Molecular and Epidemiologic Trends of Caliciviruses Associated with Outbreaks of Acute Gastroenteritis in the United States, 2000–2004

Lenee H. Blanton,1,2 a Susan M. Adams,1,2 a R. Suzanne Beard,1 Gang Wei,1,2 Sandra N. Bulens,1,2 a Marc-Alain Widdowson,1 Roger I. Glass,1 and Stephan S. Monroe1

1Respiratory and Enteric Viruses Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, and 2Atlanta Research and Education Foundation, Decatur, Georgia

Between July 2000 and June 2004, fecal specimens from 270 outbreaks of acute gastroenteritis were sent to the Centers for Disease Control and Prevention by local or state health departments for calicivirus testing. Of the 226 outbreaks that met the criteria for inclusion in the present study, caliciviruses were detected in 184 (81%) by reverse-transcription polymerase chain reaction and nucleotide sequencing. Nursing homes, retirement centers, and hospitals were the most frequently reported settings, and person-to-person contact was the most common mode of transmission, followed by foodborne spread. Overall, genogroup II norovirus (NoV) strains were the most abundant (79%), followed by genogroup I NoV strains (19%) and sapovirus (2%). Nucleotide-sequence analysis indicated a great diversity of NoV strains and implicated the emergence of one particular sequence variant in outbreaks occurring between July 2002 and June 2003. The public health impact of caliciviruses will not be fully appreciated, nor will interventions be completely evaluated, until methods to detect these viruses are more routinely used.

Caliciviruses, the most common etiologic agent of acute gastroenteritis (AGE), are estimated to cause 23 million illnesses in the United States each year [1]. These viruses are transmitted through many routes—such as fecally contaminated food or water, person-to-person contact, and contaminated environments [2–16]—and outbreaks are notoriously difficult to control. These outbreaks occur in a wide range of settings, including nursing homes, retirement centers, hospitals, restaurants, events with catered meals, schools, and cruise ships [2–16]. The calicivirus family includes 4 genera of viruses, and 2 of them, norovirus (NoV) and sapovirus (SaV), cause AGE in humans. NoVs are classified into 5 distinct genogroups, designated GI–GV, which are further divided into >25 genetic clusters [17] (D.-P. Zheng, personal communication). GI, GII, and GIV strains infect humans, GIII strains have been detected in cattle, and a GV strain was recently discovered in mice [18]. SaVs have been found only in humans and are most often associated with pediatric AGE. Although multiple NoV strains cocirculate in human populations, individual strains have dominantly emerged both in the United States and worldwide [15, 19]. In particular, strains that belong to GII/genetic cluster 4 (GII/4) have in recent years (1995–1996 and 2002–2003) been reported as epidemic strains in the United States, Ireland, England, The Netherlands, Germany, Japan, New Zealand, and Australia [4, 7–14].

In the present study, we describe epidemiologic and laboratory data from outbreaks of AGE occurring between July 2000 and June 2004 for which specimens were sent to the Centers for Disease Control and Pre-
vitation (CDC) to be tested for calicviruses, and we assess the role played by calicviruses as etiological agents of AGE outbreaks as well as their prevalences. To determine whether certain strains are associated with specific epidemiologic or clinical characteristics, we also examine the genetic diversity of NoV strains circulating in the United States. The results are compared with those of 2 previous studies [5, 6] that reviewed epidemiologic and laboratory data on AGE outbreaks reported to the CDC between January 1996 and June 2000. Also, studies from other laboratories around the world [4, 7–14] are examined, to gain additional insights into the importance of NoV infection and the circulation patterns of NoV strains.

