremendous burden, accounting for 1.45 million deaths every year and an estimated 98.5 million disability-adjusted life-years (DALYs) [1, 2]. Diarrheal illnesses can cause a wide range of symptoms that range from minor discomfort to dehydration which can result in death (reviewed in [3]). Episodes of diarrhea can also negatively affect overall health through malnutrition or weakened immunity, especially for vulnerable populations including children and the elderly [4-6]. Norovirus is estimated to be associated with 18% of all cases of acute gastroenteritis (AGE) characterized by diarrhea and vomiting [7]. In low income countries, AGE has caused more than 25% of deaths in children younger than 5 years [7]. Noroviruses are the leading cause of outbreak-associated gastroenteritis worldwide (reviewed in [8]). In both developing and developed countries, the burden of norovirus infections is estimated to be hundreds of cases per 10,000 persons; however, the number of infections is still underestimated due to the lack of diagnosis and reporting to surveillance systems (reviewed in [9]).

Norovirus in the United States

In the United States, norovirus is the leading cause of sporadic gastroenteritis in all age groups and the most common cause of foodborne illnesses (reviewed in [10-12]). Annually in the United States, there are 19–21 million cases, 56,000–71,000 hospitalizations, and 570–800 deaths attributed to norovirus [13]. Moreover, it has been

estimated that in the United States norovirus illnesses have resulted in a loss of 5,000 quality-adjusted life years (QALY) annually [14]. Although norovirus can affect all age groups, young children and the elderly are at increased risk for more severe and prolonged illness leading to hospitalization [15]. Among children <5 years of age, norovirus is the leading cause of AGE, and norovirus infections account for nearly 1 million healthcare visits per year [16]. Elderly adults ≥65 years of age are at the greatest risk for norovirus-associated death with an estimated 90% of annual norovirus-related deaths occurring in the elderly [13]. Moreover, for immunocompromised patients, norovirus is increasingly being recognized as a significant cause of chronic gastroenteritis [17].

Norovirus is the leading cause of AGE outbreaks in the United States, which have occurred in a wide variety of settings including nursing homes, retirement centers, hospitals, cruise ships, schools, restaurants, and catered events (reviewed in [18-21]). The most commonly reported settings of norovirus outbreaks in the United States are long-term care facilities including nursing homes (reviewed in [11, 22]). Health-care staff, visitors, and patients can introduce the virus, and outbreaks in these settings have been demonstrated to last several months [23]. Additionally, restaurants and catered events are commonly reported settings for norovirus outbreaks. With norovirus as the leading cause of foodborne illness (reviewed in [24, 25]), food can be contaminated with norovirus at any point during production, processing, distribution, or preparation [26]. Norovirus outbreaks have also been reported in schools, child care centers, and universities in the United States [27, 28]. Norovirus outbreaks can occur throughout the year, but seasonal patterns have been observed with increased activity during the winter months (reviewed in [13, 22, 29]).

Biology and Epidemiology of Norovirus

Noroviruses are nonenveloped, single-stranded RNA viruses in the family *Caliciviridae*. The prototype norovirus, Norwalk virus, was first identified as the causative agent of an acute nonbacterial gastroenteritis outbreak in Norwalk, Ohio in 1968 [30]. Noroviruses comprise 7 distinct genogroups (GI-GVII) with three genogroups that infect humans (GI, GII, and GIV) [31]. The most prevalent genotype, GII.4, accounts for 70% of the capsid-based genotypes [32]. The emergence of new GII.4 genotype strains are associated with periodic increases in norovirus outbreaks due to evasion of population immunity, and new strains tend to rapidly replace exiting strains in circulation causing unusually high norovirus activity (reviewed in [22, 33]). Noroviruses are highly infectious with infectious doses between 18 to 10^3 viral particles [10]. Transmission of norovirus can be both fecal-oral or vomit-oral and can occur through various routes including person-to-person, environmentally-mediated, foodborne, and waterborne transmission [34]. Noroviruses are also environmentally stable and have been found to persist on environmental surfaces during non-outbreak periods [34]. With a low infectious dose, a variety of transmission routes, and viral stability, norovirus outbreaks can be occur in a variety of settings and can be difficult to control.

Norovirus infections can cause a variety of symptoms. Illness typically begins after a short incubation period of 10–51 hours incubation period [15] with symptoms characterized by acute onset of non-bloody diarrhea, vomiting, nausea, abdominal cramps, fever, or body aches [26]. Without treatment, symptoms usually resolve after 28–60 hours in otherwise healthy individuals [35]. In young children, the elderly, and immunocompromised persons, symptoms have been observed to last for prolonged periods

from 4–6 days (reviewed in [36, 37]). Norovirus-associated deaths among the elderly have also been reported as a result of outbreaks in long-term care facilities [38].

Norovirus can be shed through a variety of mechanisms; however, it is uncertain if detection of the virus alone is a risk for transmission [26]. Norovirus is shed primarily through stool but has been detected in vomitus and mouthwash samples of individuals with AGE [39]. Peak viral shedding occurs 2–5 days after infection with nearly 100 billion viral copies per gram of feces [40]. Norovirus has been detected in fecal samples for a median of 4 weeks and up to 8 weeks after infection [40]. Although there has been documented evidence of prolonged viral shedding, it is unclear for how long these viruses are infectious after illness. Moreover, up to 30% of documented norovirus infections were asymptomatic and shedding virus at slightly lower titers than symptomatic individuals [40-42].

Surveillance Systems for Norovirus in the United States

Surveillance systems are crucial to better understanding the frequency of outbreaks, the spread of existing and emerging pathogens, the major modes of transmission, and the incidence of disease in the United States. Two national systems for norovirus outbreak surveillance are currently in place through the Centers for Disease Control and Prevention (CDC), CaliciNet and the National Outbreak Reporting System (NORS). CaliciNet was launched in 2009 and is utilized by the local, state, and federal health laboratories to monitor norovirus strains associated with outbreaks. Information on genetic sequences and epidemiology data related to norovirus outbreaks are reported to aid in linking norovirus outbreaks to specific strains and monitor for new or emerging strains of norovirus [43]. NORS was also launched in 2009 to aid in reporting all outbreaks of gastrointestinal illness,

including those resulting from foodborne, waterborne, person-to-person, environmental, and animal contact transmission [26]. NORS was implemented to improve and expand upon already existing waterborne and foodborne surveillance systems for enteric illnesses [44].

NoroSTAT, the Norovirus Sentinel Testing and Tracking network, was implemented in August 2012 to improve the efficiency, completeness, and consistency of norovirus outbreak reporting. Through a collaborative network of five state health departments (Oregon, Minnesota, Wisconsin, Ohio, and Tennessee) and the CDC, norovirus strain data from CaliciNet are rapidly linked with epidemiologic characteristics of outbreaks reported through NORS [45]. Data from these systems may help improve the attribution of norovirus disease and can be evaluated to assess the impact of new norovirus strains, the frequency of norovirus outbreaks, and the severity of norovirus outbreaks.

With these surveillance systems, more robust data for norovirus outbreaks including strain information and clinical and epidemiologic criteria have been collected for all modes of transmission; however, there are still limitations with reported surveillance data. For foodborne outbreaks reported from 2009-2012 through NORS, there was a 100-fold difference in reporting rates between the highest and lowest reporting states and some states did not report any outbreaks.[24] This drastic variation in reporting reflects the sensitivity of outbreak reporting among states rather than the incidence of disease alone. Based on the variation in reporting rates among states, it is likely that disease incidence is much higher and signifies the continued need for increased capacity in state and local health departments to investigate and report outbreaks. Moreover, the lack of reported

comprehensive information on outbreaks including the source of contamination or additional contributing factors can limit the understanding and cause of an outbreak [24].

Outbreak Investigations in the United States

In the United States, surveillance and outbreak investigations are crucial to guide measures for reducing norovirus-related illnesses, such as the promotion of hand hygiene, environmental disinfection, and the isolation of infected persons [26, 46]. Due to the highly infectious nature of norovirus, it is essential to minimize transmission and limit contamination of the environment especially in settings with great opportunities for exposures such as restaurants, hospitals, long-term care facilities, universities, or cruise ships. These practices are based on infection-control principles to minimize contact with individuals at their peak infectious periods [47].

Surveillance systems are essential to identifying potential outbreaks. Outbreaks are defined as two or more persons with similar illness resulting from a common exposure or, more broadly, any increase above the baseline of expected disease [48, 49]. Outbreak investigations are generally performed to find the source of the pathogen or to eliminate the source of infection. A small proportion of cases is often initially reported in the initial stages of outbreak investigations. Therefore, case definitions are often created to capture more persons with illness to understand the size, severity and timing of an outbreak [50]. Several case definitions may be created for an investigation, such as confirmed, probable, or suspect cases. Data are often collected from cases including descriptive characteristics such as age, race/ethnicity, occupation, recent travel, or attendance at an event [48, 49]. Clinical information are also typically sought, including specific symptoms, timing of

illness onset and recovery, and whether or not health care was sought. Based on this information, investigators can use data to distinguish the person, place, and time of illness to create an epidemic curve [50]. With an epidemic curve, the outbreak may be classified as a point-source outbreak with exposure at one point in time, a common source outbreak with continued exposure, or a propagated outbreak with initial exposures and secondary exposures [50]. This information can help investigators formulate and test hypotheses on the possible sources of infection with the use of distinct clinical and epidemiologic profiles. In addition to case data, laboratory samples may be collected for clinical testing [48, 49].

Challenges with Diagnosing Norovirus

Progression of diagnostics for norovirus has improved over the years; however, many challenges including cost, time, and necessary equipment are still issues. Diagnostic testing for norovirus with whole stool samples are preferred; although, rectal swabs and vomitus can also be used. Diagnostics for norovirus were initially performed with an electron microscope but were highly intensive and not widely available in diagnostic laboratories (reviewed in [31]). Immunochromatographic lateral flow assays for rapid diagnostics were created for a panel of various norovirus genotypes with 100% specificity; however, sensitivity of the tests were low (35-52%) and required further validation (reviewed in [51]). Additionally, broadly-reactive enzyme immunoassays (EIAs) were created but have been shown to have a sensitivity of <70% and a specificity of 90% [52]. Although the specificity is fairly high, general consensus among the scientific community is that that EIAs are only useful for rapid screening of multiple fecal samples collected during an outbreak. However, because of its low sensitivity, results with EIAs should be

interpreted with caution from sporadic cases [52]. Additionally, developing broadly reactive EIAs for norovirus has been challenging due to the number of antigenically distinct norovirus strains and the requirement for high viral load [26].

Current diagnostic testing is performed with the gold standard RT-qPCR although no commercial stand-alone norovirus RT-qPCR assay has been FDA approved, and equipment for RT-qPCR is not widespread in clinical settings. RT-qPCR assays are highly sensitive and can provide a quantitative amount of virus present in the sample to determine the viral load, have a higher throughput, better adaptability to newer strains of norovirus, and shorter turnaround times than electron microscopy, immunochromatographic lateral flow assays, or EIAs (reviewed in [31]). Although RT-qPCR is a useful assay for detecting the presence of norovirus, viral detection does not always correlate with clinical norovirus disease. Norovirus has been detected in individuals for a range of 4–8 weeks after infection [40]. Moreover, viral shedding has been documented in symptomatic and asymptomatic persons [40-42, 53, 54]. Therefore, positive RT-qPCR results in asymptomatic persons with lower viral loads could be challenging to interpret.

Clinical and Epidemiologic Criteria

In the absence of diagnostic laboratory testing or inconclusive tests, assessing outbreaks by clinical and epidemiological profiles has proven to be effective for foodborne AGE outbreaks (reviewed in [55-57]). Characteristic features of outbreaks with respect to incubation periods, duration of symptoms, and the proportion of cases that experience certain symptoms like vomiting, diarrhea, or fever are often very specific and be indicative of etiology (reviewed in [57]). Kaplan et al. demonstrated that the lack of diagnostic testing

was problematic to attributing AGE outbreaks to norovirus. Thus, they established a set of criteria to distinguish AGE outbreaks attributed to norovirus from those of bacterial etiology for all modes of transmission. Using records of gastroenteritis outbreaks among all modes of transmission in the United States from 1976–1980, Kaplan et al. worked to provide clinical and epidemiologic characteristics of Norwalk gastroenteritis [58]. The criteria characterized by Kaplan et al. have been regarded as the most useful discriminating non-laboratory based diagnostic aid in identifying norovirus outbreaks. The Kaplan criteria included: vomiting in ≥ 50% of cases in an outbreak; a mean or median incubation period of 24–48 hours; a mean or median duration of illness between 12–60 hours; and no bacterial pathogen detected in the stool culture [35].

