Molecular Surveillance of Shiga Toxigenic *Escherichia coli* O157 by PulseNet USA

PETER GERNER-SMIDT,1, 4 JENNIFER KINCAID,1 KRISTY KUBOTA,1 KELLEY HISE,1 SUSAN B. HUNTER,1 MARY-ANN FAIR,1 DAWN NORTON,1 ANN WOO-MING,2 TERRY KURZYNKSKI,3 MARK J. SOTIR,1, 4 MARCUS HEAD,3 KRISTIN HOLT,3 AND BALA SWAMINATHAN5

1 Centers for Disease Control and Prevention, National Center of Infectious Diseases, Division of Bacterial and Mycotic Diseases, Foodborne and Diarrheal Diseases Branch, Mail Stop C03, 1600 Clifton Road N.E., Atlanta, Georgia 30333; 2 Colorado Department of Health and Environment, Laboratory Services Division, 8100 Lowery Boulevard, Denver, Colorado 80220; 3 Wisconsin State Laboratory of Hygiene, 465 Henry Mall, Madison, Wisconsin 53706; 4 Wisconsin Division of Public Health, P.O. Box 2659, Madison, Wisconsin 53701; and 5 U.S. Department of Agriculture, Food Safety and Inspection Service, 950 College Road, Athens, Georgia 30605, USA

MS 04-539: Received 30 November 2004/Accepted 1 June 2005

ABSTRACT

PulseNet USA is the national molecular subtyping network system for foodborne disease surveillance. Sixty-four public health and food regulatory laboratories participate in PulseNet USA and routinely perform pulsed-field gel electrophoresis of Shiga toxigenic *Escherichia coli* isolated from humans, food, water, and the environment on a real-time basis. Clusters of infection are detected in three ways within this system: through rapidly alerting the participants in the electronic communication forum, the PulseNet Web conference; through cluster analysis by the database administrators at the coordinating center at the Centers for Disease Control and Prevention of the patterns uploaded to the central server by the participants; and by matching profiles of strains from nonhuman sources with recent human uploads to the national server. The strengths, limitations, and scope for future improvements of PulseNet are discussed with examples from 2002. In that year, notices of 30 clusters of Shiga toxigenic *E. coli* O157 infections were posted on the Web conference, 26 of which represented local outbreaks, whereas four were multistate outbreaks. Another 27 clusters were detected by central cluster detection performed at the Centers for Disease Control and Prevention, of which five represented common source outbreaks confirmed after finding an isolate with the outbreak pattern in the implicated food. Ten food isolates submitted without suspicion of an association to human disease matched human isolates in the database, and an epidemiologic link to human cases was established for six of them.

In the United States, an estimated 76 million cases of foodborne illness occur each year. Of these, approximately 110,000 infections are caused by Shiga toxigenic *Escherichia coli* (STEC), resulting in 90 deaths (9). STEC usually causes self-limiting diarrhea, which often progresses to hemorrhagic colitis. In up to 10% of cases, the infection can elicit acute renal failure and blood clotting disturbances, hemolytic uremic syndrome, or thrombotic thrombocytopenic purpura. These complications have a high case fatality rate and occur more frequently among children and the elderly. Sporadic cases, as well as outbreaks of disease caused by STEC, commonly occur throughout the world (8). From the public health standpoint, the most important serotype of STEC is *E. coli* O157:[H7], which has caused numerous outbreaks in the United States and elsewhere since the 1980s (5, 17). Ruminants are the predominant reservoir for *E. coli* O157:H7; thus, beef, especially ground meat (hamburger), has been associated with human illness. However, fruits, vegetables, unpasteurized apple juice or cider, and water contaminated with manure are also sources of infection. Finally, because the infectious dose for *E. coli* O157 is low (12, 15, 16), opportunities exist for transmission of the infection from person to person or after contact with recreational waters or animals (6, 8).

PulseNet USA is the national molecular surveillance network system for bacterial foodborne pathogens in the United States. In this system, clusters of human infections caused by bacterial pathogens with the same molecular subtype are detected, and the vehicle of an outbreak can be confirmed. The results generated in PulseNet cannot alone determine whether an outbreak is occurring; the results need to be confirmed by epidemiological data. Currently, PulseNet uses pulsed-field gel electrophoresis (PFGE) of restricted DNA as the subtyping method. The network was established in response to increasing demand for rapid high-discriminatory typing to accelerate the detection and investigation of outbreaks caused by STEC O157:H7, *Salmonella*, *Listeria monocytogenes*, and *Shigella*. Since its implementation in 1996, PulseNet has become an invaluable tool for detection, investigation, and subsequent control of outbreaks of foodborne infections in this country (13). Full national participation of all 50 state public health departments was achieved in 2001. In this article, we present the results of the investigations related to the STEC PulseNet
database in 2002 and discuss the strengths of the system, its limitations, and the scope for further improvements.