**PATIENTS, MATERIALS, AND METHODS**

The Viral Gastroenteritis Laboratory at the CDC received reports and specimens from collaborators at state and local health departments who were seeking laboratory assistance for the investigation of outbreaks of suspected viral AGE. Up to 10 specimens from each outbreak were requested and were tested for the presence of NoVs with multiple primer sets. We examined fecal specimens submitted to the CDC from suspected calicivirus outbreaks occurring in the United States or aboard cruise ships docking at US ports during the 4-year period from July 2000 through June 2004 for which no bacterial pathogen was implicated. A calicivirus season was defined as the 12-month period from July through June of each year. A suspected calicivirus outbreak was defined as an outbreak of AGE in which the clinical presentation of patients was compatible with calicivirus infection; only those outbreaks for which ≥4 fecal specimens were available for testing were considered.

Epidemiologic data for each outbreak were collected on a standardized form and submitted to the CDC with each group of specimens. For the present study, we examined the setting of the outbreak; the implicated or suspected mode of transmission, as reported by the investigator; the size of the outbreak, including both the actual number of persons ill and the number susceptible; the age, sex, and symptom profile of affected persons; and the implicated calicivirus strain. Data on all of these variables were not available for all outbreaks.

The specimens were processed for calicviruses by use of a standard method of reverse-transcription polymerase chain reaction (RT-PCR). A 10% aqueous suspension of stool or emesis in a 1:1 mixture of sterile water and Vertrel-XF (DuPont) was clarified by centrifugation at 1700 g for 10 min. RNA was then extracted from the stool or emesis suspensions by use of the NucliSens automated nucleic acid extraction system (BioMérieux), in accordance with the manufacturer’s instructions for small specimen volumes. RT-PCR was performed with primers targeted to a 213-bp area of the polymerase gene, commonly referred to as region B [3].

Outbreaks for which the region B primer set produced negative or indeterminate results were further investigated using RT-PCR with additional primer sets for region C, region D, or region 5. The region C primer set amplifies a 322-bp region of open-reading frame (ORF) 2 that encodes the capsid gene [19]. The region D primer set, described by Vinje et al. [20], can distinguish between GI and GII NoV by amplifying a 177- or a 253-bp region of ORF2, respectively. The region D primer set was modified for the GII-specific reaction by adding Q solution (Qiagen) to the PCR mix and then annealing at 44°C. The region 5 primer set amplifies a region of the RNA polymerase gene in ORF1 that yields a 319-bp product specific for NoV or a 331-bp product specific for SaV [21]. Amplified products were detected on a 3% ethidium bromide–stained agarose gel. At least 2 positive specimens from each outbreak were chosen for nucleotide sequencing; they were prepared as described elsewhere [22] and were sequenced on an automated sequencer (model 3100; Applied Biosystems), using a Big Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA Polymerase FS (Applied Biosystems). All sequences were analyzed with the GCG suite of programs (Accelrys), including the PileUp program, to create a phylogram with uncorrected distances. On the basis of the resulting phylogram, each strain was placed into a genetic cluster, using the nomenclature system of Ando et al. [17] with slight modifications (D.-P. Zheng, personal communication). An outbreak was considered to be a confirmed calicivirus outbreak when at least 2 amplicons were sequenced and assigned to a genogroup.

Because the regions of the sequence determined in the present study were small (172 bp for region B, 277 bp for region C, 142 and 214 bp for region D, and 274 and 286 bp [SaV] for region 5), they were not individually submitted to GenBank. They are available from the authors on request.

**RESULTS**

**Outbreak characteristics.** Between July 2000 and June 2004, we received specimens from 270 AGE outbreaks sent by local or state health departments. Of these outbreaks, 226 met the criteria for inclusion in the present study (44 outbreaks were excluded because <4 fecal specimens were available for testing). A total of 2412 fecal and emesis samples from these 226 outbreaks were submitted for analysis, of which 1932 (80%) were tested by RT-PCR, with a median of 8 specimens tested per outbreak.

Outbreaks were reported from 30 states, the District of Columbia, and Puerto Rico and from cruise ships docking at US ports (figure 1). Four states submitted specimens from >10 outbreaks: Georgia (39), Ohio (26), West Virginia (14), and Maryland (13). Six states submitted specimens from 6 to 10 outbreaks, and 20 states provided specimens from ≤5 outbreaks. Twenty states sent no outbreak specimens at all during the study period.

