

Laboratory-based surveillance in the molecular era: the TYPENED model, a joint data-sharing platform for clinical and public health laboratories

H G Niesters¹, J W Rossen^{1,2}, H van der Avoort³, D Baas³, K Benschop^{3,4}, E C Claas⁵, A Kroneman³, N van Maarseveen⁶, S Pas⁷, W van Pelt³, J C Rahamat-Langendoen¹, R Schuurman⁶, H Vennema³, L Verhoef³, K Wolthers⁴, M Koopmans (Marion. Koopmans@rivm.nl)^{3,7}

1. Department of Medical Microbiology, Division of Clinical Virology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
2. Laboratory for Medical Microbiology, St Elisabeth Hospital, Tilburg, the Netherlands
3. Center for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
4. Department of Medical Microbiology, Academic Medical Center, Amsterdam, the Netherlands
5. Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands
6. Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, the Netherlands
7. Department of Virology, Erasmus Medical Centre, Rotterdam, the Netherlands

Citation style for this article:

Niesters HG, Rossen JW, van der Avoort H, Baas D, Benschop K, Claas EC, Kroneman A, van Maarseveen N, Pas S, van Pelt W, Rahamat-Langendoen JC, Schuurman R, Vennema H, Verhoef L, Wolthers K, Koopmans M. Laboratory-based surveillance in the molecular era: the TYPENED model, a joint data-sharing platform for clinical and public health laboratories. *Euro Surveill.* 2013;18(4):pii=20387. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20387>

Article submitted 28 June 2012 / published on 24 January 2013

Laboratory-based surveillance, one of the pillars of monitoring infectious disease trends, relies on data produced in clinical and/or public health laboratories. Currently, diagnostic laboratories worldwide submit strains or samples to a relatively small number of reference laboratories for characterisation and typing. However, with the introduction of molecular diagnostic methods and sequencing in most of the larger diagnostic and university hospital centres in high-income countries, the distinction between diagnostic and reference/public health laboratory functions has become less clear-cut. Given these developments, new ways of networking and data sharing are needed. Assuming that clinical and public health laboratories may be able to use the same data for their own purposes when sequence-based testing and typing are used, we explored ways to develop a collaborative approach and a jointly owned database (TYPENED) in the Netherlands. The rationale was that sequence data – whether produced to support clinical care or for surveillance – can be aggregated to meet both needs. Here we describe the development of the TYPENED approach and supporting infrastructure, and the implementation of a pilot laboratory network sharing enterovirus sequences and metadata.

Introduction

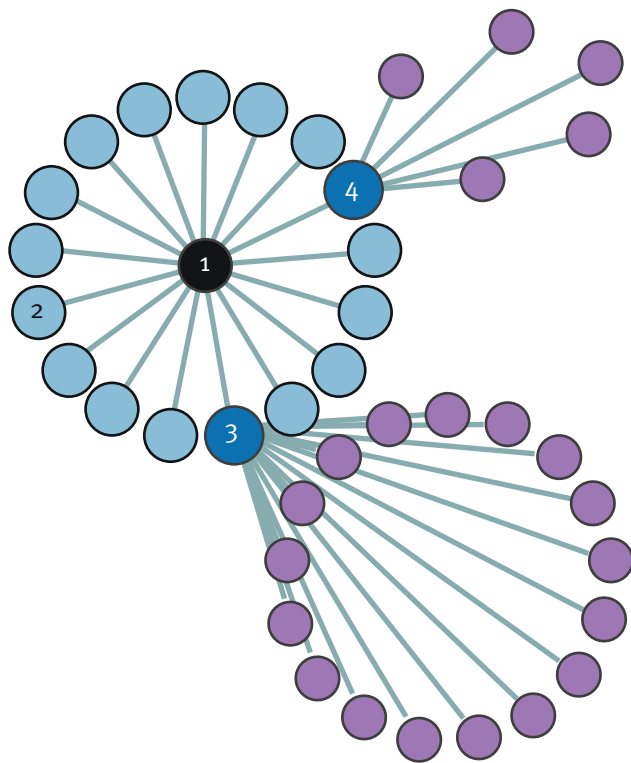
Laboratory-based surveillance is one of the pillars of monitoring infectious disease trends, which is based on data from clinical and/or public health laboratories. This type of surveillance is performed for a range of food- and waterborne, sexually transmitted and blood-borne diseases, respiratory pathogens or zoonotic pathogens and provides important input for national

and international disease surveillance, to evaluate the impact of control and prevention measures, and to detect clusters or relevant changes in pathogen presence and/or behaviour [1-3].