DESCRIPTION OF PULSENET USA

Participants in PulseNet USA include all 50 state and five city and county public health laboratories and nine food safety regulatory laboratories. Eight public health laboratories are designated PulseNet Area Laboratories and supply service and support to the participants in their area. The Foodborne and Diarrheal Diseases Branch at the Centers for Disease Control and Prevention (CDC) in Atlanta is the coordinating center of the network. The backbone of the PulseNet network is an integrated Internet-based image analysis and database software system. The initial software package, Molecular Analyst Fingerprinting Plus (Bio-Rad Laboratories, Hercules, Calif.) (13), has been replaced with BioNumerics client-server software (Applied Maths BVBA, Sint-Martens-Latem, Belgium). A client version of the software is installed in all of the participating laboratories, enabling all participants to analyze their own gels. The server version of the software is installed at CDC. Participants who have passed a certification procedure upload patterns generated in their own laboratory to the national server and compare their patterns with all patterns in the national database. Participants who have not yet been certified send TIFF images of PFGE gels produced in their laboratory to CDC, where they are analyzed and uploaded to the national server by the PulseNet staff. State public health laboratories that are unable to perform PFGE temporarily can submit isolates to their PulseNet area laboratory or to the CDC PulseNet laboratory for typing. In the PFGE procedure, all isolates of STEC are typed with the restriction endonuclease XbaI. If a cluster with the same profile is detected, the PFGE is repeated with the implicated isolates and a second enzyme, BlnI (AvrII), and, if necessary, a third enzyme, SpeI.

PulseNet handles laboratory information in three ways: (i) posts to the PulseNet Web conference; (ii) cluster analyses of patterns submitted to the national database; and (iii) comparisons of patterns of food and veterinary isolates with human isolates in the national database.

PulseNet Web conference. The PulseNet Web conference is a closed, Web-based rapid alert system for participating laboratories and epidemiologists that is based on the WebBoard software (Akiva Corp., Carlsbad, Calif.). Information posted to the Web conference is confidential and cannot be forwarded outside the network to a third party without the permission of the participant supplying the data. Participants post cluster or outbreak information, which might extend outside their state or be of general interest to participants. An outbreak posting contains summarized information about available microbiological characteristics of the outbreak strain, geography and demography of infected patients, collection date, the time frames of the specimens, identity of the person posting the message, and information on who to contact in case the other participants have information relating to the cluster. A posting always includes an image of the PFGE profile of the suspected outbreak pattern with at least the primary PulseNet restriction enzyme (XbaI). If the outbreak strain has been characterized with additional enzymes, these patterns are also posted. The decision on what makes a cluster worthy of posting and when to post information about it on the Web conference lies with the participants. If the posting laboratory has been certified and thereby has been granted access to the national database, they can also provide preliminary data comparing the outbreak profile with patterns in the national database. The CDC database team will confirm this information or supply it in cases in which it has not been posted to the Web conference. Other PulseNet participants can respond to the Web conference postings, indicating whether or not they have any recent matches to the outbreak pattern. The participants who have a match notify the epidemiologists at their site to initiate a local outbreak investigation.

Central cluster detection. The CDC database team continuously evaluates and compares patterns submitted to the national database to detect clusters that have not been posted on the Web conference as early as possible. At least once a week, the team compares all patterns submitted to the central database within the previous 60 days. When the team detects a new local cluster, they contact the submitting laboratory for more information. If the cluster represents a local outbreak that has not been investigated, the team encourages the local laboratory to initiate an outbreak investigation together with the local epidemiologists. When the CDC database team detects new clusters of patterns submitted from more than one state, they are discussed with the CDC epidemiologists. The state public health laboratories, whose isolates formed the cluster, are contacted, and if there is no clear explanation for the clustering, an outbreak investigation is initiated. The CDC database team posts a message on the Web conference as an initial step. Currently, the CDC database manager determines whether a cluster has been identified, taking into account the overall frequency of the clustered pattern in the database and the time span of the collection dates or, if these are not present, the upload dates of the patterns to the central database. No automated cluster detection algorithm is used.