The 226 outbreaks occurred in a variety of settings (table 1), but the most common locations reported were nursing homes,
Calicivirus Outbreaks in the United States

Figure 1. Distribution of outbreaks of acute gastroenteritis in the United States for which specimens were tested by the Centers for Disease Control and Prevention (CDC) between July 2000 and June 2004. For the 4 states that reported >10 outbreaks, the no. of outbreaks for which specimens were tested is shown in white circles (the black circle indicates the no. of outbreaks occurring on cruise ships docking at US ports). *States that perform diagnostic testing for noroviruses as of August 2004.

retirement centers, and hospitals (71 outbreaks [31%]), followed by restaurants and events with catered meals (64 outbreaks [28%]). Outbreaks were also reported in vacation settings, including cruise ships (37 outbreaks [16%]), and in schools and day-care centers (18 outbreaks [8%]). The median number of persons affected in all outbreaks was 45 (range, 4–547). The highest median attack rate was 50%, for restaurants and events with catered meals, whereas the median attack rate was 9% for vacation settings, including cruise ships. Person-to-person contact was the most commonly reported mode of transmission (80 outbreaks [35%]), followed by contaminated food (67 outbreaks [30%]) and water (12 outbreaks [5%]). No mode of transmission was determined for 67 outbreaks (30%).

Information on age was available for 1010 (50%) of the 2021 patients, and the median age was 47 years (range, 11 months to 102 years). Of the 1087 patients (54%) for whom information on sex was recorded, 615 (57%) were female.

Strain characterization. Of the 226 outbreaks, caliciviruses were detected in 184 (81%) by RT-PCR and nucleotide sequencing with primer sets for either region B, region C, region D, or region 5. In most confirmed outbreaks, the implicated strain was detected by RT-PCR with the region B primer set (177 outbreaks [96%]). In addition, the implicated strain was detected in 112 outbreaks (61%) with the region C primer set, in 35 outbreaks (19%) with the region D primer set, and in 10 outbreaks (5%) with the region 5 primer set. No outbreak specimens were tested using all 4 primer sets. Overall, GII NoV was the most frequently detected (79% of outbreaks), followed by GI NoV (19%); SaV was detected in 4 outbreaks (2%), 2 of which also had NoVs identified.

Of the 184 confirmed calicivirus outbreaks, 44 (24%) were associated with multiple calicivirus strains. Multiple GII NoV strains were identified in 28 outbreaks (15%), multiple GI NoV strains were identified in 28 outbreaks (15%), multiple GI NoV strains were identified in 3 outbreaks (2%), both GI and GII NoV strains were identified in 10 outbreaks (5%), both SaV and NoV strains were identified in 2 outbreaks (1%), and multiple SaV strains were identified in 1 outbreak (0.5%). A total of 278 NoV strains and 5 SaV strains were detected and sequenced during the study. By use of nucleotide-sequence comparisons, these strains were categorized into genogroups and genetic clusters (figure 2). Within a genogroup, clusters are assigned sequential numbers on the basis of the reporting of a
complete capsid protein gene sequence that differs by at least 20% in amino acid–sequence identity from other recognized sequences [17, 23].

The region B primer set amplifies a highly conserved region of ORF1 that is broadly reactive for NoV. These products are most commonly sequenced, but these region B sequences often do not allow for the unambiguous classification of NoV strains into individual clusters. Most sequences derived from amplimers only by use of the region B primer set could be designated only as GI/resolved groups or as GII/unresolved (91) (which includes GII/1, GII/4, GII/10, GII/12, and GII/13 candidate sequences), GII/unresolved (96) (which includes GII/6, GII/7, and GII/9 candidate sequences), or GII/unresolved (99) (which includes all other GII candidate sequences that significantly differ from a known cluster). To further resolve the NoV strains into clusters, additional RT-PCR primer sets, mainly for region C and region D, were used to amplify more-discriminating areas of the NoV genome.