One problem in the use of laboratory-based surveillance systems is that they require information that typically is collected at the clinical level and therefore is not focused on surveillance. For certain priority diseases, such as polio and measles, this issue has been solved by making the identification of a case notifiable, in which case the laboratory or the clinician or both are required to provide structured information for surveillance to a national or international dedicated organisation. For non-notifiable diseases, however, the need for standardisation to ensure data comparability between laboratories may be at odds with the rapid developments in clinical microbiology laboratories [4-6]. In the Netherlands, currently, diagnostic laboratories routinely submit strains or samples to reference laboratories for characterisation and typing. However, with the introduction of molecular diagnostic methods in most of the larger diagnostic centres, the distinction between diagnostic and reference laboratory functions has become less clear-cut. Multiplex real-time PCR and sequence-based detection and typing techniques may be used for clinical diagnosis, to guide treatment (by, for example, resistance profiling, strain characterisation and typing), for hospital infection control and quality management (for cluster detection). The methods and analytical tools employed for these functions potentially overlap with what is needed for national and international or cross-border surveillance. The expected introduction of next generation

FIGURE 1

Conceptual model for TYPENED, showing laboratories with different capacities



- National reference laboratory: national focal point
- Expert clinical laboratory: reference role
- Clinical laboratory: diagnostic and typing service
- Clinical laboratory: diagnostic service

TYPENED: TYPEer netwerk NEDerland [Typing network Netherlands].

1, 2, 3 and 4 represent a specialist laboratory, dealing with, for example, samples from food, water, the environment and animals.

The laboratory capacities range from routine diagnostic functions, diagnostics and typing functions, expert-level services (includes research), and national reference-level functions.

The dark circle indicates the hub from which the molecular platform infrastructure is provided (see Figure 2). Based on areas of expertise (indicated by numbers), coordination of the network activities may be delegated from the national focal point to a local laboratory, while maintaining the common infrastructure.

sequencing techniques in routine diagnostic settings within the next five years is likely to further lift the borders between the previously separated activities across disciplines and domains [7].

While international surveillance networks rely on reference laboratories, and each pathogen or pathogen group has its own network and system, often with centralised data collection, the latest developments are a challenge for these networks. As more and more clinical laboratories perform molecular testing methods, the reference laboratories become dependent on data submission by these laboratories, often with little perceived benefit for the submitting laboratories, considering the extra effort required. We anticipate increasing resistance from clinical laboratories to data requests for surveillance purposes because of these competing priorities.

Given these developments, we consider that new ways of networking of data and data sharing are needed. Assuming that clinical and public health laboratories may be able to use the same data for their own purposes when sequence-based testing and typing are used, we explored ways to develop a collaborative approach and a jointly owned database in the Netherlands. Here we describe the development of the approach and supporting infrastructure, and the implementation of a pilot laboratory network sharing enterovirus sequences and metadata.