Comparison with strains from nonhuman sources. The CDC database team routinely compares patterns of STEC isolates from food sources submitted by the U.S. Department of Agriculture, Food Safety Inspection Service, and the U.S. Food and Drug Administration with all other patterns in the national database. If a food product isolate pattern is indistinguishable from that of a contemporaneous human isolate, an epidemiologic investigation is initiated.

RESULTS OF THE SURVEILLANCE OF STEC O157 IN 2002

In 2002, a total of 37 laboratories were certified for uploading images of STEC O157 and connecting to the PulseNet U.S. national server. Thirty-three of these were state public health laboratories, one was a combined county and city laboratory, and three were federal food regulatory laboratories. Another 17 public health laboratories submit-
FIGURE 1. Normalized image of the XbaI-restricted PFGE patterns associated with the outbreak in Colorado. Pattern EXHX01.1264 is the outbreak pattern; pattern EXHX01.0047 is the related but common pattern. The arrow points to the band differing between the two patterns. The BlnI-restricted patterns of the same strains were identical.

Two hundred seven patterns of nonhuman isolates were submitted to the PulseNet STEC database in 2002. Eighty-five were submitted as part of an outbreak investigation or to confirm an association between an ill person and a food item. One hundred twenty-two isolates were submitted from food without suspicion of association to human illness. Among these, 10 possible matches were found between a human and a food isolate; in six instances, an epidemiological connection between the human and food isolates was established.

Two outbreaks illustrate the usefulness of PulseNet and practical problems encountered with PFGE-based molecular surveillance of STEC O157.

Outbreak 1. The first example was a multistate outbreak involving 15 states. In July 2002, the Colorado Department of Public Health and Environment laboratory alerted PulseNet participants on the Web conference about a cluster of illness caused by a strain of STEC O157:H7 with a PFGE profile that had never been seen before in the national database. The Colorado Department of Public Health and Environment interviewed 16 culture-confirmed case patients and determined that ground beef was consumed by all. Within 35 days, 24 response postings were on the Web conference, with more than half posted within 4 days. Notably, 47 PFGE-matched cases were identified in 10 states, and the CDC joined the epidemiologic investigation. Although the PFGE profile of the outbreak strain was rare, it only differed from a common two-enzyme combination profile in the database by a single band (Fig. 1). However, for the interviews in this outbreak, only patients infected with isolates with PFGE patterns matching the first posted pattern by both enzymes XbaI and BlnI (PulseNet pattern EXHX01.1264/EXHA26.0015) were included initially as cases. The other closely related common pattern (EXHX01.0047/EXHA26.0015) had been seen 253 times before the outbreak and was uploaded 41 times during the outbreak in July and August 2002. The outbreak consisted of 44 culture-confirmed cases with an isolate yielding the outbreak profile. Additionally, two culture-confirmed patients were considered part of the outbreak on the basis of epidemiological evidence. They were siblings of patients infected by the outbreak strain, had the same food history, and became ill at the same time as their relatives, but the isolate cultured from their stools displayed the aforementioned variant PFGE profile. Two more cases were not culture confirmed but were included in the case count on the basis of their clinical history and strong epidemiologic links.

Results of patient interviews, coupled with the isolation of the outbreak strain from a patient’s food sample and meat traceback investigations, pointed to one particular meat producer. The link to the meat producer was confirmed when the outbreak strain was found in a meat sample collected by the Food Safety Inspection Service during an investigation of positive meat samples before and independent of reports of human illness. The whole production of a single day in May had been recalled on this basis. When the PulseNet investigations had linked the meat strain to the outbreak, further investigations at the production premises were undertaken. These indicated that the entire production for a 3-month period might have been contaminated, and the recall was extended accordingly to a total of 18.6 million pounds of meat products.

Outbreak 2. The second outbreak was detected and posted on the Web conference by the Wisconsin State Laboratory of Hygiene in September. During the following month, isolates with the outbreak pattern were found in 64 patients from 13 states. There were 23 responses to the Web conference posting within 1 month, 14 of them within 2 weeks. Isolates of STEC O157:H7 with the outbreak pattern had been seen only once before in the PulseNet database. A local case control study identified an association of illness, although not statistically significant, with the purchase and consumption of ground beef during the 7 days before illness. Isolation of STEC O157:H7 from an unused portion of factory ground and packaged ground beef obtained from a Minnesota patient along with traceback of ground beef purchases of several Wisconsin case patients pointed to a single meat production facility in Wisconsin. Several states reported the isolation of the outbreak strain from opened...
packages of ground beef obtained from case patients. Traceback of one of the positive packages with a specific purchase date indicated that ground beef produced on 3 days at the plant of interest was the likely source. On the basis of the findings of the case control study, traceback investigations, and product testing, the processor voluntarily recalled 416,000 pounds of nationally distributed ground beef products. Further investigations by the U.S. Department of Agriculture, Food Safety Inspection Service, at this production plant led to the recall of an additional 2.3 million pounds of meat within a few days.