Turcios et al. reexamined the Kaplan criteria with foodborne outbreaks reported from 1998–2000 through CDC's Foodborne Outbreak Reporting System. They scrutinized the ability of the criteria to discriminate between outbreaks of norovirus or bacterial etiology from other clinical profiles like the fever-to-vomiting ratio and the diarrhea-to-vomiting ratio, proposed by Hedberg et al. and Dalton et al., respectively [59, 60]. Turcios et al. demonstrated that the Kaplan criteria still remain highly specific (99%) and moderately sensitive (68%) for norovirus detection in foodborne outbreaks of AGE [56]. Furthermore, Hedberg et al. demonstrated that the use of clinical and epidemiological profiling was effective in identifying the etiology of confirmed foodborne outbreaks and 54% of confirmed foodborne outbreaks reported to the CDC from 1982–1997 with unknown etiology shared norovirus-like clinical profiles [55].

The use of clinical and epidemiological profiling has proven to be an effective method of identifying outbreak etiologies and can help guide diagnoses for outbreak

investigations and surveillance (reviewed in [55, 57]). Similar with other uses of surveillance data, reporting of clinical and epidemiologic profiles must be systematic and complete for use of clinical and epidemiologic profiles. Additionally, distinguishing characteristics of pathogen specific clinical profiles is useful in guiding outbreak investigations and prompt implementation of pathogen-specific prevention measures. Moreover, identifying likely pathogens by clinical profiling can guide diagnostic testing and can aid laboratory diagnostics by indicating the appropriate tests (reviewed in [55]).

Classification and Regression Tree Modeling

Modeling the clinical and epidemiological criteria of norovirus outbreaks could provide additional insights into clinical symptoms and characteristics among reported outbreaks in NORS. The use of classification and regression tree (CART) modeling in public health has been established as a useful method to identify mutually exclusive and exhaustive groups to identify common characteristics in a population (reviewed in [61-63]). CART modeling analysis has demonstrated the finding of optimal clinical predictors that successfully estimated the probability of outcomes (reviewed in [61, 62]). CART modeling is a nonparametric statistical technique that can be used with both categorical and continuous dependent variables to solve classification and regression problems (reviewed in [64-66]). With CART modeling the dependent variable Y is explained by a set of independent predictors X where [$X=(X_1, X_2, \ldots X_i)$]. CART determines the best predictors that subsets the dependent variable into homogenous subsets using a greedy algorithm and recursively partitions based on the Gini index, misclassification rate, or entropy [64, 65]. Moreover, when predictor values are missing, the CART model

inherently chooses surrogate variables that best represent the missing value. Additionally, tree building with CART models strongly parallels stepwise regression where predictors are included one at a time in successive order. However, unlike stepwise regression, CART modeling does not order the predictors linearly. Instead, predictors that best differentiate outcomes are represented near the top of the tree model. CART models can often be overfitted to the training data used for modeling, but the tree model can be pruned to reduce overfitting [64-66]. With CART modeling, more representative clinical and epidemiological characteristics of reported outbreaks in NORS could be identified to distinguish norovirus outbreaks from other outbreaks of AGE. Furthermore, identification of better clinical and epidemiologic criteria would improve the identification of norovirus attributable AGE outbreaks.

Conclusion

Worldwide, noroviruses are the leading cause of AGE outbreaks. In the United States, norovirus is the leading cause of foodborne outbreaks and AGE outbreaks with peak outbreak activity in the winter months. CaliciNet and NORS were established in the United States to improve surveillance for norovirus by documenting outbreaks and novel strains of norovirus. For norovirus outbreak investigations, laboratory detection of norovirus is available with RT-qPCR; however, this diagnostic method may not be commonly available or rapid enough for effective implementation of control measures. In the absence of laboratory testing, clinical and epidemiologic profiles have been demonstrated to be useful in differentiating etiologic agents in outbreak investigations. Notably, Kaplan et al. characterized clinical and epidemiologic criteria to distinguish norovirus from other agents

that cause acute gastroenteritis. Reevaluation of these criteria by Turcios et al. with foodborne outbreaks in the United States demonstrated that these criteria are highly specific and moderately sensitive. With the advent of new surveillance systems and collection of data on all modes of transmission, it is necessary to reevaluate the Kaplan criteria with more comprehensively reported outbreaks in the United States.

Chapter II. Manuscript

Introduction

Annually, an estimated 179 million cases of acute gastroenteritis (AGE), defined as diarrhea or vomiting, occur in the United States [25]. In the United States, gastrointestinal disease of unknown etiology are attributed to an estimated 70,000 hospitalizations and 1,600 deaths annually [67]. Gastrointestinal diseases can be acquired from a range of viruses, bacteria, parasites, toxins, chemicals, or other noninfectious agents. Although there are no specific treatments for viral gastroenteritis, identification of the causative agent especially during an outbreak investigation is critical for preventative measures to limit the spread of disease. Moreover, it is important to distinguish between viral and bacterial or parasitic etiologies, as there are some specific treatments available for infections with bacterial or parasitic agents [68].

In 2009, the Centers for Disease Control and Prevention (CDC) launched a new national surveillance system, the National Outbreak Reporting System (NORS) that improved and expanded upon the two existent food and waterborne disease surveillance systems (reviewed in [26, 69]). NORS allows for local, state, and territorial health departments to report on all outbreaks of foodborne and waterborne disease regardless of etiology. Moreover, NORS provides a national surveillance system for all pathways of AGE outbreaks in the United States, including those that are spread through direct person-to-person contact, animal contact, contaminated environments, and other or unknown transmission routes. Detailed information on temporal trends, specific pathogens, and

exposure pathways provides a greater understanding of AGE epidemics and can help guide appropriate interventions for current and future outbreaks [24].

Even though NORS is an effective system for reporting, 51% of outbreaks reported through NORS reported outbreaks had no confirmed etiology identified [69]. This can most often be attributed to the lack of clinical specimen collection for diagnostic testing. Additionally, passive reporting through NORS is also subject to variability between states and among outbreaks with different exposures and methods of transmission [69]. Many other factors could also influence the absence of identification of outbreak etiology.

Although diagnostic testing is an effective method of identifying etiologic agents of AGE outbreaks, specimens may not be collected, complete testing may not occur, or lack of diagnostic equipment in clinical or local health department laboratories may hinder detection of an etiologic agent. AGE outbreaks often have pathogen-specific clinical symptoms and epidemiologic profiles. Use of these clinical and epidemiologic profiles can help in identifying AGE outbreaks of unknown etiology. Additionally, the use of these clinical-epidemiologic profiles can facilitate investigations and expedite implementation of control measures. [57]

Norovirus is the leading cause of AGE outbreaks in the United States and causes an average of 19–21 million total cases each year [13, 69]. Transmission of norovirus can occur from person-to-person, through contaminated food or water, or interaction with contaminated surfaces (reviewed in [26, 34]). If specimens are collected during AGE outbreaks, the preferred method of norovirus detection is RT-qPCR, a relatively resource-intensive laboratory test [31]. Rapid commercial enzyme immunoassays (EIAs) have been developed to detect norovirus but have inadequate sensitivity and are not recommended for

diagnosis of individual cases [31]. Consequently, single cases of norovirus are often not reported and undetected due to the lack of a routine clinical diagnostic assay.

In 1982, Kaplan et al. established a set of criteria to distinguish outbreaks caused by norovirus from outbreaks caused by bacterial etiologies [58]. The criteria include vomiting in ≥50% cases in an outbreak, an average incubation period of 24–48 hours, an average duration of illness of 12–60 hours, and the lack of identification of a bacterial etiology from stool culture. These criteria have proven to be an effective profile to differentiate acute gastroenteritis outbreaks of Norovirus-like etiology [56].

Examining Kaplan criteria and other potential clinical and epidemiologic characteristics among lab confirmed and suspected norovirus outbreaks in NORS could help provide a data-driven profile for use by public health practitioners during outbreak investigations. Determining the frequency with which these Kaplan criteria and other characteristics are reported in NORS can also provide insights into the feasibility and utility of such profiles. Furthermore, assessment of other clinical and epidemiologic criteria such as the proportion of cases with bloody stools, diarrhea, or fever could provide a better profile to distinguish norovirus than Kaplan criteria [70].

In the absence of laboratory confirmation, distinguishing an outbreak etiology by clinical and epidemiological criteria would be advantageous. Classification and regression tree (CART) models have demonstrated to be a useful statistical method of making predictions from surveillance data for public health analyses (reviewed in [62, 63, 66]). CART models identify mutually exclusive and exhaustive subgroups by repeated partition of the dataset set based on shared characteristics. Several epidemiologic studies have assessed risk factors for morbidity from specific diseases, developed screening and

diagnostic tools, and assessed predictors for medical procedures using CART modeling (reviewed in [63, 71]).

The goals of this study aims (1) to compare clinical and epidemiologic characteristics of norovirus to non-viral outbreaks reported through NORS, (2) reevaluate the performance of the Kaplan criteria, and (3) utilize CART modeling to identify an alternate clinical and epidemiologic profile to better distinguish norovirus from non-viral etiology outbreaks.

Methods

Data analysis was performed on all outbreaks reported in the National Outbreak Reporting System (NORS) that occurred during 2009–2012 (N=10,023). Reporting information was collected from all 50 U.S. states and the District of Columbia. Outbreaks from all-modes of transmission (foodborne, person-to-person, environmental, animal contact, and indeterminate/unknown) excluding waterborne were analyzed. Finalized outbreaks, (i.e., those no longer under investigation) in NORS with date of first illness of Jan. 1, 2009 through Dec. 31, 2012 were included. Information from the "General Section" of the NORS form with median incubation period, median duration of illness, and signs or symptoms were used in addition to the "Laboratory Section" with information on etiology and laboratory confirmation [NORS form in appendices]. This study did not use human subjects or identifying information and did not require IRB approval.

Classification of Outbreak Etiology

Classification of known etiology in outbreaks reported in NORS was evaluated by variables in the etiology portion of the NORS form. Outbreaks were initially classified by the etiologies reported by genus name in the "Laboratory Section" of the NORS form. If one or more etiologies were reported by genus name, then the outbreak etiology was "known". If no etiology was reported or missing for this variable, then the outbreak etiology was "unknown".

Classification of single etiology was also evaluated by the number of etiologies reported by genus name. Outbreaks that reported only one etiology by genus name were considered "single etiology" outbreaks. Etiologies that accounted for $\geq$ 1% of single

etiology outbreaks reported in NORS included: *Clostridium* spp., *Campylobacter* spp., *Escherichia* spp., *Shigella* spp., *Salmonella* spp., and norovirus. Single etiology outbreaks that accounted for less than 1% of NORS outbreaks were grouped as "other" outbreaks. Outbreaks that reported two or more etiologies by genus name were designated as "multiple" etiology outbreaks.

Classification of laboratory confirmation for NORS reported outbreaks was assessed by the "Confirmed outbreak etiology" portion in the "Laboratory Section" in the NORS form [see appendices]. If outbreaks provided an etiology by genus name and indicated the outbreak reported a confirmed outbreak etiology then the outbreak was "laboratory confirmed". If outbreaks reported an etiology by genus name but did not report a confirmed outbreak etiology, then the outbreak was designated as "suspected" etiology.

For analysis of clinical and epidemiologic characteristics, outbreaks were classified as laboratory confirmed norovirus, suspected norovirus, confirmed non-viral etiology outbreaks, and unknown etiology outbreaks. Laboratory confirmed norovirus outbreaks (N=2,939) were defined as single etiology outbreaks with laboratory confirmation that reported norovirus. Suspected norovirus outbreaks (N=1,321) were defined as single etiology outbreaks that reported norovirus but did not have laboratory confirmation. Non-viral etiology outbreaks (N=1,544) were defined as single etiology outbreaks with laboratory confirmation but excluded any viral etiologies such as: Astrovirus, Hepatitis A virus, Other-Virus, Rotavirus, and Sapovirus. Non-viral etiology outbreaks without a laboratory confirmation were excluded from analysis. Unknown etiology outbreaks were defined as outbreaks that reported no etiology.