The most common cluster found with the region B primer set was GII/unresolved (91) (71%), followed by GII/unresolved (96) (9%) and GI/3 (6%). When the region C primer set was used, the most common cluster identified was GII/4 (81%), followed by GII/3 (7%) and GII/10 (3%). In 42 outbreaks (19%), all specimens tested were negative by RT-PCR or were unsuccessfully sequenced. Classification of all detected strains into clusters by use of the region C and region D primer sets showed that GII/4 strains appeared to become epidemic strains over the last 2 years, being detected in 16% of outbreaks during the 2000–2001 calicivirus season, in 11% during the 2001–2002 season, in 61% during the 2002–2003 season, and in 60% during the 2003–2004 season. The strains detected during this period showed great diversity, and, of the 25 NoV clusters currently recognized in humans [17] (D.-P. Zheng, personal communication), 16 were identified in the present study (figure 2).

Molecular analysis of representative strains from 184 outbreaks sequenced using the region B primer set yielded 157 sequivars (distinct nucleic acid sequences). When analyzed on the basis of the 172-bp region B sequence, 43 (24%) of 180 strains from NoV outbreaks were identical to the sequivar provisionally referred to as “Farmington Hills” [15]. After further sequence analysis on the basis of the 277-bp region C sequence, these strains could be divided into 31 distinct sequivars that were closely related to Farmington Hills (figure 3).

The frequency of outbreaks was plotted by month and indicated a distinct winter seasonality (figure 4), but outbreaks did occur throughout the year. The number of outbreaks peaked between December and March of each calicivirus season, although no distinct peak was observed during the 2001–2002 season. Outbreaks of both GI and GII strains occurred throughout each season; however, between July 2002 and June 2004, GII strains accounted for 86% of all confirmed outbreaks, compared with 69% during the period from July 2000 through June 2002. Of note, GII/4 strains were detected in 61% of all confirmed outbreaks occurring between July 2002 and June 2004, compared with only 14% during the period from July 2000 through June 2002.

We identified several associations between epidemiologic characteristics and strains from particular calicivirus clusters. GII/4 strains were detected more frequently in confirmed calicivirus outbreaks occurring in closed or semiclosed settings than were other strains—they were detected in 65% of confirmed calicivirus outbreaks occurring in nursing homes, retirement centers, and hospitals; in 38% of outbreaks occurring in schools and day-care centers; and in 58% of outbreaks occurring in vacation settings, including cruise ships. In addition, GII/4 strains were significantly more common in confirmed calicivirus outbreaks involving person-to-person transmission (55%) than they were in those resulting from foodborne transmission (18%) (P < .001). The remaining outbreaks occurring in other settings and in those transmitted via food, water, or person-to-person contact were associated with a wide distribution of GI and GII strains, and no other significant epidemiologic associations with cluster types were evident. In addition, no geographic clustering of sequivars was observed.
In 2002, a sharp increase in the number of NoV outbreaks was observed both on cruise ships and on land [15]. Through nucleotide-sequence comparisons in region B, one NoV cluster was found to predominate, the above-mentioned Farmington Hills sequivar [15]. During the 2002–2003 calicivirus season, the Farmington Hills sequivar was responsible for 36% of all confirmed outbreaks and 44% of all GII outbreaks. However, during the 2003–2004 calicivirus season, the former percentage dropped to 17% of all confirmed outbreaks. Most of the 28 outbreaks caused by strains of the Farmington Hills sequivar occurred in closed settings: 10 (36%) occurred in vacation settings, including cruise ships; 9 (32%) occurred in nursing homes, retirement centers, and hospitals; 4 (14%) occurred in schools and day-care centers, 3 (11%) occurred in restaurants and events with catered meals; and 2 (7%) occurred in community-based outbreaks. Transmission by person-to-person contact (12 [43%]) was most frequently reported, followed by foodborne (4 [14%]) and waterborne (1 [4%]) spread. In outbreaks occurring on cruise ships and in other vacation settings, the Farmington Hills sequivar was significantly more common than all other GI and GII strains (33% vs. 15%; \( P < .04 \)).

