Using Molecular Epidemiology To Trace Transmission of Nosocomial Norovirus Infection

Faizel H. A. Sukhrie,1,2* Matthias F. C. Beersma,2 Albert Wong,1 Bas van der Veer,1 Harry Vennema,1 Jolanda Bogerman,2 and Marion Koopmans1,2

Laboratory for Infectious Diseases and Perinatal Screening, National Institute for Public Health and the Environment, Bilthoven,1 and Departments of Virology, Erasmus Medical Center, Rotterdam,2 Netherlands

Received 15 July 2010/Returned for modification 15 October 2010/Accepted 2 December 2010

Nosocomial norovirus (NoV) infection is common and may lead to complications in vulnerable hospitalized patients. Understanding sources and modes of transmission of noroviruses within health care settings will support the design of evidence-based strategies for reducing introduction and further spread. We sequenced a highly variable segment of the genome to identify possible clusters in patients with and without acute gastroenteritis who were hospitalized in the period 2002-2007. Admission and sampling dates were used to separate patients with nosocomial infection from those without nosocomial infection. Epidemiological clustering retrieved 22 clusters, defined as ≥2 patients with nosocomial infection on the same ward within 5 days. In total, 264 patients (of 2,458 tested) were diagnosed with NoV infection, and 61% of the patient strains could be genotyped. Of those, 51% (n = 82) belonged to GI.4, 34% (n = 54) belonged to GI.3, and 15% (n = 24) belonged to other genotypes (GI.5, GI.2, GII.3, GII.4, and GII.7). In children’s wards, GII.3 strains were associated with nosocomial spread more often than other viruses were, whereas in adults this was the case for GI.4 strains. Sequence alignment recognized 11 new clusters based on identical P2 domains (4 GI.3 and 7 GI.4 clusters), involving patients in different wards. This increased the total number of recognized clusters by 50%. Five of these clusters involved at least one outpatient, providing a possible target for improvement of infection control. We concluded that the use of sequence-based typing should be considered for identifying hidden nosocomial clusters of NoV infections within health care settings.

Noroviruses (NoV) belong to the family Caliciviridae and are the most common cause of acute viral gastroenteritis worldwide (11). Noroviruses have a positive-sense RNA genome with an average length of 7.5 kb (4, 8). Noroviruses are genetically highly variable and are classified into 5 genogroups (GI, GII, GIII, GIV, and GV), 3 of which are found in humans (12, 24).

NoV are usually transmitted from person to person but may also spread via contaminated surfaces, food, and water (17). NoV outbreaks are common, particularly affecting health care institutions such as nursing homes and hospitals, but their impact and modes of transmission have not been assessed systematically (1, 6, 18). Previously, we described a high frequency of nosocomial infections by comparing the times of diagnosis and dates of hospitalization of newly diagnosed patients (1). There is evidence for increased health expenditures and possible complications in high-risk patients for nosocomial norovirus infections, showing that studies are required to develop effective methods for reducing nosocomial infections (23). A study examining the efficacy of control strategies found that implementation within 3 days after the first cases was the only factor that significantly reduced the size and duration of NoV outbreaks in nursing homes, regardless of the infection control protocol that was followed (5). Furthermore, another study monitoring gastroenteritis outbreaks in England demonstrated the potential effectiveness of ward closure in hospitals (15). This shows that timely detection of nosocomial spread is a key determinant of successful control activities (7, 9, 21). We therefore investigated the possible use of molecular typing in addition to routine monitoring for nosocomial infections to detect transmission pathways of norovirus in a hospital environment. Sequencing of the norovirus P2 domain, which is located in the ORF 2 capsid gene, was used to link patients with identical strains into clusters (24, 25). This approach identified possible clusters that would be missed by standard epidemiological cluster analysis.