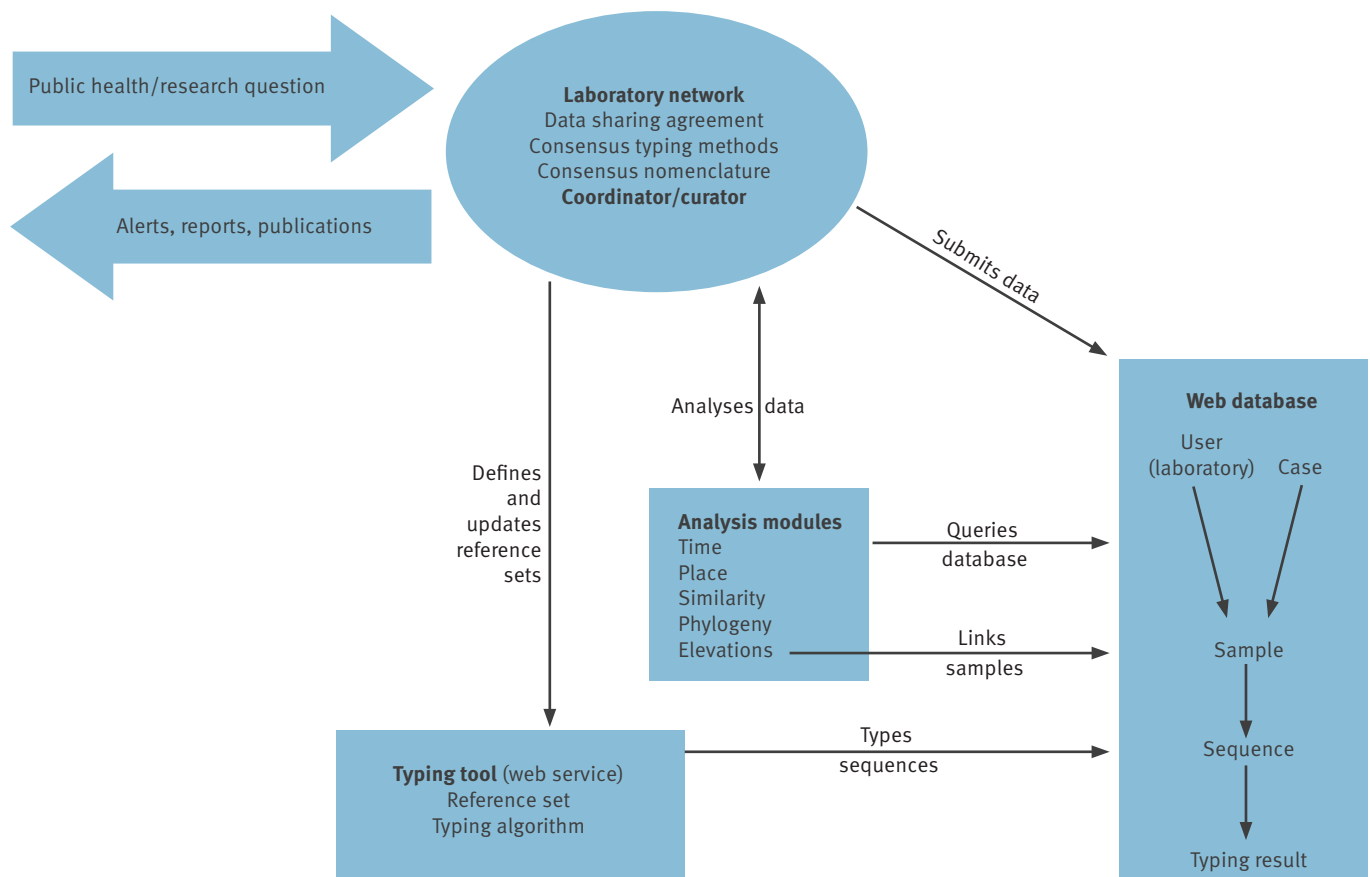
Methods

Partnership

An initiative set up by a group of opinion leaders in microbiology in the Netherlands to draw attention to the changing needs of and demands placed on clinical laboratories and the need for standardisation to ensure data comparability and sharing between laboratories. Within this initiative, called TYPENED (TYPEer netwerk NEDerland [Typing network Netherlands]), two pilots were started in 2009: one for bacterial typing and one for viruses. In the VIRO-TYPENED pilot, five universities and one regional laboratory collaborated with the National Institute of Public Health and the Environment (RIVM) to develop a new model of collaboration for virology based on sequence information gathered in the routine diagnostic setting. These laboratories have all been long-term suppliers of surveillance information, by sending to RIVM isolates or clinical specimens as well as clinical information for a number of viruses such as influenza A virus, norovirus, enterovirus, rotavirus and hepatitis A, B and E viruses. All participating laboratories have molecular diagnostic testing facilities and perform sequencing as part of their routine diagnostics for specific clinical or research questions on one or more of these pathogens (Figure 1, first ring of clinical laboratories surrounding the national reference laboratory). Using a centralised database structure at the national reference laboratory level, expert clinical laboratories can still have

FIGURE 2

Conceptual model for data sharing platform for TYPENED collaboration between the national public health institute and clinical laboratories in a laboratory network



TYPENED: TYPeer netwerk NEDerland [Typing network Netherlands].