Classification of Clinical and Epidemiologic Characteristics

Clinical and epidemiologic characteristics reported in the "Signs and Symptoms" portion in the "General" section of the NORS form [see appendices] were examined among laboratory confirmed norovirus (N=2,939), suspected norovirus (N=1,321), laboratory confirmed non-viral (N=1,573) and unknown etiology (N=3,694) outbreaks. Various characteristics related to AGE were identified including the proportion of cases with bloody stools, the proportion of cases with diarrhea, the proportion of cases with fever, the proportion of cases with vomiting, the proportion of cases with fever divided by the proportion of cases with vomiting (fever-to-vomit ratio), the proportion of cases with diarrhea divided by the proportion of cases with vomiting (diarrhea-to-vomiting ratio), the median incubation period, and the median duration of illness. Proportions for symptoms were calculated by dividing the number of cases with the symptom by the total number of cases for whom the symptom was available. For analysis, the median incubation period and the median duration of illness were converted from days or minutes to hours. Analysis of clinical and epidemiologic characteristics was performed with SAS, version 9.4 (SAS Institute Inc., Cary, North Carolina).

Analysis of Outbreak Reporting Misclassification

NORS reporting practices were evaluated by examining whether the reported responses to "Etiology Known" agreed with the classification of known etiology stated earlier in the *Classification of Outbreak Etiology* section. Outbreaks that reported yes to "Etiology known" and listed one or more etiologies were considered concordant. Similarly, outbreaks that reported no or had a missing response to "Etiology known" and listed no

etiology were considered concordant Outbreaks that reported yes to "Etiology known" but did not report any etiology by genus name were considered discordant and misclassified. Similarly, outbreaks that reported no or missing to "Etiology known" but reported at least one genus name were considered discordant and misclassified. Cohen's Kappa statistic [72] was performed to test agreement between these classifications.

Analyses of Outbreak Characteristics

Distributions of reported clinical and epidemiologic characteristics were examined for confirmed norovirus outbreaks, suspected norovirus, non-viral outbreaks, and unknown etiology outbreaks. Since the distributions for the clinical and epidemiologic characteristics were non-parametric, medians and interquartile (IQR) ranges were reported and Kruskal-Wallis tests were used to assess differences in distributions [73]. Post-hoc analyses with Steel-Dwass all-pairs comparison tests were performed to determine individual significance compared to characteristics of confirmed norovirus outbreaks [74].

For confirmed norovirus and non-viral outbreaks, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated to evaluate diagnostic performance of the Kaplan criteria collectively and each of its component characteristics individually (i.e., the proportion of cases with vomiting ≥50%, the median incubation period was between 24–48 hours, and the median duration of illness was between 12–60 hours). Evaluation of Kaplan et al.'s fourth criterion of a negative bacterial culture was excluded from this study to focus solely on clinical and epidemiologic criteria. In addition to the Kaplan criteria, the fever-to-vomiting ratio ≤ 1 and the diarrhea-to-vomiting ratio < 2.5 proposed by Hedberg et al. and Dalton et al., respectively, to

differentiate outbreaks due to enterotoxigenic *Escherichia coli* from those due to norovirus were evaluated [59, 60]. Outbreaks with missing information for each characteristic were excluded from analysis. OpenEpi, version 3.03 **(**Dean AG, Sullivan KM, Soe MM.), was used to calculate the sensitivity, specificity, and positive and negative predictive with their respective 95% confidence intervals for each characteristic. Likelihood ratios were also calculated to assess the diagnostic value of each characteristic, where values close to 1 are considered less useful [75].

Classification and Regression Tree Modeling

An optimal CART profile was created based on confirmed norovirus and non-viral outbreaks using the function *rpart* in R, version 3.1.1 (R Foundation for Statistical Computing, Vienna Austria). Because there were nearly twice as many confirmed norovirus outbreaks than confirmed non-viral outbreaks, bias in predictor selection was a concern. In order to reduce bias when creating the CART model, a random sample of 1,000 confirmed norovirus and 1,000 non-viral outbreaks were used for the model training data set. For each outcome, candidate clinical and epidemiologic characteristics that were missing in more than half of the outbreaks were excluded. Clinical and epidemiologic characteristics that were included in the training data set and assessed with the CART model included: the proportion of cases with bloody stools, the proportion of cases with diarrhea, the proportion of cases with fever, the proportion of cases with vomiting, the fever-to-vomit ratio, and the diarrhea-to-vomit ratio. Criteria for the best model selection consisted of low cross validation error relative to tree size, Cohen's kappa statistic, and the

likelihood ratio. The cross validation error was minimized by adjusting the cost complexity parameter to find the optimal tree size.

In order to determine how well the CART predictors performed with outbreak investigation data, the CART predictors were assessed among laboratory confirmed norovirus outbreaks and other non-viral outbreaks that had complete information for all of the significant CART model predictors. Performance of the CART model predictors was then evaluated by Cohen's kappa statistic for agreement. Additionally, CART model predictors were evaluated by sensitivity, specificity, PPV, NPV, and likelihood ratios. Suspect norovirus outbreaks and unknown etiology outbreaks were then evaluated with the CART predictors to assess the proportion of outbreaks that were likely attributed to norovirus based on our clinical and epidemiologic profile.

Results

Misclassification of Reported Outbreaks

To determine the misclassification of reported outbreaks, responses to "Etiology known" were compared to our definition of known etiology where one or more etiologies were listed by genus [Table 1]. Ninety-five outbreaks reported yes to "Etiology known" but did not list any etiology by genus. Two hundred and seventy-eight outbreaks reported no to "Etiology known" but provided one or more suspected etiologies by genus. Twenty-one outbreaks reported a missing "Etiology known" but listed 1 suspected etiology by genus. Overall, for outbreaks reported in NORS from 2009–2012, 3.9% (N=394) outbreaks were discordant and considered misclassified.

Agreement of NORS reported "Etiology known" to our definition of known etiology was assessed by Cohen's kappa statistic [Figure 1]. Of the 6,124 outbreaks that reported yes to "Etiology known", 6,029 were designated correctly as known etiology based on the number of etiologies provided. Similarly, of the 3,899 outbreaks that reported no or missing "Etiology known", 3,600 outbreaks were concordant with our definition of unknown etiology. Two hundred and ninety-nine outbreaks reported no or did not report "Etiology known" when etiologies were listed which could possibly reflect the misinterpretation of known etiology. Although some misreporting occurred, the overall agreement with Cohen's kappa was 0.92 (95% CI: 0.91-0.92) depicting a high level of agreement in reporting practices and generally appropriate interpretation of known and unknown etiologies.

In order to further assess misclassification of reporting by etiology (including both suspected and confirmed), responses to "Etiology known" were compared to reported etiologies by genus name [Table 2]. Laboratory confirmed and suspected norovirus outbreaks were the most misclassified outbreaks reported in NORS (N=218), although norovirus outbreaks were also the most frequently reported outbreaks in NORS (N=4,260, 42.5%). The highest rate of misclassification was seen with *Clostridium* spp, for which 19 (14%) of 133 outbreaks were misclassified. *Escherichia* spp. confirmed and suspected outbreaks were the least misclassified etiology (N=3).

Clinical and Epidemiologic Characteristics

Distributions of eight clinical and epidemiological characteristics used to distinguish norovirus outbreaks were assessed to determine the frequency of characteristics reported through NORS and how well each characteristic discerned from confirmed norovirus [Table 3]. Overall, the median incubation period was the least frequently reported characteristic for laboratory confirmed norovirus, suspected norovirus, laboratory confirmed non-viral, and unknown etiology outbreaks where a range of 2.2% to 5.4% of outbreaks reported that information. Among laboratory confirmed norovirus outbreaks, the proportion of cases with bloody stools was the second least reported characteristic (25.8%) followed by the fever-to-vomiting ratio (51.2%). However, for non-viral outbreaks, the proportion with bloody stools was reported in over half of the outbreaks (53.6%). There were no significant differences between suspected norovirus and confirmed norovirus outbreaks among any of the characteristics assessed, except for the fever-to-vomiting ratio. In contrast, non-viral outbreaks had different characteristic

distributions for all characteristics compared to laboratory confirmed norovirus (p-value < 0.0001). For unknown etiology outbreaks, the distributions for median incubation period, the median duration of illness, the proportion of cases with diarrhea, and the proportion of cases with fever were significantly different than distributions of laboratory confirmed norovirus outbreaks (p-value < 0.01). Conversely, the distributions for the proportion of cases with bloody stools, the proportion of cases with vomiting, the fever-to-vomiting ratio, and the diarrhea-to-vomiting ratio were not significantly distinguishable between unknown etiology outbreaks and laboratory confirmed norovirus outbreaks (p-value $\geq$ 0.05). Overall, suspected norovirus outbreaks shared clinical and epidemiologic characteristic distributions similar to those of laboratory confirmed norovirus except for the fever-to-vomiting ratio, whereas other non-viral outbreaks were statistically different from laboratory confirmed norovirus outbreaks; half of the characteristics for unknown etiology outbreaks looked similar to laboratory confirmed norovirus.

Clinical and Epidemiologic Profiles

To determine how well the Kaplan criteria, individual characteristics of the Kaplan criteria, fever-to-vomiting ratio, and diarrhea-to-vomiting ratio discriminate between confirmed norovirus and non-viral outbreaks, indices for the number of outbreaks with complete information for the criteria, the likelihood ratio, sensitivity, specificity, positive predictive value, negative predictive value, and were assessed [Table 4]. With NORS data, the Kaplan criteria were the most specific index with 100% specificity (95% CI: 83.2%-100%) but only a 63.9% sensitivity (95% CI: 54.5%-72.3%); a likelihood ratio was undefined due to the 100% specificity. Moreover, only 108 (3.7%) confirmed norovirus

outbreaks and 19 (1.2%) non-viral outbreaks had complete information for the Kaplan criteria. Among the individual components of the Kaplan criteria, the median duration of illness performed well with high likelihood ratio of 12.9, 79.6% sensitivity (95% CI: 77.2%-81.8%), and 93.8% specificity (95% CI: 89.3%-96.5%). Additionally, 1,192 (40.6%) laboratory confirmed norovirus outbreaks reported information for the duration of illness, but only 178 (11.3%) of non-viral outbreaks reported information for the duration of illness. The fever-to-vomiting ratio had a high sensitivity (97.8%, 95% CI: 96.9%-98.4%), a likelihood ratio of 2.3, and were reported in approximately 50% for both norovirus and non-viral outbreaks.

Classification and Regression Tree Modeling

Clinical and epidemiologic characteristics were evaluated among laboratory confirmed norovirus and laboratory confirmed non-viral outbreaks with classification and regression tree modeling to examine alternate characteristics and cut-points with NORS reported outbreaks. Various models were evaluated with the Cohen's kappa statistic and the likelihood ratio [see appendices]. The best model, selected based on the highest likelihood ratio of 12.5, evaluated the following predictors: the proportion of cases with bloody stools, the proportion of cases with diarrhea, the proportion of cases with fever, the proportion of cases with vomiting, the duration of illness, the fever-to-vomiting ratio, and the diarrhea-to-vomiting ratio. Through CART modeling, three significant predictors were selected with a cost complexity parameter of 0.02 that best distinguished norovirus from non-viral outbreaks [Figure 2]. The three predictors included fever-to-vomit ratio < 1,

proportion of cases with vomiting ≥ 0.34, and proportion of cases with bloody stools < 0.12.

Performance of Kaplan Criteria and CART Predictors

Among laboratory confirmed norovirus and laboratory confirmed non-viral outbreaks, the CART predictors performed better than the Kaplan criteria based on the Cohen's kappa statistic and the proportion of outbreaks with complete information for all criteria. The CART predictors had a high Cohen's kappa statistic (0.78, 95% CI: 0.75-0.81) demonstrating substantial agreement of laboratory confirmed norovirus and non-viral outbreaks with outbreaks that fit the criteria. Moreover, 706 (24.9%) laboratory confirmed norovirus outbreaks and 604 (39.1%) laboratory confirmed non-viral outbreaks had complete information for the criteria. The CART predictors were also effective in identifying norovirus outbreaks with an 86.0% sensitivity (95% CI: 83.2%-88.3%), a 93.1% specificity (95% CI: 90.7%-94.8%), and a likelihood ratio of 12.5. In comparison, the Kaplan criteria had only fair agreement with the Cohen's kappa statistic (0.34, 95% CI: 0.22-0.48). Furthermore, only 108 (3.7%) laboratory confirmed norovirus outbreaks and 19 (1.2%) laboratory confirmed non-viral outbreaks reported information for all of the Kaplan criteria.

