Trace-Back and Trace-Forward Tools Developed Ad Hoc and Used During the STEC O104:H4 Outbreak 2011 in Germany and Generic Concepts for Future Outbreak Situations

Armin A. Weiser,¹ Stefan Gross,¹ Anika Schielke,¹ Jan-Frederik Wigger,¹ Andrea Ernert,¹ Julian Adolphs,¹ Alexandra Fetsch,¹ Christine Müller-Graf,¹ Annemarie Käsbohrer,¹ Olaf Mosbach-Schulz,² Bernd Appel,¹ and Matthias Greiner^{1,3}

Abstract

The Shiga toxin-producing *Escherichia coli* O104:H4 outbreak in Germany in 2011 required the development of appropriate tools in real-time for tracing suspicious foods along the supply chain, namely salad ingredients, sprouts, and seeds. Food commodities consumed at locations identified as most probable site of infection (outbreak clusters) were traced back in order to identify connections between different disease clusters via the supply chain of the foods. A newly developed relational database with integrated consistency and plausibility checks was used to collate these data for further analysis. Connections between suppliers, distributors, and producers were visualized in network graphs and geographic projections. Finally, this trace-back and trace-forward analysis led to the identification of sprouts produced by a horticultural farm in Lower Saxony as vehicle for the pathogen, and a specific lot of fenugreek seeds imported from Egypt as the most likely source of contamination. Network graphs have proven to be a powerful tool for summarizing and communicating complex trade relationships to various stake holders. The present article gives a detailed description of the newly developed tracing tools and recommendations for necessary requirements and improvements for future foodborne outbreak investigations.

Introduction

 \mathbf{F}^{ROM} Мау то July 2011, a large Shiga toxin-producing Escherichia coli (STEC) serotype O104:H4 outbreak осcurred in Europe (mainly Germany) with 2,987 reported and symptomatic STEC cases, an unusually high number (855) of hemolytic uremic syndrome (HUS) complications, and in total 53 (STEC, 18; HUS, 35) fatal cases (Askar et al., 2011; Frank et al., 2011; RKI, 2011). In the aftermath, an international workshop held by the Robert Koch Institute (RKI) in November 2011 addressed lessons learned and future research needs (STEC Workshop Reporting Group, 2012). As the identification of the probable vehicle and source was unsuccessful through microbiological analysis, but through the restaurant cohort study of the RKI and the trace-back and trace-forward data analysis of the Task Force EHEC (Appel et al., 2011; Bernard et al., 2011; Buchholz et al., 2011; Federal Institute for Risk Assessment, 2011a; LAVES and NLGA, 2011; RKI, 2011), the importance of tools and concepts for trade network investigations was addressed as one major issue for future outbreak investigations. In general, it was emphasized that one should use this outbreak as an example to learn as much as possible to improve the handling of future outbreak situations (STEC Workshop Reporting Group, 2012).

Therefore, we describe here the information technology tools for trace-back and trace-forward investigation of food supply chains that were developed *ad hoc* during this large STEC outbreak to identify the vehicle of the pathogen and to stop the outbreak. Requirements for improving the elucidation of future outbreaks are also presented.

Materials and Methods, and Results

General trace-back and trace-forward strategy

A trace-back investigation is the method used to determine and document the full distribution and production chain in order to identify the sources of a product that has been

¹Federal Institute for Risk Assessment (BfR), Berlin, Germany.

²European Food Safety Authority (EFSA), Parma, Italy.

³University of Veterinary Medicine Hannover (TiHo), Hannover, Germany.

Trace-forward

FIG. 1. Scheme of tracing strategies: (1) Disease clusters are identified (by public health authorities). (2) Trace-back is performed in order to identify common nodes of different clusters (straight solid arrow). (3) Trace-forward is performed in order to identify yet unknown or potential clusters (straight dotted arrow). This strategy is usually followed within the context of an ongoing process that involves company audits, local sampling, and further trace-back and trace-forward investigations.

implicated in a foodborne outbreak (EFSA, 2011b). Without reliable information on the contaminated vehicle, the first step is usually the selection of cases complying with a case definition. In more complex situations, outbreak clusters are chosen if available that may allow microbiological investigation of retained food samples and/or interviews with human cases to identify food commodities or recipe components as candidates for trace-back. If a common source (e.g., producer or supplier) can be identified, it might be possible to investigate the entry site and probable source of the contamination (Fig. 1). Subsequently, trace-forward from this common network node in direction to the customer is a strategy to detect yet unknown outbreak clusters or critical nodes (e.g., grocery stores) where contaminated food commodities might still be available (EFSA, 2011b).

Structure of data collecting tools

Up to the German STEC O104:H4 outbreak in 2011, no systematic data collecting tool for an appropriate outbreak tracing along complex food chains and analysis of trade data was available in Germany. Therefore, a new data collecting and exchange format together with new data analyzing tools had to be developed.