Because of the sudden emergence and predominance of a single NoV sequivar, the clinical characteristics of the outbreaks were reviewed. We compared the frequency of diarrhea and vomiting in outbreaks caused by the Farmington Hills sequivar with those in outbreaks caused by other sequivars. We found that the patients in the 28 outbreaks caused by the Farmington Hills sequivar were slightly more likely to be ill with diarrhea than were the patients in the 89 outbreaks due to other sequivars (87% vs. 80%; \( P = .03 \)) but that there was no signif-
Figure 3. Phylogenetic analysis of norovirus Farmington Hills outbreak sequivars, July 2000 through June 2004. The dendrogram was generated by neighbor joining on the basis of uncorrected distances calculated for 277 bp of region C, using the Distances and GrowTree programs of the GCG sequence-analysis suite (Accelrys). Sequivars from outbreaks associated with the Farmington Hills strain on the basis of region B sequences plus 1 prototype strain from each defined cluster were analyzed in region C. *Although Melksham was used as the prototype strain for genogroup II/genetic cluster 2 (GII/2), it shares 92.8% nucleotide identity with Snow Mountain when the region C primer set is used.

DISCUSSION

NoVs remain the most common cause of outbreaks of AGE in the United States. Our present results, combined with those of previous reports [5, 6], provide data on 549 outbreaks of AGE that have been investigated for calciviruses by the CDC over a period of 8 years. Of these, 471 outbreaks (86%) were confirmed to be attributable to NoVs by RT-PCR, demonstrating that NoVs are the leading cause of outbreaks of nonbacterial AGE in the United States. The setting profile for NoV outbreaks...
Figure 4. Distribution of gastroenteritis outbreaks \( (n = 226) \) by month and associated calicivirus strain, July 2000 through June 2004. The norovirus (NoV) genogroup (GI or GII) of implicated strains, sapoviruses (SaVs), and outbreaks for which all tested specimens were negative by reverse-transcription polymerase chain reaction are shown. *Refers to those outbreaks for which both GI and GII strains were detected; **Includes 2 outbreaks for which SaVs and NoVs were implicated.

changed over this 8-year period. From 1996 to 2000, the most common settings reported were restaurants and events with catered meals. Since 2000, the most common venues for an outbreak have been nursing homes, retirement centers, hospitals, and vacation settings, including cruise ships. The increase in the number of outbreaks at these latter settings may be the result of enhanced epidemiologic investigations, improved collection of specimens for NoV testing, and more-sensitive laboratory diagnostics. One explanation for the decrease in the number of NoV-related foodborne outbreaks investigated by the CDC is likely the increased number of states performing NoV diagnostics and reporting a greater number of foodborne outbreaks to the CDC’s Electronic Foodborne Outbreak Reporting System (eFORS). NoV was reported to eFORS in 37\% \( (n = 164) \) of all foodborne outbreaks with a confirmed etiology in 2000, in 34\% \( (n = 150) \) in 2001, in 40\% \( (n = 197) \) in 2002, and in 34\% \( (n = 139) \) in 2003, making it the most common cause of laboratory-confirmed AGE in this database [24]. In addition, in 2003, NoV was suspected in 139 (21\%) of 664 foodborne outbreaks of unknown etiology [24].