The Kaplan criteria and CART predictors were also evaluated with suspected norovirus and unknown etiology outbreaks to determine the proportion of outbreaks that each would attribute to norovirus. Among suspected norovirus outbreaks, 324 (24.5%) had complete information for the CART predictors compared to the 24 (2.2%) with complete information for the Kaplan criteria. Among outbreaks of unknown etiology, 762 (20.6%)

had complete information for the CART predictors compared to 121 (3.3%) that had complete information for the Kaplan criteria. Additionally, application of the CART profile to suspected norovirus outbreaks determined that 262 (80.9%) outbreaks fit the criteria, which was similar to the proportion of laboratory confirmed norovirus outbreaks that fit the criteria (86.0%). Among unknown etiology outbreaks, 516 (67.7%) fit the criteria for the suggested CART predictors for norovirus.

Discussion

When evaluating clinical and epidemiologic profiles of AGE outbreaks reported in NORS, we found that the Kaplan criteria performed well with 100% specificity in distinguishing laboratory confirmed norovirus from laboratory confirmed non-viral outbreaks, but could not be applied to a majority outbreaks due to a lack of reported data. Moreover, the CART predictors performed better with Cohen's kappa statistic than the Kaplan criteria in distinguishing norovirus outbreaks from non-viral outbreaks and were reported in a greater number of outbreaks. Lastly, application of the CART predictors to unknown etiology outbreaks reported in NORS suggested the majority of those may in fact be attributable to norovirus. These findings suggest an alternative set of clinical and epidemiologic criteria to the Kaplan criteria that can be used to distinguish norovirus from other non-viral outbreaks based on frequently reported clinical and epidemiologic characteristics.

Among outbreaks reported in NORS, the Kaplan criteria were highly specific (100%), moderately sensitive (63.9%), but rarely reported (<5%). These findings were consistent with those from a similar evaluation by Turcios et al., in which they found the Kaplan criteria were highly specific (98.6%) and moderately sensitive (68.2%) [56]. High specificity of the criteria could be attributed to the combination of the criteria with specific cut-off values that must be met including: the incubation period from 24-48 hours, the duration of illness 12-60 hours, and the proportion of cases with vomiting ≥ 50% [58]. However, these criteria may be too specific for public health practice and may not always be available during outbreak investigations. Accurate exposure information can be difficult to determine for viral gastroenteritis where incubation periods are often short,

making it difficult to distinguish between primary and secondary cases in an outbreak [76]. In addition to the criteria being only moderately sensitive, only 2.9% (N=276) of the outbreaks evaluated (N=9484) had information reported for all of the Kaplan criteria. When evaluating the Kaplan criteria, the high specificity of the criteria were useful in distinguishing norovirus from other non-viral outbreaks but lack of information for each criterion limited their use to a small percentage of outbreaks with complete information for all criteria.

The CART predictors, including the fever-to-vomit ratio <1, the proportion of cases with vomiting ≥0.34, and the proportion of cases with bloody stools <0.12, performed better statistically in distinguishing laboratory confirmed norovirus outbreaks from non-viral outbreaks than the Kaplan criteria among NORS reported outbreaks. The CART predictors had higher sensitivity in detecting norovirus outbreaks (86.0%) and still relatively high specificity (93.1%) in distinguishing laboratory confirmed norovirus outbreaks from laboratory confirmed non-viral outbreaks. Improved sensitivity of norovirus detection could be attributed to the CART predictor's lower cut-off value of 0.34 for the proportion of cases with vomiting in an outbreak. This cut-off was lower than Kaplan's proposed 50% or more cases with vomiting [35], which may be the result of the criterion being used in conjunction with the fever-to-vomiting ratio < 1. Additionally, compared to the Kaplan criteria, CART predictors were reported in over eight times as many outbreaks. The increased reporting of CART predictors is at least partly due to training the CART model with predictors that tended to report information and inclusion of only those predictors that had less than 50% missing values. Studies have illustrated bias in variable importance measures where potential predictors differ [77, 78]. Therefore,

bias was minimized by excluding predictors with ≥ 50% missing and training the model with an equal random sample of each outcome. Overall, the CART predictors were effective in distinguishing more laboratory confirmed norovirus outbreaks from laboratory confirmed non-viral outbreaks within NORS compared to the Kaplan criteria.

When applying the CART predictors to outbreaks reported in NORS, 80.9% of suspected norovirus and 67.7% of unknown etiology outbreaks would be attributed to norovirus. With NORS, no current estimates are available for unknown etiology outbreaks attributed to norovirus for all modes of transmission [69]. For suspected norovirus outbreaks and unknown etiology outbreaks, it is possible that misclassification with the predictors could occur for a variety of reasons. It is likely that a small proportion of reported outbreaks could have viral etiologies that exhibit similar clinical and epidemiologic characteristics to norovirus, including incubation period, diarrhea, fever, and vomiting [76, 79-82]. Viral pathogens are the most common cause of gastroenteritis in industrialized countries [83-85], and without diagnostic testing, other viral etiologies including sapovirus, rotavirus, astrovirus, or enteric adenovirus may have similar clinical and epidemiologic characteristics as norovirus including fever, diarrhea, and vomiting [76, 79-82]. However, the CART predictors have the potential for some false positives with a 93.1% specificity and some false negatives among norovirus outbreaks with outlying clinical and epidemiologic characteristics given the sensitivity of 86.0%. This misclassification was observed with 14% of laboratory confirmed norovirus outbreaks that do not fit the CART predictor criteria. Lastly, incomplete reporting of clinical and epidemiologic criteria among all cases in outbreaks could potentially bias the performance

of these predictors if the proportion of cases with symptoms reported does not fully represent the total number of cases in an outbreak.

There were several strengths and limitations to this study. A major limitation of the NORS data set was the lack of complete data reported for clinical and epidemiologic characteristics of interest especially for the AGE characteristics evaluated in the study. Although outbreaks reported through NORS did not have complete data for these characteristics, the CART model was still able to distinguish significant predictors to differentiate norovirus from non-viral outbreaks with the use of surrogate variables. We were also unable to directly compare the Kaplan criteria and the CART predictors by likelihood ratios due to the 100% specificity of the Kaplan criteria. However, we were able to compare the performance of alternative CART models by Cohen's Kappa statistic, the likelihood ratio, and the proportion of outbreaks with all information to ensure that we selected the best model [see appendices]. Lastly, due to the small number of viral outbreaks and lack of reported clinical and epidemiologic characteristic information, we were unable to differentiate norovirus outbreaks from those of other viral outbreaks and had to exclude them from our analysis

In conclusion, predictors from the CART model were the most effective clinical and epidemiologic profile to differentiate a larger proportion of norovirus from non-viral outbreaks with NORS reported data from 2009 to 2012. Although the Kaplan criteria still remain highly specific in identifying norovirus etiology among NORS reported outbreaks, a majority of the reported outbreaks lacked complete information to make a diagnosis with strictly those criteria alone. In the absence of laboratory testing, clinical and epidemiologic criteria have proven to be an effective alternative in ascribing norovirus etiology. With

ongoing improvements in surveillance and increased reporting of outbreaks, these alternative criteria can aid in the diagnosis of norovirus in outbreak investigations and lead to more targeted implementation of control measures.

Tables and Figures

Table 1. Misclassification of reported outbreaks by "Etiology known" and the number of etiologies reported by genus in NORS 2009-2012

No. Etiologies Reported by Genus	"Etiology known"			Total No. Outbreaks[4]
	Yes[1]	No[2]	Missing[3]	
0	95 (2.6%)	3,413 (92.4%)	187 (5.1%)	3,695 (36.9%)
1	5,865 (95.2%)	272 (4.4%)	21 (0.3%)	6,158 (61.4%)
2	147 (96.7%)	5 (3.3%)	0 (0.0%)	152 (1.5%)
3	12 (92.3%)	1 (7.7%)	0 (0.0%)	13 (0.1%)
4	2 (100.0%)	0 (0.0%)	0 (0.0%)	2 (0.0%)
5	3 (100.0%)	0 (0.0%)	0 (0.0%)	3 (0.0%)
Total[5]	6,124 (61.1%)	3,691 (36.8%)	208 (2.1%)	10,023

[1]Outbreaks reporting "Yes" to "Etiology Known" with row percent
[2]Outbreaks reporting "No" to "Etiology Known" with row percent
[3]Outbreaks with no reported response to "Etiology Known" with row percent
[4]Total number of outbreaks reported with column percent
[5]Total number of outbreaks reported with row percent
Blue cells indicate discordant responses to "Etiology known" and the number of etiologies reported by genus

"Etiology known"	Etiology		
	Known (+)	Unknown (-)	
Yes (+)	6,029	95	6,124
No & Missing (-)	299	3,600	3,899
	6,328	3,695	10,023
Cohen's Kappa (95% CI)		0.92 (0.91,0.92)	

Figure 1. Agreement of outbreaks by known etiology and reported responses to "Etiology known" with outbreaks in NORS 2009-2012.

Table 2. Misclassification of reported outbreaks by "Etiology known" by the reported etiology in NORS 2009-2012

Reported Etiology	"Etiology known"			Total No. Outbreaks[4]
	Yes[1]	No[2]	Missing[3]	
Norovirus	4,042 (94.9%)	203 (4.8%)	15 (0.4%)	4,260 (42.5%)
Clostridium spp.	114 (85.7%)	19 (14.3%)	0 (0.0%)	133 (1.3%)
Salmonella spp.	680 (98.7%)	6 (0.9%)	3 (0.4%)	689 (6.9%)
Campylobacter spp.	171 (97.7%)	4 (2.3%)	0 (0.0%)	175 (1.7%)
Shigella spp.	226 (97.8%)	5 (2.2%)	0 (0.0%)	231 (2.3%)
Escherichia spp.	214 (98.6%)	2 (0.9%)	1 (0.5%)	217 (2.2%)
Other[5]	418 (92.3%)	33 (7.3%)	2 (0.4%)	453 (4.5%)
Multiple[6]	164 (96.5%)	6 (3.5%)	0 (0.0%)	170 (1.7%)
Unknown	95 (2.6%)	3,413 (92.4%)	187 (5.1%)	3,695 (36.9%)
Total[7]	6,124 (61.1%)	3,691 (36.8%)	208 (2.1%)	10,023

[1]Outbreaks reporting "Yes" to "Etiology Known" with row percent

[2]Outbreaks reporting "No" to "Etiology Known" with row percent

[3]Outbreaks with no reported response to "Etiology Known" with row percent

[4]Total number of outbreaks reported with column percent

[5]Multiple indicates outbreaks with more than one etiology

[6]Other includes: Astrovirus (N=2), *Bacillus* spp. (N=38), *Brucella* spp. (N=1), Ciguatoxin (N=38), *Cryptosporidium* spp. (N=56), *Cyclospora* spp. (N=4), *Enterococcus* spp.(N=1), *Giardia* spp. (N=32), Hepatitis (N=10), Histamine (N=8), *Listeria* spp. (N=18), Mycotoxins (N=10), Other (N=16), Other-Bacterium (N=12), Other-Chemical (N=20), Other-Virus (N=36), Paralytic shellfish (N=2), Pesticides (N=2), Plant/Herbal Toxin (N=1), Rotavirus (N=20), Sapovirus (N=12), Scombroid toxin (N=40), *Streptococcus* spp. (N=1), *Trichinella* spp. (N=3), *Vibrio* spp. (N=28), *Yersinia* spp. (N=5)

[7]Total number of outbreaks reported with row percent

Blue cells indicate discordant responses to "Etiology known" and reported etiology

Table 3. Distribution of acute gastroenteritis (AGE) clinical and epidemiologic characteristics among confirmed norovirus, suspected norovirus, confirmed non-viral, and unknown etiology outbreaks in NORS 2009-2012