FIG. 2. Database structure. The database consists of five tables that are relationally interconnected by the gray shaded items.

(1) Data collecting tool based on Excel. For a stepwise and consistent trace-back and trace-forward investigation, the data collecting tool had to fulfil three major requirements: (1) it should be easy to handle for a wide variety of users, (2) data should be importable into a centralized database, and (3) it should be possible to forward the data electronically (e.g., via e-mail) to specialists ensuring a rapid working flow.

Based on experiences made during initial trace investigations during the outbreak, an Excel sheet had been designed to collect all information that is essential for a lot-specific traceback and trace-forward. The layout of the tables follows the requirement for the Food Business Operator to document all trade data one-step-up (items received) and one-stepdown (items delivered) according to Regulation No. 178/2002 (European Commission, 2002). The original Excel template with example data and a short manual is provided in the Supplementary Data (available online at www.liebertpub .com/fpd).

(2) Relational JAVA-based database HSQLDB. Hyper-SQL database (HSQLDB: http://hsqldb.org) is an open-source relational database that is Java-based and was used for the newly designed tracing tool. The structure of the relational database (Fig. 2) is closely related to the Excel structure. The database consists of five tables. One central table is "COM-PANY" containing the ID and address-ID of the respective food business operator. For each operator the address details are stored in the table "ADDRESS." The product range of the operator is stored in table "COMPANY_ARTICLE." The table "ARTICLE_DELIVERY" contains the information on particular deliveries of the corresponding operator. And, finally the table "DELIVERY_DELIVERY" connects two supplies with each other. This simple database structure allows the tracing of supply chains with indefinite length. All relationships and further details on the gathered data are illustrated in Figure 2.

Capabilities of the tools for food tracing investigations in outbreak situations

Data collection and import into the database. Available data can be filled into the Excel sheet by the industry/ authority/institution responsible for the trace-back and/or trace-forward investigation. As a prerequisite, the pre-filled tables are sent directly to the companies or competent authorities that have to complete the missing data. Regarding the import into the database, the number and sequence of Excel sheets are irrelevant.

During the import step into the database, the data is tested for correctness and consistency, and redundant data are eliminated. These tests include checks of correct food operator names and addresses, and safe and sound assignment of supply chains, respectively. Each food operator detail (e.g., zip code, street, town) must be unique-which is in comparison to primary data collection (as done by Excel sheets) rarely true in reality. Zip codes are checked for actually existing. Tests for correct food operator assignment include a similarity search for addresses, which can be re-checked interactively. The same similarity algorithm is applied for product names, product numbers, and lot numbers. Additionally, tests were performed for plausible time frames (i.e., the chronology within a supply chain): outgoing deliveries within the same lot have to occur after the incoming deliveries. Furthermore, quantities of each product were checked for (1) getting smaller along the supply chain and (2) being of the same order of magnitude for incoming and outgoing deliveries/storage at each company.

Analysis and visualization of the data. The most selfevident task on the data analysis is the visualization of the supply chains as a network graph. The more data that is collected, the less analysis of the trading network is manageable without computerized tools. Database filters were developed to look at specific categories related to products (e.g., product name, lot number, or food operator name). If the filter criteria are met, then each node of the supply chain is included in the output, even if a node itself does not directly fulfil the filter criteria. For example, one food operator is chosen as filter, and then information on all network nodes directly or indirectly connected by supply chains to this food operator is given as output. An arbitrary number of filters may be put together, resulting in a graph that is more manageable as more filters are set.

The identification of common nodes is straightforward: selecting specific nodes in the network (e.g., the outbreak clusters) and desired filters (e.g., lot number 48088) results in a list of common nodes and their connections among each other within the network.

All filtered data can be exported using standardized data formats (*.xls, *csv, *.xml, *.txt) as interface to other software environments (e.g., R or graphviz) for further analysis. Besides the node identity and edge direction, these sheets can provide further information on product quantity balances of each node, quantity units for each delivery along the network edges, delivery dates, and start and end points of each supply chain.

Visualization of the supply network was accomplished in multiple ways by importing the filtered output format into visualization tools which translated them into graphs or projections. For the R package "network" (Butts, 2008) (http://statnet.org; Fig. 3), a specific graphical style has been developed in order to enhance the visual recognizability of network graphs. In this context, priority was given to visualize important network nodes with many connections to other nodes (i.e., hubs; e.g., producers, suppliers, and clusters), the edges (trade) linking these hubs, and the direction and amount of product (e.g., seeds or sprouts) traded.