MATERIALS AND METHODS

Data collection and fecal specimens. Data on norovirus-positive cases diagnosed between 2002 and 2007 were retrieved from the database of the hospital laboratory and grouped as nosocomial cases, outpatient cases, and community-acquired cases (1). We used a conservative estimate to ensure high specificity by considering the possibility of nosocomial transmission only if a patient was diagnosed with NoV infection for the first time >4 days after admission. Patients who tested positive for NoV 0 to 1 day after admission were defined as having community-acquired cases. Patients with NoV-positive stools diagnosed 2 to 4 days after hospitalization were classified as indeterminate. On the basis of the >4-day cutoff, 22 nosocomial clusters had previously been obtained using epidemiological criteria (defined as ≥2 patients with nosocomial infection with NoV on the same ward within 5 days) (1). Background data listing the age of the patient, sex of the patient, ward where the patient was hospitalized, date of admission, and dates of hospitalization were used to identify nosocomial cases.

Outline. Stored fecal specimens (stored at –80°C) were retrieved, viral RNA was extracted, and strains were typed using a two-step approach. First, viruses were assigned to a genotype by sequencing region A of the polymerase gene (22). Subsequently, the corresponding P2 domains in the capsid gene were sequenced, using a specific P2 primer set for each genotype (24). This approach was necce-
sary because the genetic diversity of norovirus P2 regions is so high that a single set of primers has an inherently low sensitivity.

RNA extraction and RT. Fecal samples were suspended (200 mg/200 μl) in Hanks’ medium (800 μl) containing penicillin and were clarified for 30 min at 3,000 rpm and 4°C (8,000 relative centrifugal force [RCF]; Eppendorf 4515R centrifuge). Two hundred microliters of the supernatant was transferred to a Magna Pure LC plate for reverse transcription-PCR (RT-PCR) (total nucleic acid extraction program, performed according to the manufacturer’s instructions) with an elution volume of 50 μl (Roche Diagnostics GmbH). For genotyping, 20 μl of RNA extract was reverse transcribed to cDNA with random hexamers by use of a MultiScribe reverse transcriptase kit (Applied Biosystems) according to the manufacturer’s instructions. All of the obtained threshold cycle (CT) values in our study correspond with real-time PCR results, as previously mentioned (1).

Genotyping and P2 domain sequencing. A seminested PCR was performed on the polymerase region (region A) to type the NoV strains. First-round PCRs were performed with the primer set FW-JV12 (ATACCACTATGATGCAGATTAT) and COGREV (TCGACGCCATCTTCATTCACA) (10); 10 μl cDNA was added to a 40-μl reaction mix containing 5 μl PCR buffer, 1 μl (each) of forward and reverse primers (70 pmol), 2 μl MgCl2, 25.5 μl Adest, 1 μl deoxynucleoside triphosphates (dNTPs) (10 mM), and 0.5 μl Hotstart Taq polymerase (2 units). The second-round PCR (seminested) was performed with primer set JV12-FW and JV15-REV (CTCATCCAYCTRAACATNGNYTCYTGTG). Two microliters of the first PCR product was added to a 48-μl reaction mix containing 5 μl PCR buffer, 1 μl (each) of forward and reverse primers (70 pmol), 2 μl MgCl2, 37.5 μl Adest, 1 μl dNTPs (10 mM), and 0.5 μl Hotstart Taq polymerase (2 units). Both PCRs were performed using a Gene Amp 9700 thermal cycler (Applied Biosystems) under the following cycling conditions: 96°C for 15 min, 40 cycles of 96°C, 52°C, and 72°C for 1 min each, and 72°C for 10 min. PCR products were loaded on 2% agarose gels (stained with Syber Safe). When the target band was observed (approximately 650 bp), the PCR products were purified with ExoSAP-IT (USB Corporation, Cleveland, OH) (2 μl ExoSAP-IT for 5 μl PCR product), followed by sequencing with the same primers used for PCR, using the ABI Prism BigDye Terminator 3.1 approach (Applied Biosystems model 3730 DNA analyzer), with denaturation at 96°C for 10 s and 25 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. After assignment of the genotype, type-specific primers were used to sequence a 794- to 818-nucleotide (nt) target covering the P2 domain (24). This was done only for the two most common genotypes, GII.3 and GII.4, for which sufficient background data were available for sequence similarity comparisons. Primers and PCR conditions were as described previously (24).