Characteristic[1]	N (%) with characteristic	Median	IQR (Q1, Q3)[2]	P-value[3]
Median incubation Period (hrs)				
Confirmed norovirus	156 (5.3%)	30.0	(24.0, 37.0)	REF
Suspected norovirus	43 (3.3%)	30.0	(24.0-34.0)	0.92
Non-viral	33 (2.1%)	60.0	(48.0, 120.0)	<.0001
Unknown	200 (5.4%)	24.0	(11.3, 34.0)	<.0001
Median duration of illness (hrs)				
Confirmed norovirus	1,192 (40.6%)	48.0	(24.0, 48.0)	REF
Suspected norovirus	378 (28.6%)	42.5	(24.0, 48.0)	0.63
Non-viral	177 (11.5%)	144.0	(96.0, 204.0)	<.0001
Unknown	1,184 (32.1%)	36.0	(24.0, 48.0)	<.0001
Proportion of cases with bloody stools				
Confirmed norovirus	759 (25.8%)	0.0	(0.0, 0.0)	REF
Suspected norovirus	355 (26.9%)	0.0	(0.0, 0.0)	0.56
Non-viral	827 (53.6%)	0.3	(0.04, 0.50)	<.0001
Unknown	911 (24.7%)	0.0	(0.0, 0.0)	0.29
Proportion of cases with diarrhea				
Confirmed norovirus	2,200 (74.9%)	0.86	(0.75, 0.98)	REF
Suspected norovirus	944 (71.5%)	0.88	(0.74, 1.00)	0.30
Non-viral	1,261 (81.7%)	1.00	(0.99, 1.00)	<.0001
Unknown	2,506 (67.8%)	0.94	(0.75, 1.00)	<.0001
Proportion cases with fever				
Confirmed norovirus	1,542 (52.5%)	0.22	(0.08, 0.40)	REF
Suspected norovirus	686 (51.9%)	0.26	(0.09, 0.45)	0.05
Non-viral	1,012 (65.5%)	0.58	(0.33, 0.83)	<.0001
Unknown	1,473 (39.9%)	0.18	(0.01, 0.43)	0.006
Proportion of cases with vomiting				
Confirmed norovirus	2,164 (73.6%)	0.72	(0.58, 0.87)	REF
Suspected norovirus	919 (69.6%)	0.71	(0.56, 0.89)	0.69
Non-viral	1,031 (66.8%)	0.39	(0.22, 0.60)	<.0001
Unknown	2,369 (64.1%)	0.75	(0.49, 1.00)	0.8
Fever-to-Vomiting Ratio				
Confirmed norovirus	1,506 (51.2%)	0.31	(0.12, 0.56)	REF
Suspected norovirus	665 (50.3%)	0.39	(0.17, 0.64)	0.002
Non-viral	795 (51.5%)	1.33	(1.00, 2.00)	<.0001
Unknown	1,329 (36.0%)	0.33	(0.05, 0.71)	0.84
Diarrhea-to-Vomiting Ratio				
Confirmed norovirus	2,139 (72.8%)	1.33	(1.00, 1.44)	REF
Suspected Norovirus	908 (68.7%)	1.12	(1.00, 1.50)	0.8
Non-viral	899 (58.2%)	2.00	(1.44, 3.40)	<.0001
Unknown	2,214 (59.9%)	1.00	(1.00, 1.67)	0.88

[1]Characteristics examined in laboratory confirmed norovirus outbreaks (N=2,939), suspected norovirus outbreaks (N=1,321), laboratory confirmed non-viral outbreaks (N=1,544), and unknown etiology outbreaks (N=3,694)
[2]IQR (Q1, Q3) is the interquartile range where Q1 is the 25 percentile and Q2 is the 75 percentile
[3] P-values were obtained by Kruskal-Wallis Tests with post-hoc Steel, Dwass comparisons to laboratory confirmed norovirus outbreaks

Table 4. Kaplan criteria and clinical and epidemiologic characteristics used to discriminate between norovirus or non-viral etiology outbreaks with NORS 2009-2012

Clinical and Epidemiologic Characteristics	Confirmed norovirus[1]	Confirmed non-viral[2]	Likelihood Ratio	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive Predictive Value, % (95% CI)	Negative Predictive Value, % (95% CI)
Kaplan et al[3]							
No. outbreaks that fit the criteria	69 (63.9%)	0 (0.0%)	Undefined	63.9 (54.5-72.3)	100 (83.2-100)	100 (94.7-100)	32.8 (22.1-45.6)
No outbreaks that did not fit the criteria	39 (36.1%)	19 (100.0%)					
Total No. outbreaks with all criteria	108 (3.7%)	19 (1.2%)					
Median Duration of Illness (hrs)							
12-60 hrs	949 (79.6%)	11 (6.2%)	12.9	79.6 (77.2-81.8)	93.8 (89.3-96.5)	98.9 (98.0 -99.4)	40.7 (36.1-45.6)
not 12-60 hrs	243 (20.4%)	167 (93.8%)					
Total No. outbreaks with all criteria	1,192 (40.6%)	178 (11.3%)					
Proportion with Vomiting							
≥ 50 %	1,857 (85.8%)	438 (41.7%)	2.1	85.8 (84.3-87.2)	58.5 (55.5-61.4)	80.9 (79.3-82.5)	66.8 (63.7-69.7)
< 50 %	307 (14.2%)	612 (58.3%)					
Total No. outbreaks with all criteria	2,164 (73.6%)	1,050 (66.8%)					
Median Incubation Period (hrs)							
24-48 hrs	117 (75.0%)	14 (41.2%)	1.8	75.0 (67.7-81.1)	58.8 (42.2-73.6)	89.3 (82.9-93.5)	33.9 (23.1-46.6)
not 24-48 hr	39 (25.0%)	20 (58.8%)					
Total No. outbreaks with all criteria	156 (5.3%)	34 (2.2%)					
Fever-to-Vomiting Ratio							
≤ 1	1,451 (97.8%)	343 (42.8%)	2.3	97.8 (96.9-98.4)	57.2 (53.7-60.6)	80.9 (79.0-82.6)	93.3 (90.7-95.2)
> 1	33 (2.2%)	458 (57.2%)					
Total No. outbreaks with all criteria	1,484 (50.5%)	801 (50.9%)					
Diarrhea-to-Vomiting Ratio							
< 2.5	2,050 (95.8%)	558 (61.3%)	1.6	95.8 (94.9-96.6)	36.7 (35.6-41.9)	78.6 (77.0-80.1)	79.8 (75.8-83.3)
≥ 2.5	89 (4.2%)	352 (38.7%)					
Total No. outbreaks with all criteria	2,139 (72.8%)	910 (57.9%)					

[1]Laboratory confirmed norovirus outbreaks (N=2,939)
[2]Laboratory confirmed non-viral outbreaks (N=1,573)
[3]Kaplan criteria includes vomiting ≥ 0.5 affected persons, median incubation period of 24-48 hours, and median duration of illness of 12-60 hours

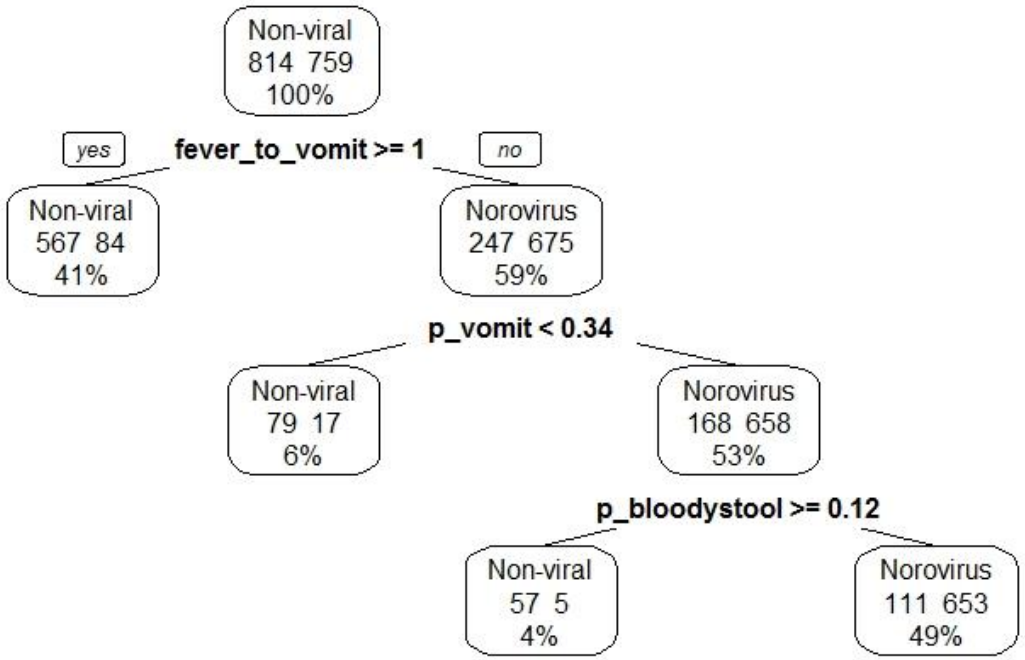


Figure 2. Classification and regression tree (CART) model of significant clinical and epidemiologic predictors for distinguishing confirmed norovirus outbreaks from confirmed non-viral etiology outbreaks with NORS 2009-2012. Each rectangular partition represents a node in the classification and regression tree. Within each node, the most frequent outcome is displayed first, followed by the number of outbreaks in each outcome (norovirus & non-viral), and the percentage of the most frequent outcome of the node. Text in bold represents significant clinical and epidemiologic predictors selected by the model, and its significant cut-off value. "Fever_to_vomit" represents the proportion of cases with fever divided by the proportion of cases with vomiting (fever-to-vomit ratio). "P_vomit" represents the proportion of cases with vomiting. "P_bloodystool" represents the proportion of cases with bloody stool. In the terminal nodes of the tree, outcomes represent the relative frequencies of *yes* and *no* answers to the predictors.

Table 5. Kaplan criteria and classification and regression tree (CART) predictors performance with NORS reported outbreaks 2009-2012

Clinical and Epidemiologic Profiles	Confirmed Norovirus[1]	Suspected Norovirus [2]	Confirmed non-viral[3]	Unknown[4]	Cohen's Kappa (95% CI)†	Likelihood Ratio†	Sensitivity, % (95% CI) †	Specificity, % (95% CI) †	Positive Predictive Value, % (95% CI) †	Negative Predictive Value, % (95% CI)†
Kaplan et al[5]										
No. outbreaks that fit the criteria	69 (63.9%)	12 (41.4%)	0 (0.0%)	35 (28.9%)	0.34 (0.22-0.48)	Undefined	63.9 (54.5-72.3)	100 (83.2-100)	100 (94.7-100)	32.8 (22.1-45.6)
No outbreaks that did not fit the criteria	39 (36.1%)	17 (58.6%)	19 (100.0%)	86 (71.1%)						
Total No. outbreaks with all criteria	108 (3.7%)	29 (2.2%)	19 (1.2%)	121 (3.3%)						
CART predictors[6]										
No. outbreaks that fit the criteria	607 (86.0%)	262 (80.9%)	42 (7.0%)	515 (67.7%)	0.78 (0.75-0.81)	12.5	86.0 (83.2-88.3)	93.1 (90.7-94.8)	93.5 (91.4-95.2)	85.0 (82.1-87.5)
No outbreaks that did not fit the criteria	99 (14.0%)	62 (19.1%)	562 (93.0%)	246 (32.3%)						
Total No. outbreaks with all criteria	706 (24.0%)	324 (24.5%)	604 (39.1%)	761 (20.6%)						

[1]Laboratory confirmed norovirus outbreaks (N=2,939)

[2]Suspected Norovirus outbreaks (N=1,321)

[3]Laboratory confirmed non-viral outbreaks (N=1,573)

[4]Unknown etiology outbreaks (N=3,694)

[5]Kaplan criteria includes vomiting ≥ 0.5 affected persons, median incubation period of 24-48 hours, and median duration of illness of 12-60 hours

[6]Classification and Regression Tree (CART) predictors include: fever-to-vomiting < 1, proportion of vomiting ≥ 0.34, and the proportion of bloody stools < 0.12

†Values determined by comparing confirmed norovirus with confirmed non-viral outbreaks

Chapter III. Public Health Implications

Public Health Implications

- Application of the predictors from our CART model could help distinguish norovirus outbreaks in the absence of laboratory testing.
- The Kaplan criteria are still highly specific for diagnosing norovirus outbreaks; however, the lack of information on incubation period and duration of illness limit their use for outbreaks reported through NORS.
- Awareness of the poorly reported characteristics could influence outbreak investigators to collect more information on these characteristics and improve reporting to aid in better diagnosis.
- For common-source or propagated outbreaks where the incubation periods cannot be assessed, the CART predictors could be better suited for identifying a likely norovirus etiology.
- Based on outbreaks reported through NORS, proportions of cases with fever and vomiting are often reported and can be used more often in outbreak investigations than other potential criteria.
- The use of the CART predictors in addition to the Kaplan criteria could aid in identifying likely norovirus etiology in the absence of laboratory diagnostics.
- With improved identification of norovirus in an outbreak setting, prompt and targeted public health action can be implemented to prevent and control propagation of the disease in an outbreak.