Although the diversity of NoV strains varied from season to season, an increase in the frequency of GII/4 strains and the Farmington Hills sequivar was observed from July 2002 and June 2004, similar to findings in other countries [4, 7–14], including Ireland, England, Germany, Japan, The Netherlands, New Zealand, and Australia. All of these surveillance studies used RT-PCR and nucleotide sequencing as a detection method, and some also used electron microscopy, hybridization techniques, and/or EIA as additional detection methods. In Hungary and Spain, GII/1 and GII/2 strains, respectively, were most commonly associated with outbreaks. In all countries except Hungary and Spain, GII/4 strains were the most commonly circulating outbreak strains. Also, in 6 of these 9 countries, outbreaks most commonly occurred in closed settings, including hospitals, nursing homes, and/or schools. In New Zealand and Japan, the most common settings were restaurants and events with catered meals; in Ireland, only outbreaks from hospitals and residential homes were included in the study. These findings may reflect biological differences, such as a mild increase in the severity of disease, or the requirement for a lower infectious dose. In addition, further research may be needed to determine the virulence of the Farmington Hills sequivar, its persistence in the environment, and its susceptibility to disinfectants. Of note, SaVs are typically associated with AGE in children, yet the 4 outbreaks in which SaVs were implicated in the present study occurred among adults on 2 cruise ships, at a restaurant, and at a conference.

There are several limitations to the present study. Because of the passive nature of the reporting system, it is not known whether any bias in selection toward certain settings exists.
Reports from other countries, such as the United Kingdom [25], suggest that foodborne NoV outbreaks are less common than we found in this study. Because foodborne outbreaks of AGE are reportable nationally in the United States, these may be investigated more thoroughly than are nonfoodborne outbreaks. Underreporting of NoV outbreaks remains a problem. Because people recover within 2–3 days, they may not report an illness to health authorities or go to a physician. Unlike the situation for bacterial agents, which are detected by the testing of specimens routinely submitted to clinical laboratories, no routine, commercially available test for NoV exists, and so ill persons who submit a stool specimen are not tested for NoV. Health departments are often reluctant to investigate outbreaks of virus-like illnesses, because of the difficulty in diagnosing and implementing successful interventions. Use of the region 5 primer set, which detects both NoVs and SaVs, was not universally applied to all outbreaks; therefore, the true incidence of coinfection with NoV and SaV could be higher than that reported here. The result of these factors is that the true burden of disease caused by NoVs in the United States remains unknown. Our study presents only the results for specimens tested at the CDC; yet 46 state and 8 county public health laboratories are performing RT-PCR diagnostics for NoV.

Through improved diagnostic methods, NoVs have been increasingly identified in a greater number of outbreaks. Improved and more-rapid detection of NoVs has been made possible by the introduction of a new RT-PCR primer set, for region D, and a real-time RT-PCR platform. However, the capacity for NoV testing exists solely in state and select local and reference laboratories and, therefore, is not currently available in clinical laboratories. Only through improved specimen collection, thorough investigations of AGE outbreaks, and more-extensive viral testing will we obtain a more-complete estimate of the incidence of NoV infection in the United States. A recent study estimated that, if all specimens were tested for viruses, one-half of all foodborne outbreaks in the United States would be attributed to NoV [16]. The data reported here confirm that NoVs are overwhelmingly the most common cause of outbreaks of nonbacterial AGE in the United States. The recently discovered murine NoV [26] may lead to greater knowledge of the biological characteristics of caliciviruses. Future studies will provide a more-comprehensive understanding of calicivirus outbreaks and could lead to improved strategies for their control and prevention.

Acknowledgments

We thank all of the local and state health departments that conduct epidemiologic and laboratory investigations of acute gastroenteritis, both for conducting such investigations and for their continued collaboration. We also thank Du-Ping Zheng, for use of newly discovered NoV clusters; our Epidemic Intelligence Service officers and laboratory personnel, especially Du-Ping Zheng and Leslie Hadley, who helped in outbreak investigations; and Claudia Chesley, for providing editorial comments.

References

20. Vinje J, Hamidjaja RA, Sobsey MD. Development and application of a capsid VPI (region D) based reverse transcription PCR assay for...