Sequence analysis. The obtained sequences were entered and aligned in Bionumerics (software package 5.1; Applied Maths) and were typed with a genotyping tool for noroviruses (http://www.rivm.nl/mpf/norovirus/typingtool). The sequences were connected with the available background data file listing age, sampling date, date of discharge from the hospital, and location (ward) where the patient stayed while hospitalized (19). The P2 domain sequences, with an average length of 550 nt, were compared using the neighbor-joining method (TREECON for Windows) to identify patients who had identical sequences in order to create molecular clusters. Sets of identical sequences were defined as clusters (Fig. 1). The community-acquired cases not only served as background sequence data in the comparison but also were used to link with nosocomial cases in order to detect the introduction of strains into the hospital.
### RESULTS

In total, 264 patients (of 2,458 that were evaluated) tested positive for noroviruses during the 5-year period. Among these, 61% of the infecting strains (n = 160) could be genotyped. Viruses belonging to GII.3 (34%; n = 54) and GII.4 (51%; n = 82) were identified most commonly, followed by viruses of GII.7 (9%; n = 15), GII.2 (4%; n = 6), and G1.6B/II.17 (2%; n = 3). The samples that could not be genotyped were retested using different diagnostic PCRs. Mean Ct values did not differ between samples that could or could not be genotyped.

Overall, 48% (n = 128) of NoV-positive patients most likely had hospital-acquired infection, according to the cutoff. Patients with newly diagnosed cases (17%; n = 44) had an onset of illness within 2 to 4 days after admission, but the exact source of infection could not be established. Finally, 35% (n = 92) of patients tested NoV positive 0 to 1 day after admission and were classified as having community-acquired cases. In Table 1, the proportions of nosocomial cases for several groups are shown, based on age and genotype. These proportions varied from 0 (for GII.7 strains in adults) to 0.737 (for GII.3 strains in children younger than 5 years of age). As shown in Table 1, genotype GII.3 in children represented a large proportion of nosocomial infections, whereas for the group of patients of >5 years of age, this was the case for GII.4 strains (Table 2). Testing the relationship between the proportion of nosocomial transmission and genotypes, on one hand, and age, on the other hand, showed that the proportion of nosocomial transmission was significantly different in the older age group but not in children. Overall, nosocomial NoV was observed more commonly in young children. Viruses of genotype II.3 were found more often in young children.

### DISCUSSION

We describe the results of a systematic evaluation of patients diagnosed with NoV in a large hospital between 2002 and 2007 to look for evidence of nosocomial outbreaks through sequence-based clustering of cases. This approach was done as part of a study aimed at mapping the sources of virus introduction that may be amenable to intervention strategies, as NoV outbreaks in hospitals may have significant health impacts. The use of sequence analysis in this study identified 11
clusters that had not been recognized through earlier defined epidemiological clustering analysis (1), increasing the number of probable nosocomial clusters by 50%. Almost half of these involved links with a patient who had visited the hospital but was not admitted, suggesting the introduction of virus into wards through staff movement or contaminated surfaces. Since we used a rather conservative cutoff for the definition of nosocomial infection, we may have underestimated the prevalence.

We analyzed the virus diversity in relation to date of hospitalization. This provided the opportunity to compare the diversity of the viruses in nosocomial patients with that of viruses causing illness in the community. This comparison is essential, as widespread community outbreaks may occur, in which case finding an identical sequence in the hospital may not signify a hospital-acquired event. An example where this occurred is cluster 8 in Fig. 1A, showing 2 apparently connected patients with community-acquired illness. However, since all other community-acquired cases were distinct, this strengthened the support for the observed approach and thus the clusters that were identified.

Our findings clearly show the limitations of commonly used epidemiological clustering, where these clusters would not be noticed. In this study, patients were identified as possible linked cases when they were hospitalized within the same wards and within the same time frame (5 days).

A limitation of our study is the number of samples for which the genotype could not be determined. Since mean Ct values did not differ between stool samples with and without genotyping results, the quantity of virus in the original specimens is not an explanation. A reasonable explanation could be the different PCRs used for diagnosis and genotyping: the former uses a smaller amplicon size, and fragmentation of RNA during preparation and freeze-thawing could preferentially influence the genotyping PCR, with its longer target fragment. Alternatively, it is possible that the nontypeable samples contained different norovirus genotypes, but we could not find any evidence of this possibility.