- Better clinical and epidemiologic criteria could also improve reporting and surveillance of norovirus outbreaks and aid in determining a more accurate proportion of reported outbreaks attributable to norovirus in the United States.

References

1. Lozano, R., et al., *Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010.* The Lancet, 2013. **380**(9859): p. 2095-2128.
2. Murray, C.J., et al., *Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010.* The lancet, 2013. **380**(9859): p. 2197-2223.
3. Scallan, E., et al., *Prevalence of diarrhoea in the community in Australia, Canada, Ireland, and the United States.* International Journal of Epidemiology, 2005. **34**(2): p. 454-460.
4. Neal, K.R., J. Hebden, and R. Spiller, *Prevalence of gastrointestinal symptoms six months after bacterial gastroenteritis and risk factors for development of the irritable bowel syndrome: postal survey of patients*. Vol. 314. 1997. 779.
5. Helms, M., et al., *Short and long term mortality associated with foodborne bacterial gastrointestinal infections: registry based studyCommentary: matched cohorts can be useful*. Vol. 326. 2003. 357.
6. Guerrant, R.L., et al., *Diarrhea as a cause and an effect of malnutrition: diarrhea prevents catch-up growth and malnutrition increases diarrhea frequency and duration.* The American journal of tropical medicine and hygiene, 1992. **47**(1 Pt 2): p. 28-35.
7. Ahmed, S.M., et al., *Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis.* The Lancet Infectious Diseases. **14**(8): p. 725-730.
8. Patel, M.M., et al., *Noroviruses: a comprehensive review.* Journal of Clinical Virology, 2009. **44**(1): p. 1-8.
9. Belliot, G., et al., *The Burden of Norovirus gastroenteritis: an important foodborne and healthcare-related infection.* Clinical Microbiology and Infection, 2014.
10. Teunis, P.F., et al., *Norwalk virus: how infectious is it?* Journal of medical virology, 2008. **80**(8): p. 1468-1476.
11. Blanton, L.H., et al., *Molecular and epidemiologic trends of caliciviruses associated with outbreaks of acute gastroenteritis in the United States, 2000–2004.* Journal of Infectious Diseases, 2006. **193**(3): p. 413-421.
12. Hall, A.J., et al., *Epidemiology of foodborne norovirus outbreaks, United States, 2001–2008.* Emerging infectious diseases, 2012. **18**(10): p. 1566.
13. Hall, A.J., et al., *Norovirus disease in the United States.* Emerg Infect Dis, 2013. **19**(8): p. 1198-1205.
14. Hoffmann, S., M.B. Batz, and J.G. Morris Jr, *Annual cost of illness and quality-adjusted life year losses in the United States due to 14 foodborne pathogens.* Journal of Food Protection®, 2012. **75**(7): p. 1292-1302.

15. Glass, R.I., U.D. Parashar, and M.K. Estes, *Norovirus gastroenteritis.* New England Journal of Medicine, 2009. **361**(18): p. 1776-1785.
16. Payne, D.C., et al., *Norovirus and medically attended gastroenteritis in US children.* New England Journal of Medicine, 2013. **368**(12): p. 1121-1130.
17. Bok, K. and K.Y. Green, *Norovirus gastroenteritis in immunocompromised patients.* New England Journal of Medicine, 2012. **367**(22): p. 2126-2132.
18. Anderson, A.D., et al., *Multistate Outbreak of Norwalk-like Virus Gastroenteritis Associated with a Common Caterer.* American Journal of Epidemiology, 2001. **154**(11): p. 1013-1019.
19. Fankhauser, R.L., et al., *Epidemiologic and Molecular Trends of "Norwalk-like Viruses" Associated with Outbreaks of Gastroenteritis in the United States.* Journal of Infectious Diseases, 2002. **186**(1): p. 1-7.
20. Widdowson, M.-A., et al., *Outbreaks of Acute Gastroenteritis on Cruise Ships and on Land: Identification of a Predominant Circulating Strain of Norovirus—United States, 2002.* Journal of Infectious Diseases, 2004. **190**(1): p. 27-36.
21. Control, C.f.D. and Prevention, *Norovirus activity--United States, 2006-2007.* MMWR. Morbidity and mortality weekly report, 2007. **56**(33): p. 842.
22. Zheng, D.-P., et al., *Molecular epidemiology of genogroup II-genotype 4 noroviruses in the United States between 1994 and 2006.* Journal of clinical microbiology, 2010. **48**(1): p. 168-177.
23. Cieslak, P., et al., *Recurring norovirus outbreaks in a long-term residential treatment facility-Oregon, 2007.* Morbidity and Mortality Weekly Report, 2009. **58**(25): p. 694-698.
24. Hall, A.J., et al., *Vital Signs: Foodborne Norovirus Outbreaks—United States, 2009–2012.* MMWR. Morbidity and mortality weekly report, 2014. **63**(22): p. 491.
25. Scallan, E., et al., *Foodborne illness acquired in the United States—major pathogens.* Emerg Infect Dis, 2011. **17**(1).
26. Hall, A.J., et al., *Updated norovirus outbreak management and disease prevention guidelines*. 2011: US Department of Health and Human Services, Centers for Disease Control and Prevention.
27. Lyman, W.H., et al., *Prospective study of etiologic agents of acute gastroenteritis outbreaks in child care centers.* The Journal of pediatrics, 2009. **154**(2): p. 253-257.
28. Roberts, C., et al., *Norovirus outbreaks on three college campuses-California, Michigan, and Wisconsin, 2008.* Morbidity and Mortality Weekly Report, 2009. **58**(39): p. 1095-1100.
29. Ahmed, S.M., B.A. Lopman, and K. Levy, *A systematic review and meta-analysis of the global seasonality of norovirus.* PloS one, 2013. **8**(10): p. e75922.
30. Kapikian, A.Z., et al., *Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis.* Journal of virology, 1972. **10**(5): p. 1075-1081.

31. Vinjé, J., *Advances in laboratory methods for detection and typing of norovirus.* Journal of clinical microbiology, 2014: p. JCM. 01535-14.
32. Tran, T.H., et al., *Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: global distribution of genogroups, genotypes and GII. 4 variants.* Journal of Clinical Virology, 2013. **56**(3): p. 269-277.
33. Cannon, J.L., et al., *Herd immunity to GII. 4 noroviruses is supported by outbreak patient sera.* Journal of virology, 2009. **83**(11): p. 5363-5374.
34. Lopman, B., et al., *Environmental transmission of norovirus gastroenteritis.* Current opinion in virology, 2012. **2**(1): p. 96-102.
35. Kaplan, J.E., et al., *The frequency of a Norwalk-like pattern of illness in outbreaks of acute gastroenteritis.* American journal of public health, 1982. **72**(12): p. 1329-1332.
36. Rockx, B., et al., *Natural history of human calicivirus infection: a prospective cohort study.* Clinical Infectious Diseases, 2002. **35**(3): p. 246-253.
37. Lopman, B.A., et al., *Clinical manifestation of norovirus gastroenteritis in health care settings.* Clinical Infectious Diseases, 2004. **39**(3): p. 318-324.
38. Trivedi, T.K., et al., *Hospitalizations and mortality associated with norovirus outbreaks in nursing homes, 2009-2010.* Jama, 2012. **308**(16): p. 1668-1675.
39. Kirby, A., et al., *Detection of norovirus in mouthwash samples from patients with acute gastroenteritis.* Journal of Clinical Virology, 2010. **48**(4): p. 285-287.
40. Atmar, R.L., et al., *Norwalk virus shedding after experimental human infection.* Emerging infectious diseases, 2008. **14**(10): p. 1553.
41. Graham, D.Y., et al., *Norwalk virus infection of volunteers: new insights based on improved assays.* Journal of Infectious Diseases, 1994. **170**(1): p. 34-43.
42. Phillips, G., et al., *Diagnosing norovirus-associated infectious intestinal disease using viral load.* BMC infectious diseases, 2009. **9**(1): p. 63.
43. Vega, E., et al., *Novel surveillance network for norovirus gastroenteritis outbreaks, United States.* Emerg Infect Dis, 2011. **17**(8): p. 1389-1395.
44. Gould, L.H., et al., *Surveillance for foodborne disease outbreaks—United States, 1998–2008.* MMWR Surveill Summ, 2013. **62**(2): p. 1-34.
45. Leshem, E., et al., *Effects and clinical significance of GII. 4 Sydney norovirus, United States, 2012–2013.* Emerg Infect Dis, 2013. **19**(8): p. 1231-8.
46. Sickbert-Bennett, E.E., et al., *Comparative efficacy of hand hygiene agents in the reduction of bacteria and viruses.* American journal of infection control, 2005. **33**(2): p. 67-77.
47. Harris, J., B. Lopman, and S. O'Brien, *Infection control measures for norovirus: a systematic review of outbreaks in semi-enclosed settings.* Journal of Hospital Infection, 2010. **74**(1): p. 1-9.
48. Nelson, K.E. and C.M. Williams, *Infectious disease epidemiology*. 2012: Jones & Bartlett Publishers.

49. Reingold, A.L., *Outbreak investigations--a perspective.* Emerging infectious diseases, 1998. **4**(1): p. 21.
50. Torok, M., et al., *Epidemic curves ahead.* Focus on Field Epidemiology, 2004. **1**.
51. Ambert-Balay, K. and P. Pothier, *Evaluation of 4 immunochromatographic tests for rapid detection of norovirus in faecal samples.* Journal of Clinical Virology, 2013. **56**(3): p. 278-282.
52. Costantini, V., et al., *Diagnostic accuracy and analytical sensitivity of IDEIA Norovirus assay for routine screening of human norovirus.* Journal of clinical microbiology, 2010. **48**(8): p. 2770-2778.
53. Ozawa, K., et al., *Norovirus infections in symptomatic and asymptomatic food handlers in Japan.* Journal of clinical microbiology, 2007. **45**(12): p. 3996-4005.
54. Phillips, G., et al., *Risk factors for symptomatic and asymptomatic norovirus infection in the community.* Epidemiology and infection, 2011. **139**(11): p. 1676-1686.
55. Hedberg, C., et al., *The use of clinical profiles in the investigation of foodborne outbreaks in restaurants: United States, 1982–1997.* Epidemiology and infection, 2008. **136**(01): p. 65-72.
56. Turcios, R.M., et al., *Reevaluation of Epidemiological Criteria for Identifying Outbreaks of Acute Gastroenteritis Due to Norovirus: United States, 1998–2000.* Clinical Infectious Diseases, 2006. **42**(7): p. 964-969.
57. Hall, J., et al., *Epidemiologic profiling: evaluating foodborne outbreaks for which no pathogen was isolated by routine laboratory testing: United States, 1982–9.* Epidemiology and infection, 2001. **127**(03): p. 381-387.
58. KAPLAN, J.E., et al., *Epidemiology of Norwalk gastroenteritis and the role of Norwalk virus in outbreaks of acute nonbacterial gastroenteritis.* Annals of Internal Medicine, 1982. **96**(6_part_1): p. 756-761.
59. Hedberg, C.W. and M.T. Osterholm, *Outbreaks of food-borne and waterborne viral gastroenteritis.* Clinical microbiology reviews, 1993. **6**(3): p. 199-210.
60. Dalton, C., et al., *Outbreaks of enterotoxigenic Escherichia coli infection in American adults: a clinical and epidemiologic profile.* Epidemiology and Infection, 1999. **123**(01): p. 9-16.
61. Patel, R.B., et al., *Demographic and Clinical Predictors of Mortality from Highly Pathogenic Avian Influenza A (H5N1) Virus Infection: CART Analysis of International Cases.* PloS one, 2014. **9**(3): p. e91630.
62. Rovlias, A. and S. Kotsou, *Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables.* Journal of neurotrauma, 2004. **21**(7): p. 886-893.
63. Lemon, S., et al., *Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression.* Annals of Behavioral Medicine, 2003. **26**(3): p. 172-181.