Laboratories submit data to a joint database. The data comprise sequence data and background data on the sample and the patient (case). All sequences are typed automatically. A set of online analysis modules is available for all participants to mine the data. Data can be analysed for trends or clusters in time and place. Sequence data can be analysed for similarity and phylogenetic clustering. Elevations are identified through an automatic cluster detection algorithm based on both sequence information and epidemiological parameters.

their own network activities with collaborating local diagnostic laboratories (Figure 1, showing the group of diagnostic laboratories that refer molecular typing data to the laboratories indicated by '3' or '4').

Selection of pilot pathogens

An inventory was made of the currently used typing methods in the six clinical laboratories and the public health laboratory participating in VIRO-TYPENED using a structured online questionnaire. Participants were asked to list those viruses for which they had typing methods operational in their laboratories and the purpose of those typing applications, and to indicate for which viruses they would like to see joint action, and at which level. The options provided were: (i) the exchange of protocols, control reagents and

quality-control panels; (ii) a centralised reference data collection; (iii) a common database; and (iv) no collaboration considered necessary. The purpose of this inventory was to identify areas for which there was a common need, as well as areas where joint action was not considered advantageous. A second part of the inventory asked about methods used and the frequency of typing in each laboratory.

Molecular platform database

In order to achieve efficiency and continuity, a generic database infrastructure for sharing of molecular typing data and metadata was developed at RIVM between 2008 and 2011. The platform consists of a web database and a set of analysis modules. The database can be configured for a specific pathogen, at the request

of a laboratory network, which also appoints a coordinator or curator. User types can be defined, coupled with tailored access rights. The two central entities are sample and sequence. A minimal dataset can be defined by the network, based on the questions addressed, coupled with a feasibility assessment. This dataset minimally comprises time and place, but can be complemented with additional epidemiological or clinical metadata specific to the targeted organism. Besides online data entry forms, the platform provides a bulk upload option using Microsoft Excel and FASTA formats.

All sequences submitted to the database are automatically typed in a standardised way using a web-based typing tool [8]. Sequence data can be analysed by carrying out built-in similarity searches using the BLAST algorithm, and by generating pie charts, incidence plots, geographical maps and phylogenetic trees (neighbour-joining clustering method, with a two-parameter Kimura nucleotide-substitution model, with or without bootstrapping). The added value of a database like this – compared with the database of GenBank [9], in which laboratories all over the world share their sequences – is threefold. Firstly, the data are more comparable because of the agreed typing region and the standardised typing results and secondly, the data are shared before laboratories have decided to make them publicly available, for example, through GenBank. The third important advantage is the linked, standardised set of epidemiological and clinical data with each sequence, which allows in-depth analysis. A description of the components and functions of the molecular platform is shown in Figure 2.

Pilot study: enteroviruses

On the basis of the inventory results, the seven laboratories agreed to start the pilot with enterovirus as a test pathogen. A minimum dataset was agreed, including age and sex of patient, type of sample from which the virus was detected, whether the patient was hospitalised, travel history (by country visited), clinical symptoms in broad categories (skin, neurological, respiratory, enteric). For each patient, at least one sequence of the major capsid protein VP1 gene has to be provided of the agreed genomic region (nucleotides 2,604–2,909 NC_001612, CVA16). In addition, samples that could not be typed as an enterovirus but were typed as poliovirus-like, were sent to the enterovirus section of the Center for Infectious Disease Control at RIVM, as part of the enterovirus surveillance programme in place, to document the absence of wild-type poliovirus circulation.

Data sharing and confidentiality agreement

Participants worked with a confidentiality agreement, consenting to the use of the data to provide surveillance overviews and alerts and to the right to publish the data, with proper acknowledgement, in case of public health emergencies. All participants can access

and download the data, but they cannot be used without the consent of the data provider.

Enterovirus diagnostics and sequencing

Each laboratory used a laboratory-developed test, adapted from the protocol described by Nix et al. [10] (2006) for the detection of enteroviruses. One laboratory used an additional protocol described by McWilliam Leitch et al. [11] for cerebrospinal fluid samples. All laboratories participated in an external proficiency testing programme organised through Quality Control for Molecular Diagnostics (QCMD), Glasgow, United Kingdom, an International Organization of Standardization (ISO) 17043-accredited organisation. Amplification of the 5' non-coding region of enterovirus was performed at the individual participating laboratory.