**DISCUSSION**

**Real-time aspect.** The PulseNet Web conference functioned as a very effective tool for rapidly alerting participants about disease clusters and potential outbreaks and for rapidly assessing the magnitude and scope of clusters and outbreaks in 2002. The participants routinely used the Web conference for these purposes. Responses to Web conference postings were usually prompt. Both positive and negative responses were equally useful for the assessments. However, for the Web conference to function optimally, participants must type their isolates as soon as they are received, analyze the results, and disseminate the results to the network. This did not always happen. Several public health laboratories tended to do the typing of clinical STEC O157 isolates in batches (i.e., they waited until they had a sufficient number of isolates to fill a gel so as to conserve consumable supplies and use human resources efficiently). This delay resulted in loss of the real-time capacity. Additionally, participants did not always post results promptly.

In outbreak 1, clinical isolates were submitted to the Colorado public health laboratory an average of 3.5 days (range, 1 to 7 days) after collection. Isolates required an average 6.7 days (range, 4 to 13 days) to assess purity of the culture and for confirmatory biochemical testing, serotyping, Shiga toxin testing, PFGE typing, and submission of the results to the national database. Several factors contributed to the long local investigation time: (i) in Colorado, 57% of isolates submitted as probable O157 isolates were not confirmed, necessitating identification followed by serogrouping and -typing rather than real-time PFGE testing when the isolate was received; (ii) PFGE typing was conducted for two restriction enzymes when a cluster was detected, requiring multiple separate runs; (iii) for certain patients, isolate repeat runs were required because a second cluster with a one-band difference was being analyzed at the same time; and (iv) the Colorado laboratory staff did not have direct access to the CDC online database and had to e-mail the patterns to CDC for analysis. A temporary shortage in CDC PulseNet database management staff created a backlog of uploaded patterns from gel images submitted by e-mail. An increase in the number of laboratories, which are certified for uploading E. coli patterns has increased to 46.

**PFGE differentiation criterion.** Outbreak 1 also illustrates a problem when subtyping results are used exclusively to sort outbreak-related infections from unrelated infections. A strain can undergo changes during an outbreak, causing minor differences in the number or position of the DNA fragments in a PFGE gel (11). However, if one or more outbreak pattern variants happen to match frequently encountered PFGE patterns in the database, this occurrence would complicate the investigation and the identification of the point source of infection. According to the criteria proposed by Tenover et al. (14), strains are closely related if their PFGE profiles differ from each other in up to six band positions. For practical purposes when working with clonal foodborne organisms like STEC O157, this criterion does not work. In outbreak 1, in accordance with established PulseNet protocol, only patients with isolates displaying patterns indistinguishable from the original outbreak pattern were included in the initial case interviews. Thirty-seven isolates of the outbreak pattern were uploaded to the national server in July and August. During the same time period, 41 isolates of a related but frequently encountered pattern were also uploaded. Many of the patients infected with this frequently encountered strain were not likely to have been part of the outbreak and might have acquired their infection from other sources. Inclusion of these patients as part of this specific outbreak could significantly have weakened the epidemiologic association of the illnesses with the implicated ground beef. At the same time, epidemiologic evidence supported the inclusion of two siblings as part of the outbreak, even though their isolates had the frequently encountered pattern. This example illustrates the complexities involved in interpretation of outbreak investigation data and underscores the importance of considering all the available bacteriological and epidemiological evidence before making decisions.

**Microbiological confirmation of the source of an outbreak.** In outbreak 2, the finding of the outbreak strain in opened packages of ground beef from the same producer was an important argument for the recall. It is unusual to consider the finding of an outbreak strain in an open package of a suspect food as sufficient evidence to warrant a recall because the outbreak strain might have been transferred to the open package from an external source after it was opened. However, the chance of finding a pattern as rare as the one displayed by the outbreak strain in packages contaminated from different external sources from unrelated households in separate geographic locations is very low. A common source of the strain (i.e., contamination at the production plant in this situation) is a much more likely explanation. This evidence was additionally supported by traceback from a number of patients who had indicated consumption of ground beef and further microbiological evidence collected at the plant. The meat producer was convinced by the compilation of evidence and therefore initiated a voluntary recall.