64. Therneau, T.M. and E.J. Atkinson, *An introduction to recursive partitioning using the RPART routines*. 1997, Technical Report 61. URL http://www. mayo. edu/hsr/techrpt/61. pdf.
65. Strobl, C., J. Malley, and G. Tutz, *An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests.* Psychological methods, 2009. **14**(4): p. 323.
66. De'ath, G. and K.E. Fabricius, *Classification and regression trees: a powerful yet simple technique for ecological data analysis.* Ecology, 2000. **81**(11): p. 3178-3192.
67. Scallan, E., et al., *Foodborne Illness Acquired in the United States—Unspecified Agents.* Emerging Infectious Diseases, 2011. **17**(1): p. 16-22.
68. Bull, R.A., et al., *Emergence of a new norovirus genotype II. 4 variant associated with global outbreaks of gastroenteritis.* Journal of clinical microbiology, 2006. **44**(2): p. 327-333.
69. Hall, A.J., et al., *Acute Gastroenteritis Surveillance through the National Outbreak Reporting System, United States.* Emerging Infectious Diseases, 2013. **19**(8): p. 1305-1309.
70. Lopman, B.A., et al., *Two epidemiologic patterns of norovirus outbreaks: surveillance in England and Wales, 1992–2000.* Emerging infectious diseases, 2003. **9**(1): p. 71.
71. Breiman, L., et al., *Classification and regression trees*. 1984: CRC press.
72. Viera, A.J. and J.M. Garrett, *Understanding interobserver agreement: the kappa statistic.* Fam Med, 2005. **37**(5): p. 360-363.
73. Kruskal, W.H. and W.A. Wallis, *Use of Ranks in One-Criterion Variance Analysis.* Journal of the American Statistical Association, 1952. **47**(260): p. 583-621.
74. Neuhäuser, M. and F. Bretz, *Nonparametric All-Pairs Multiple Comparisons.* Biometrical Journal, 2001. **43**(5): p. 571-580.
75. McGee, S., *Simplifying likelihood ratios.* Journal of general internal medicine, 2002. **17**(8): p. 647-650.
76. Lee, R.M., et al., *Incubation periods of viral gastroenteritis: a systematic review.* BMC infectious diseases, 2013. **13**(1): p. 446.
77. Strobl, C., et al., *Bias in random forest variable importance measures: Illustrations, sources and a solution.* BMC Bioinformatics, 2007. **8**: p. 25-25.
78. Hothorn, T., K. Hornik, and A. Zeileis, *Unbiased recursive partitioning: A conditional inference framework.* Journal of Computational and Graphical statistics, 2006. **15**(3): p. 651-674.
79. Sai, L., et al., *Epidemiology and clinical features of rotavirus and norovirus infection among children in Ji'nan, China.* Virol J, 2013. **10**(302,422).
80. Le Guyader, F.S., et al., *Aichi virus, norovirus, astrovirus, enterovirus, and rotavirus involved in clinical cases from a French oyster-related gastroenteritis outbreak.* Journal of Clinical Microbiology, 2008. **46**(12): p. 4011-4017.

81. Tran, A., et al., *Prevalence of rotavirus, adenovirus, norovirus, and astrovirus infections and coinfections among hospitalized children in northern France.* Journal of clinical microbiology, 2010. **48**(5): p. 1943-1946.
82. Logan, C., J.J. O'Leary, and N. O'Sullivan, *Real-time reverse transcription PCR detection of norovirus, sapovirus and astrovirus as causative agents of acute viral gastroenteritis.* Journal of virological methods, 2007. **146**(1): p. 36-44.
83. Caul, E.O., *Viral gastroenteritis: small round structured viruses, caliciviruses and astroviruses. Part I. The clinical and diagnostic perspective.* Journal of clinical pathology, 1996. **49**(11): p. 874-880.
84. de Wit, M.A.S., et al., *Sensor, a Population-based Cohort Study on Gastroenteritis in the Netherlands: Incidence and Etiology.* American Journal of Epidemiology, 2001. **154**(7): p. 666-674.
85. Tompkins, D., et al., *A study of infectious intestinal disease in England: microbiological findings in cases and controls.* Communicable disease and public health/PHLS, 1999. **2**(2): p. 108-113.

Appendices

A. NORS form

General

National Outbreak Reporting System

Foodborne Disease Transmission, Person-to-Person Disease Transmission, Animal Contact

This form is used to report enteric foodborne, person-to-person, and animal contact-related disease outbreak investigations. This form has 5 sections, General, Etiology, Settings, Animal Contact, and Food, as indicated by tabs at the top of each page. ***Complete the General and Etiology tabs for all modes of transmission and complete additional sections as indicated by the mode of transmission.*** *Please complete as much of all sections as possible.*

CDC USE ONLY

CDC Report ID _____ State Report ID _____

Form Approved
OMB No. 0920-0004

Primary Mode of Transmission ***(check one)***

- ☐ Food *(complete General, Etiology, and Food tabs)*
- ■ Water *(complete CDC 52.12)*
- ☐ Animal contact *(complete General, Etiology, and Animal Contact tabs)*
- ☐ Person-to-person *(complete General, Etiology, and Settings tabs)*
- ☐ Environmental contamination other than food/water *(complete General, Etiology, and Settings tabs)*
- ☐ Other/Unknown *(complete General, Etiology, and Settings tabs)*

Investigation Methods ***(check all that apply)***

- ☐ Interviews only of ill persons
- ☐ Case-control study
- ☐ Cohort study
- ☐ Food preparation review
- ☐ Water system assessment: Drinking water
- ☐ Water system assessment: Nonpotable water
- ☐ Treated or untreated recreational water venue assessment
- ☐ Investigation at factory/production/treatment plant
- ☐ Investigation at original source (e.g., farm, water source, etc.)
- ☐ Food product or bottled water traceback
- ☐ Environment/food/water sample testing
- ☐ Other

Comments

Dates (mm/dd/yyyy)

Date first case became ill *(required)* ____/_____/_____ Date last case became ill ____/_____/_____

Date of initial exposure____/_____/_____ Date of last exposure____/_____/_____

Date of report to CDC *(other than this form)* ____/_____/_____

Date of notification to State/Territory or Local/Tribal Health Authorities ____/_____/_____

Geographic Location

Reporting state: _____

- ☐ Exposure occurred in multiple states
- ☐ Exposure occurred in a single state, but cases resided in multiple states

Other states: _____

Reporting county: _____

- ☐ Exposure occurred in multiple counties in reporting state
- ☐ Exposure occurred in a single county, but cases resided in multiple counties in reporting state

Other counties: _____

City/Town/Place of exposure: _____
(*Do not include proprietary or private facility names*)

Primary Cases

Number of primary cases				**Sex** *(number or percent of the primary cases)*	
Lab-confirmed primary cases	#		Male	#	%
Probable primary cases	#		Female	#	%
Estimated total primary cases	#		Unknown	#	%

Primary Case Outcomes	**# Cases**	**Total # of cases for whom info is available**		**Age** *(number or percent of the primary cases)*				
Died	#	#	<1 year	#	%	20–49 years	#	%
Hospitalized	#	#	1–4 years	#	%	50–74 years	#	%
Visited Emergency Room	#	#	5–9 years	#	%	≥ 75 years	#	%
Visited health care provider (excluding ER visits)	#	#	10–19 years	#	%	Unknown	#	%

CDC 52.13 Rev. 05/2013 National Outbreak Reporting System CS235353 1

General

Incubation Period, Duration of Illness, Signs or Symptoms for Primary Cases Only

Shortest		Min, Hours, Days	Shortest		Min, Hours, Days
Median		Min, Hours, Days	Median		Min, Hours, Days
Longest		Min, Hours, Days	Longest		Min, Hours, Days
Total # of cases for whom info is available			Total # of cases for whom info is available		
☐ Unknown incubation period			☐ Unknown duration of illness		

Signs or Symptoms *(*Refer to terms from appendix, if appropriate, to describe other common characteristics of cases.)*

Feature	# Cases with signs or symptoms	Total # of cases for whom info is available
Vomiting		
Diarrhea		
Bloody stools		
Fever		
Abdominal cramps		
HUS		
Asymptomatic		
*		
*		
*		

Secondary Cases

Mode of secondary transmission *(check all that apply)*	Number of secondary cases	
☐ Food ☐ Water ☐ Animal contact ☐ Person-to-person ☐ Environmental contamination other than food/water ☐ Other/Unknown	Lab-confirmed secondary cases	#
	Probable secondary cases	#
	Estimated total secondary cases	#
	Estimated total cases (Primary + Secondary)	#

Environmental Health Specialists Network *(if applicable)*

EHS-Net Evaluation ID: 1.) _____ 2.) _____ 3.) _____ 4.) _____

Traceback *(for food and bottled water only, not public water)*

☐ Please check if traceback conducted

Source name *(if publicly available)*	Source type *(e.g., poultry farm, tomato processing plant, bottled water factory)*	Location of source		Traceback Comments
		State	Country	

Recall

☐ Please check if any food or bottled water product was recalled

Type of item recalled:

Comments:

Reporting Agency

Agency name: _____ E-mail: _____

Contact name: _____ Phone no.: _____

Contact title: _____ Fax no.: _____

General Remarks *Briefly describe important aspects of the outbreak not covered above. Please indicate if any adverse outcomes occurred in special populations (e.g., pregnant women, immunocompromised persons.)*

Etiology Section – complete for all modes of transmission except Water

Etiology known? ☐ Yes ☐ No

If etiology is *unknown*, were patient specimens collected? ☐ Yes ☐ No ☐ Unknown

If *yes*, how many specimens collected? (*provide numeric value*) _____

What were they tested for? *(check all that apply)* ☐ Bacteria ☐ Chemicals/Toxins ☐ Viruses ☐ Parasites

Etiology ***(Name the bacterium, chemical/toxin, virus, or parasite. If available, include the serotype and other characteristics such as phage type, virulence factors, and metabolic profile. Confirmation criteria available at http://www.cdc.gov/outbreaknet/references_resources/guide_confirming_diagnosis.html or MMWR 2000/Vol. 49/SS-1/App. B)***

Genus	Species	Serotype/Genotype	Confirmed outbreak etiology	Other characteristics	Detected in*	# Of Lab-Confirmed cases
			☐ yes			
			☐ yes			
			☐ yes			
			☐ yes			

*Detected in *(choose all that apply)*: 1 - patient specimen 2 - food specimen 3 - environment specimen 4 - food worker specimen

Isolates/Strains ***(For bacterial pathogens, provide a representative for each distinct pattern. For viral pathogens, provide CaliciNet key, outbreak number, sequenced region, and genotype for each distinct strain.)***

State Lab ID/ CaliciNet Key	CDC PulseNet or CaliciNet Outbreak Number	CDC PulseNet Pattern Designation for Enzyme 1	CDC PulseNet Pattern Designation for Enzyme 2	CaliciNet Sequenced Region/Other Molecular Designation 1	CaliciNet Genotype/ Other Molecular Designation 2

☐ Camp ☐ Hotel ☐ Private setting (residential home) ☐ School
☐ Child day care ☐ Nursing home ☐ Religious facility ☐ Ship
☐ Community-wide ☐ Prison or detention facility ☐ Restaurant ☐ Workplace
☐ Hospital ☐ Other, please specify: _____

Attack rates for major setting of exposure

Group *(based on setting)*	Estimated exposed in major setting*	Estimated ill in major setting	Crude attack rate *[(estimated ill / estimated exposed) x 100]*
residents, guests, passengers, patients, etc.			
staff, crew, etc.			