Genotype assignment using a standardised sequence-based typing tool

Upon entering of sequences into the database, an automated algorithm was run to assign the genotype. This tool has been validated against most currently known picornaviruses and has been shown to correlate highly with the serotype assignment [8].

Results

Questionnaire information

In addition to enterovirus, the seven participating laboratories indicated that they performed systematic genotyping for influenza virus (n=7), hepatitis B virus (n=6) and hepatitis C virus (n=5), primarily related to monitoring of treatment. Some laboratories also typed parechoviruses (n=5), rhinoviruses (n=3), hepatitis E virus (n=3), norovirus (n=2), hepatitis A virus (n=2), cytomegalovirus (n=2), herpes simplex virus (HSV) (n=2), adenovirus (n=2), human immunodeficiency virus (HIV) (n=3), as well as hepatitis B virus and hepatitis C virus for specific research or clinical study-related questions. A need for a more structured collaboration between the laboratories, possibly including the operation of a joint reference database, was indicated by the majority of respondents regarding influenza virus, parechovirus, rhinovirus and hepatitis B virus. For the less commonly used typing approaches, a need for collaboration was expressed for hepatitis viruses A, C and E. Given the consensus that a type of collaborative network would meet a need, a pilot TYPENED database was set up for enteroviruses.

Pilot enterovirus database

As of 1 May 2012, a total of 651 human enterovirus (HEV) sequences were submitted to the TYPENED database, representing all enterovirus-positive clinical samples that were successfully sequenced at six of the collaborating laboratories from 1 January 2010 to 31 December 2011. Most of the sequences belonged to HEV-A (n=168; 25.8%) and B (n=466; 71.6%), whereas only a few belonged to HEV-C (n=6; 0.9%) and D (n=6; 0.9%). Following automatic typing of the sequences submitted

to the TYPENED database, it appeared that some of the viruses that were enterovirus positive in the molecular diagnostic assay appeared to be a rhinovirus A (n=5; 0.8%), most probably due to the cross-reactivity of the primers used for detection. In addition, three poliovirus sequences were identified within the HEV-C set: all three isolates were obtained from children from the former Netherlands Antilles (Curaçao and Sint Maarten), where oral polio vaccines were used.

The laboratories that submitted the sequences received samples from laboratories all over the Netherlands. Although the numbers per serotype were not always very large, some clusters of serotypes over time could be observed (detailed data not shown). For example, of the 48 CV-A9 sequences submitted, 43 were found in samples collected from May to August 2010 with a clear peak (n=38) in June and July. In addition, five of the six EV-D68 sequences were found in samples collected from August to November 2010; 46 of the 65 E-7 sequences were found in samples collected from May to August 2011 and 51 of the 69 E-25 sequences were found in samples collected from August to December 2011.

Discussion

We have described a data-sharing concept that combines the capacities of clinical and public health laboratories in the Netherlands in a database to which all laboratories have equal and full access. After initial discussions to align expectations and develop a code of conduct, all laboratories were able to share a first set of historical data within two months. One of the triggers for the development of this concept was the concern that current enterovirus surveillance which is based on cell culture isolation is no longer the preferred method for enterovirus detection at hospital level and information obtained through other typing methods would not be captured centrally [12].

We managed to get consensus on the typing protocol and a data sharing agreement between the central public health laboratory (RIVM), large university laboratories and some large general hospitals that are geographically dispersed, thus potentially enabling broad coverage of surveillance of viruses of common interest. Within the enterovirus pilot, all sequences generated in two years by six of the seven collaborating laboratories were shared.

One pitfall of a consensus typing method may be that some viruses will be missed if they are not detected in the particular molecular test. This is of concern, given that the previously common practice of viral culture, which could serve as a safety net, is diminishing very rapidly. Most laboratories maintain these culture facilities only to grow control material for molecular assays. Since RNA viruses diverge rapidly, there is a need to get updated full-length sequences, not only for epidemiological reasons but also to keep diagnostic assays based on molecular testing up to date. At present, the

availability of whole genome sequences is limited, but with next generation sequencing techniques rapidly coming within reach of academic and even clinical laboratories, this situation will change quickly.