**Manual or automated cluster detection.** The CDC centralized cluster detection process was particularly helpful in identifying small and diffuse clusters and led to the
recognition of five outbreaks in 2002 not recognized through the Web conference system. The cluster detection process is presently not automated. The PulseNet database team members perform the cluster analysis manually. Thus, the cluster detection process is subjective and highly dependent on the experience of the team members. This limitation can result in the detection and investigation of too many clusters or too few, with the possibility that outbreaks might be overlooked if the clusters are not detected. The PulseNet centralized cluster detection process could be significantly enhanced with a cluster detection algorithm similar to the cumulative sum-based algorithm (CUSUM) developed for detecting *Salmonella* serotype clusters of illness (4). However, the cumulative sum algorithm requires 3 to 5 years of high-quality retrospective data for the algorithm baseline. The PulseNet STEC O157 database is just beginning to meet this requirement. Additionally, the cumulative sum algorithm will run slowly on an image database because of the complexity of the data, and the cluster signals would be of doubtful quality. It could be implemented effectively and efficiently with the use of the PulseNet standardized pattern designations in a manner analogous to serotype designations. But to implement this, all PFGE patterns submitted to PulseNet databases would have to be named as soon as they were uploaded. Because the personnel dedicated to PulseNet databases at CDC has not kept pace with the exponential increase in the PFGE patterns submitted to PulseNet national databases, only approximately two-thirds of the more than 3,000 profiles submitted were named in 2002. An automated pattern-naming feature for the PulseNet customized version of BioNumerics is currently being developed to solve this problem. Once this feature is implemented, an cumulative sum algorithm will be implemented to facilitate automated computer-assisted cluster detection in PulseNet.

**Minimizing the time before an outbreak is detected.**

The PulseNet cluster detection system for STEC O157 efficiently detects local and multistate outbreaks and provides early warning of potential outbreaks even when the number of cases is very small. This is in sharp contrast to the situation before PulseNet was implemented. Figure 2 shows the epidemic curves of two STEC O157 outbreaks, one occurring in the western states in 1993 (1, 16), before the implementation of PulseNet, and the other occurring in 2002 (outbreak 1, this article). In the absence of routine molecular surveillance of foodborne pathogen isolates, it took 39 days after the first person fell ill to recognize the outbreak in 1993. An alert pediatrician detected the outbreak on clinical grounds. It took 10 to 12 more days to identify the source. The net result was that 726 people became ill after eating contaminated hamburgers at the implicated fast-food chain and four children died. In contrast to the 1993 situation, the PulseNet participant in Colorado detected the 2002 multistate outbreak 18 days after the first case patient became ill at a time when there were only 15 cases.

**Low number of food isolates.**

The food regulatory agencies submit a very low number of food isolate PFGE patterns because of the low level of contamination of food with STEC (3). Submissions that are often triggered by a suspicious link between a human patient and a specific food item more frequently confirm the source and support further actions by regulatory agencies to prevent further exposure of consumers to a contaminated food item.
New typing methods. PFGE is a comparative typing method by nature. In PulseNet, it has proven efficient as an outbreak detection and controlling tool when rigorously standardized between the participating laboratories. However, PFGE results are hardly portable, the typing procedure is not cheap, and it is rather slow; it takes 1 to 3 days from when the strain has been isolated until the PFGE profile has been obtained. Therefore, PulseNet has also initiated the development of alternative methods. These should ideally be universally applicable, more rapid, cheap, and as discriminatory as PFGE. The subtypes generated should be definitive, portable, and easily comparable. Different PulseNet laboratories are currently exploring one such method, multilocus variable number of tandem repeats analysis (MLVA) (7, 10), and more are under way. Before they can be implemented, they will need to be thoroughly validated so as not to lose the many years of information generated by PFGE the day that method is abandoned. The new methods have not been fully developed so far, and the validation process has barely begun. Therefore, it is expected that PFGE will remain the “gold standard” for PulseNet for at least 5 years.

CONCLUSIONS

The PulseNet system has served as an effective and efficient early warning system for detecting, investigating, and controlling foodborne disease outbreaks caused by STEC O157. However, the overall incidence of STEC O157 infections in the United States has only just recently started to decline (2). This highlights the complexities involved in prevention of foodborne infections and underscores the need for concerted and coordinated action by decision makers, microbiologists, epidemiologists, and other experts in government agencies and the food industry.

REFERENCES