In the current approach, we used stringent selection based on 100% similarity among strains defining a link (24, 25). This may be too stringent, as NoV is rapidly evolving and mutations are accumulated rapidly (20). Therefore, allowing one or even two nucleotide differences between sequences could potentially increase the sensitivity of outbreak detection. However, this remains to be proven, as few studies have addressed the evolution of NoV over different chains of transmission (3, 20).

Interestingly, we found that the proportion of nosocomial infections seemed to depend on the particular strain involved. In particular, the GII.3 strains showed a significantly larger proportion of nosocomial infection, regardless of age, than the other genotypes. This illustrates the complexity of NoV epidemiology, showing that NoV should not be viewed as “a” virus but rather as a group of related viruses with different properties. This comes as no surprise given the huge diversity of NoV: viruses belonging to GII.7 and GII.4 are quite distinct. Taking these differences into account, one could possibly speculate that each specific genotype could be associated with a particular disease burden. As shown within the family *Picornaviridae*, another family of positive-strand RNA viruses, genetically related viruses can cause quite distinct spectra of diseases (13). In addition to the virus genotype, age group needs to be factored in: GII.3 strains were found more often in children and in nosocomial cases, in contrast to GII.4 strains, for which this age difference was not found.

Our findings suggest differences in either susceptibility or severity of GII.3 infection for different age groups, as has been described for group A rotaviruses (16). A plausible explanation would be the development of herd immunity, given the widespread circulation of these viruses. For GII.4 strains, rapid evolution of viruses into new antigenic variants has been shown to be an explanation for repeated epidemics involving all age groups (14).

The age-related probability of transmission in a health care setting is something to be aware of. The generally higher rate of nosocomial infection in the young is easily explained by hygienic conditions: young children may wear diapers, and the handling thereof is associated with higher exposure to stools. Without proper hand-washing hygiene, this may constitute a greater risk of transmission. A second factor could be that viral loads are higher in young children, as has been observed for

---

**Table 3. Overview of molecular clustering versus epidemiological clustering**

<table>
<thead>
<tr>
<th>Molecular cluster</th>
<th>Presence of epidemiological cluster</th>
<th>Ward(s)</th>
<th>Presence within 5 days</th>
<th>No. of nosocomial infections</th>
<th>No. of indeterminate cases</th>
<th>No. of community-acquired cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
<td>Different</td>
<td>Yes</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>Different</td>
<td>No</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
<td>Different</td>
<td>No</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Same</td>
<td>Yes</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>Different</td>
<td>Yes</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>Different</td>
<td>Yes</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>Same</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>Different</td>
<td>Yes</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>Different</td>
<td>No</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>Different</td>
<td>No</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>No</td>
<td>Different</td>
<td>No</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>Yes*</td>
<td>Same</td>
<td>Yes</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>Yes*</td>
<td>Same</td>
<td>Yes</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>No</td>
<td>Same</td>
<td>No</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

* Cluster was identified by both methods, but the size (number of patients) differed.
other viruses against which immunity develops. Be that as it may, nonviral factors such as behavior (e.g., hygiene) seem to be important in contributing to transmission.

In conclusion, we show the usefulness of molecular information as a basis for detecting transmission events in the hospital setting. We show that the use of molecular typing may increase the early detection of clusters by 50%, and we were able to identify introductions from the outpatient department. This indicates that a careful review of movements of people between outpatient clinics and wards could potentially identify areas for improvement. The significantly increased proportions of nosocomial transmission of GII.4 and GII.3 strains compared with those of NoV belonging to other genotypes show that an early warning system that rapidly identifies an increasing prevalence of new variants of these genotypes could be used to guide enhanced infection control policies.

ACKNOWLEDGMENTS

We thank Georgina Aron for her assistance during sample collection and Erwin Duizer, Annelies Kroneman, and Linda Verhoef for their useful comments and feedback.

This study was financed by ZonMw, Netherlands.

REFERENCES