The same system is currently being set up for a number of other viruses for which collaboration was valued according to the questionnaire – with parechovirus, norovirus and hepatitis E virus on the priority list [13-15]. Sequence-based characterisation is becoming more common within the larger diagnostic centres: the availability of sequence-based information will assist both the clinicians and diagnostic laboratories as well as the public health laboratories.

The concept of TYPENED in the Netherlands has been shown to be an effective means of close collaboration and the participating laboratories are willing to extend this collaboration to other targets. Furthermore, by using sequencing technologies, a more in-depth analysis of circulating strains can be carried out, as individual sequences can be analysed, instead of serotypes. Sequences have a much higher discriminatory power, as most sequences within one serotype will be different from each other, thus facilitating, for example, the tracing of transmission patterns. Sequence techniques are particularly valuable for viruses that are difficult to grow. In an economic climate with shrinking budgets, it may prove difficult for facilities to perform sequencing for diagnostic and epidemiological purposes, although it is expected that large centres will continue to perform routine sequencing. The TYPENED model seeks to maximise the use of data generated both in clinical and public health laboratories, for clinical care and for surveillance purposes. The harmonisation of typing protocols and sharing of data with a more extensive group of laboratories, or even cross-border centres, will be a next step.

References

1. Bean NH, Martin SM. Implementing a network for electronic surveillance reporting from public health reference laboratories: an international perspective. *Emerg Infect Dis.* 2001;7(5):773-9.
2. Emori TG, Gaynes RP. An overview of nosocomial infections, including the role of the microbiology laboratory. *Clin Microbiol Rev.* 1993;6(4):428-42.
3. Workman MR, Wall PG, Tearle P, O'Mahony M, Brunton WA. Active surveillance of health and safety in microbiology laboratories. *Commun Dis Rep CDR Rev.* 1995;5(4):R54-6.
4. Deyde VM, Sampath R, Gubareva LV. RT-PCR/electrospray ionization mass spectrometry approach in detection and characterization of influenza viruses. *Expert Rev Mol Diagn.* 2011;11(1):41-52.
5. Endimiani A, Hujer KM, Hujer AM, Kurz S, Jacobs MR, Perlin DS, et al. Are we ready for novel detection methods to treat respiratory pathogens in hospital-acquired pneumonia? *Clin Infect Dis.* 2011;52 Suppl 4:S373-83.
6. Fournier-Wirth C, Coste J. Nanotechnologies for pathogen detection: Future alternatives? *Biologicals.* 2010;38(1):9-13. Epub 2010 Jan 15.
7. Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, et al. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerg Infect Dis.* 2012;18(11):e1.
8. Kroneman A, Vennema H, Deforche K, v d Avoort H, Peñaranda S, Oberste MS, et al. An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol.* 2011(2):121-5.
9. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Res.* 2013;41(D1):D36-42.
10. Nix WA, Oberste MS, Pallansch MA. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. *J Clin Microbiol.* 2006;44(8):2698-704.
11. McWilliam Leitch EC, Harvala H, Robertson I, Ubbilos I, Templeton K, Simmonds P, et al. (2009). Direct identification of human enterovirus serotypes in cerebrospinal fluid by amplification and sequencing of the VP1 region. *J Clin Virol.* 44(2):119-24.
12. Manzara S, Muscillo M, La Rosa G, Marianelli C, Cattani P, Fadda G. Molecular identification and typing of enteroviruses isolated from clinical specimens. *J Clin Microbiol.* 2002;40(12):4554-60.
13. Wolthers KC, Benschop KS, Schinkel J, Molenkamp R, Bergevoet RM, Spijkerman IJ, et al. Human parechoviruses as an important viral cause of sepsislike illness and meningitis in young children. *Clin Infect Dis.* 2008;47(3):358-63.
14. Nelson KE, Kmush B, Labrique AB. The epidemiology of hepatitis E virus infections in developed countries and among immunocompromised patients. *Expert Rev Anti Infect Ther.* 2011;9(12):1133-48.
15. Verhoef L, Williams KP, Kroneman A, Sobral B, van Pelt W, Koopmans M; et al. Selection of a phylogenetically informative region of the norovirus genome for outbreak linkage. *Virus Genes.* 2012;44(1):8-18.