*e.g., number of persons on ship, number of residents in nursing home or affected ward

Other settings of exposure ***(choose all that apply)***

☐ Camp ☐ Hotel ☐ Private setting (residential home) ☐ School
☐ Child day care ☐ Nursing home ☐ Religious facility ☐ Ship
☐ Community-wide ☐ Prison or detention facility ☐ Restaurant ☐ Workplace
☐ Hospital ☐ Other, please specify: _____

Animal Contact Section – complete for animal contact primary mode of transmission

Setting of exposure	Type of animal	

3

Food

Contributing Factors *(check all that contributed to this outbreak)*

☐ Contributing factors unknown

Contamination Factor

☐ C1 ☐ C2 ☐ C3 ☐ C4 ☐ C5 ☐ C6 ☐ C7 ☐ C8 ☐ C9 ☐ C10 ☐ C11 ☐ C12 ☐ C13 ☐ C14 ☐ C15 ☐ C-N/A

Proliferation/Amplification Factor *(bacterial outbreaks only)*

☐ P1 ☐ P2 ☐ P3 ☐ P4 ☐ P5 ☐ P6 ☐ P7 ☐ P8 ☐ P9 ☐ P10 ☐ P11 ☐ P12 ☐ P-N/A

Survival Factor

☐ S1 ☐ S2 ☐ S3 ☐ S4 ☐ S5 ☐ S-N/A

The confirmed or suspected point of contamination *(check one)*

☐ Before preparation ☐ Preparation

If *'Before Preparation'*: ☐ Pre-Harvest ☐ Processing ☐ Unknown

Reason suspected *(check all that apply)*

☐ Environmental evidence	☐ Laboratory evidence
☐ Epidemiologic evidence	☐ Prior experience makes this a likely source

Was food-worker implicated as the source of contamination? ☐ Yes ☐ No

If *yes*, please check only one of the following:

- ☐ Laboratory ***and*** epidemiologic evidence
- ☐ Epidemiologic evidence
- ☐ Laboratory evidence
- ☐ Prior experience makes this a likely source

School Questions

(Complete this section only if "school" is checked in either sections "Location where food was prepared" or "Location of exposure (where food was eaten)").

1. Did the outbreak involve a single or multiple schools?

☐ Single

☐ Multiple (number of schools____)

2. School characteristics *(for all involved students in all involved schools)*

a. Total approximate enrollment

____(*number of students*)

☐ Unknown or undetermined

b. Grade level(s)

☐ Preschool

☐ Grade school *(grades K-12)*

Please check all grades affected: ☐ K ☐ 1st ☐ 2nd ☐ 3rd ☐ 4th ☐ 5th ☐ 6th ☐ 7th ☐ 8th ☐ 9th ☐ 10th ☐ 11th ☐ 12th

☐ College/university/technical school

☐ Unknown or Undetermined

c. Primary funding of involved schools

☐ Public

☐ Private

☐ Unknown

3. Describe the preparation of the implicated item: *(check all that apply)*

- ☐ Heat and serve *(item mostly prepared or cooked off-site, reheated on-site)*
- ☐ Served a-la-carte
- ☐ Serve only *(preheated or served cold)*
- ☐ Cooked on-site using primary ingredients
- ☐ Provided by a food service management company
- ☐ Provided by a fast-food vendor
- ☐ Provided by a pre-plate company
- ☐ Part of a club or fundraising event
- ☐ Made in the classroom
- ☐ Brought by a student/teacher/parent
- ☐ Other (*describe in General Remarks*)
- ☐ Unknown or Undetermined

4. How many times has the state, county or local health department inspected this school cafeteria or kitchen in the 12 months before the outbreak?*

- ☐ Once
- ☐ Twice
- ☐ More than two times
- ☐ Not inspected
- ☐ Unknown or Undetermined

*If multiple schools are involved, please answer according to the most affected school.

5. Does the school have a HACCP plan in place for the school feeding program?*

- ☐ Yes
- ☐ No
- ☐ Unknown or Undetermined

*If multiple schools are involved, please answer according to the most affected school.

5

6. Was implicated food item provided to the school through the National School Lunch/Breakfast Program?

☐ Yes
☐ No
☐ Unknown or Undetermined

If *yes*, was the implicated food item donated/purchased by:

☐ USDA through the Commodity Distribution Program
☐ The state/school authority
☐ Other (*describe in General Remarks*)
☐ Unknown or Undetermined

Ground Beef

1. What percentage of ill persons (*for whom information is available*) ate ground beef raw or undercooked? _____ %

2. Was ground beef case-ready? ☐ Yes ☐ No ☐ Unknown
(Case-ready ground beef is meat that comes from a manufacturer packaged for sale that is not altered or repackaged by the retailer.)

3. Was the beef ground or reground by the retailer?
☐ Yes ☐ No ☐ Unknown
If *yes*, was anything added to the beef during grinding (*such as shop trim or any product to alter the fat content*)?: _____

Additional *Salmonella* Questions
(Complete this section for Salmonella outbreaks)

1. Phage type(s) of patient isolates:

_____ *if RDNC* then include #* _____

_____ *if RDNC* then include #* _____

_____ *if RDNC* then include #* _____

_____ *if RDNC* then include #* _____

* Reacts, Does Not Conform

Eggs

1. Were eggs *(check all that apply)*

☐ in shell, unpasteurized?
☐ in shell, pasteurized?
☐ packaged liquid or dry?
☐ stored with inadequate refrigeration during or after sale?
☐ consumed raw?
☐ consumed undercooked?
☐ pooled?

2. Was Salmonella enteritidis found on the farm? ☐ Yes ☐ No ☐ Unknown

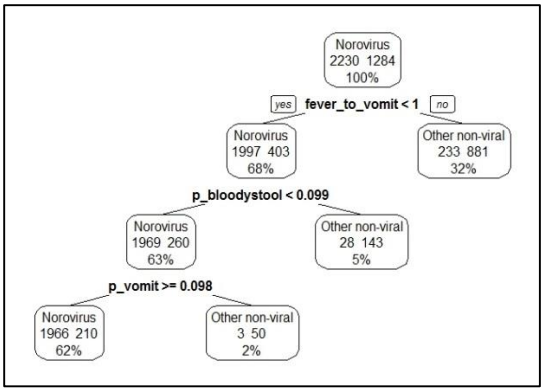
Egg Comment *(e.g., eggs and patients isolates matched by phage type)*: _____

Figure A. The National Outbreak Reporting System (NORS) form for foodborne, animal contact, person-to-person, environmental, or other transmission is shown. This form allows local and state health departments to report outbreaks to the Centers for Disease Control and Prevention for surveillance measures.

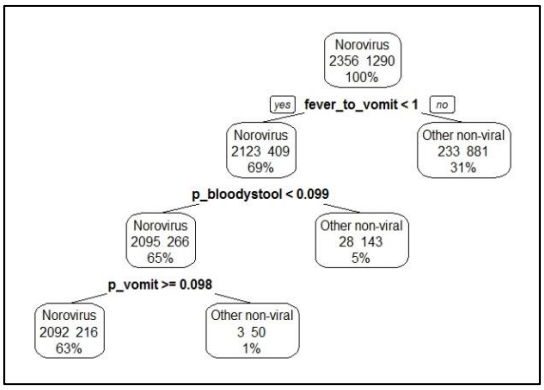
B. Alternative CART models

Alternative Classification and Regression Tree (CART) Models

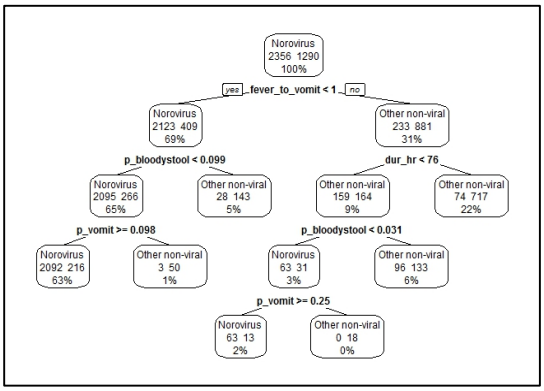
CART Models	Predictors Included	Significant Predictors	CP	Cohen's Kappa	Likelihood Ratio	Outbreaks with all information: Confirmed Norovirus	Other non-viral	Sensitivity	Specificity
Model 1	p_bloodystool	fever_to_vomit < 1	0.01	0.77 (0.73-0.80)	9.2	612	57	86.9 (84.0-89.0)	90.6 (88.0-92.6)
	P_diarrhea	p_bloodystool < 0.099				94	547		
	p_fever	p_vomit ≥ 0.098				24.0%	39.1%		
	fever_to_vomit								
	diarrhea_to_vomit								
Model 2	p_bloodystool	fever_to_vomit < 1	0.02	0.77 (0.73-0.80)	9.2	612	57	86.9 (84.0-89.0)	90.6 (88.0-92.6)
	P_diarrhea	p_bloodystool < 0.099				94	547		
	p_fever	p_vomit ≥ 0.098				24.0%	39.1%		
	fever_to_vomit								
	diarrhea_to_vomit								
	dur_hr								
Model 3	p_bloodystool	fever_to_vomit < 1	0.01	0.86 (.80-0.92)	11.6	253	8	95.5 (92.3-97.4)	91.8 (84.6-95.8)
	P_diarrhea	p_bloodystool < 0.099				12	89		
	p_fever	p_vomit ≥ 0.098				9.0%	6.3%		
	fever_to_vomit								
	diarrhea_to_vomit	fever_to_vomit > 1							
	dur_hr	dur_hr < 76							
		p_bloodystool < 0.031							
		p_vomit ≥ 0.25							
Model 4	p_bloodystool	fever_to_vomit < 1	0.02	0.78 0(.75-0.81)	12.5	607	42	86.0 (83.2-88.3)	93.1 (90.7-94.8)
(random sample 1000)	P_diarrhea	p_vomit ≥ 0.34				99	562		
	p_fever	p_bloodystool < 0.12				24.0%	39.1%		
	fever_to_vomit								
	diarrhea_to_vomit								
	dur_hr								
	(same if excluded dur_hr)								
Model 5	p_bloodystool	fever_to_vomit < 1	0.01	0.77 (0.70-0.85)	6.5	246	14	92.8 (89.1-95.4)	85.7 (77.2-91.2)
(random sample 1000)	P_diarrhea	p_vomit ≥ 0.34				19	83		
	p_fever	p_bloodystool < 0.12				9.0%	6.3%		
	fever_to_vomit								
	diarrhea_to_vomit	fever_to_vomit ≥ 1							
	dur_hr	dur_hr < 78							



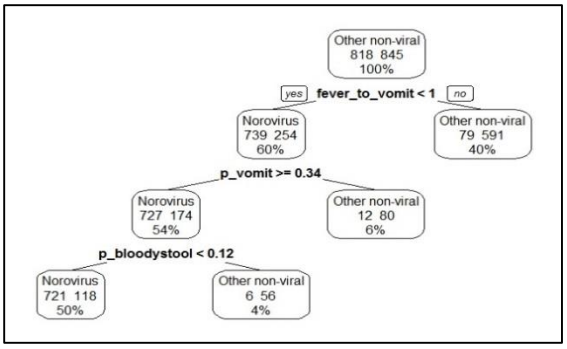
Model 1



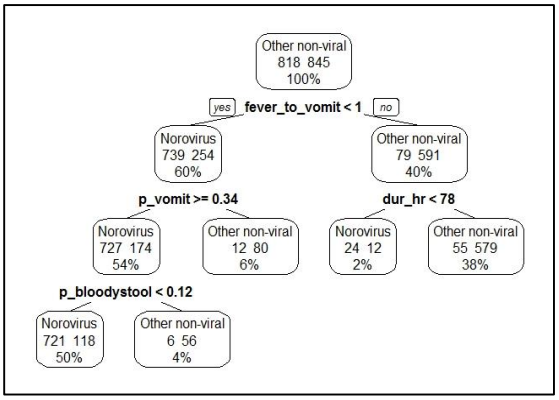
Model 2



Model 3



Model 4



Model 5

Figure B. Alternative classification and regression tree (CART) models created with different predictors and cost complexity parameters. Classification and regression tree (CART) model of significant clinical and epidemiologic predictors for distinguishing confirmed norovirus outbreaks from confirmed non-viral etiology outbreaks with NORS 2009-2012. Each rectangular partition represents a node in the classification and regression tree. Within each node, the most frequent outcome is displayed first, followed by the number of outbreaks in each outcome (norovirus & non-viral), and the percentage of the most frequent outcome of the node. Text in bold represents significant clinical and epidemiologic predictors selected by the model, and its significant cut-off value. "Fever_to_vomit" represents the proportion of cases with fever divided by the proportion of cases with vomiting (fever-to-vomit ratio). "P_vomit" represents the proportion of cases with vomiting. "P_bloodystool" represents the proportion of cases with bloody stool. "Dur_hr" represents the duration of illness reported. In the terminal nodes of the tree, outcomes represent the relative frequencies of *yes* and *no* answers to the predictors.