Other visualization opportunities were realized with the graphviz visualization software (http://www.graphviz.org; Fig. 4) and as a geographical projection on county or even country level into Google Earth (http://www.google.de/intl/ de/earth/index.html). The latter was achieved by generating a file in kml format that can be easily read by Google Earth (Fig. 5).

Results during the STEC O104:H4 outbreak 2011

Based on information collected and analyzed by the RKI (Federal Institute for Risk Assessment, 2011b; Frank *et al.*, 2011; RKI, 2011) in several case-control and cohort studies, a list of 91 potential food items was generated by the German Task Force EHEC (Bernard *et al.*, 2011), which was the basis for the tracing analysis. Due to the historical involvement of sprouts in several STEC outbreaks (Como-Sabetti *et al.*, 1997; Breuer *et al.* 2001; Ferguson *et al.* 2005; Michino *et al.* 1999; Taormina *et al.* 1999), they were one important food item group on that list; the other groups were herbs, small leaves, lettuces, onion/leek, and fruit vegetables (e.g., tomatoes).



FIG. 3. Combined network of all relevant fenugreek seed and sprout supplies. Supplier network of the STEC O104:H4 outbreak in 2011; results of combined forward and backward tracing of the supply chain. Here, all the companies are represented that have come in contact with the suspected batch of seeds or produced sprouts (created with the R-Package "network," available at http://statnet.org) (Butts, 2008).





FIG. 5. County-based geographical projection of the known distribution of sprouts from the horticultural farm in lower Saxony using Google Earth. The horticultural farm in Lower Saxony is located in the county shown in light gray. Sprouts produced there were delivered into the counties shown in white.

When France reported a cluster of patients with bloody diarrhea, caused by the identical outbreak strain, the same tracing approach was used on the European level to investigate the seeds used for sprouting. A list of six suspicious seed types were generated by the EFSA Task Force (EFSA, 2011b). Finally, the seed supply chains were traced back to December 2009 and the sprout supply chains to mid-April 2011.

Network graphs showed that a horticultural farm in Lower Saxony and several first-step post-suppliers were hubs in the sprout and seed trading network linked to many outbreak clusters (Fig. 3). Furthermore, the trading patterns of the six seed types showed that, within the seed trading network in Germany, there are only four to five hubs ensuring the integrity of the whole network consisting of more than 400 companies (Fig. 4). However, the German and French outbreak could only be connected via one specific lot (48088) of fenugreek seeds imported from Egypt in December 2009 (Fig. 3). From these seeds, sprouts were produced at the horticultural farm in Lower Saxony and by private households in France, thus making fenugreek seeds the most likely underlying source of the outbreak (Bernard *et al.*, 2011; EFSA, 2011a, b). The trace-back/trace-forward

investigations revealed that the same lot of fenugreek seeds was traded via suppliers from Germany, the United Kingdom, and France (Fig. 3) and was split into two different supply chains 16 months before the outbreak event took place. The geographical projection (Fig. 5) indicates that all counties in Germany where sprouts were produced at the horticultural farm in Lower Saxony (county with dark gray border) were traded to. This projection gives also an impression of the geographical scale that can be affected by products of only one business operator (see also geographical projection of STEC cases in RKI [2011]).

Discussion

During investigations of foodborne outbreaks (animaland plant-based vehicles), much emphasis has been laid on case interviews and microbiological analysis of food samples. When there has been a failure to detect the microbiological agent in food, a trace-back strategy combined with epidemiological analysis is a suitable method to identify the source or at least a common node of intersection, as was the case in the German EHEC outbreak in 2011 (Askar et al., 2011; Buchholz et al. 2011). The challenging point is the acquisition and structuring of the trading data. Trading information stems at best from well-designed company databases or at worst from heterogeneous paper delivery receipts and invoices. The approach developed during the STEC O104:H4 outbreak event 2011 in Germany is based on Excel sheet templates with product and trading information connected to a new central database structure. This combination guarantees that data providers-mainly local authorities-are able to supply necessary information electronically, as Excel is in widespread use and is familiar to most users. The final structure of the Excel template (e.g., trading connections represented as triads-i.e., one step back, one step forward) was designed according to Regulation No. 178/2002 (European Commission, 2002). It accelerates information gathering, as both trace-back and trace-forward information are acquired simultaneously with just one request of the food business operator. This is especially important in saving time during disease outbreaks. To compensate for the disadvantages of Excel-based information exchange (manual data input is time consuming and error-prone), a new database infrastructure has been developed that supports data correction and integration. The method presented here-using software products that are free of charge-proved to be quick and effective. Available information was automatically double-checked and analyzed immediately. The proper identification of the origin and vehicle of an outbreak is essential for a timely response (i.e., mitigation or recalls) during foodborne disease outbreaks. It took approximately 3 weeks, including data gathering, to develop and implement the tool. The collection of data from one retailer, for example, took about 1-3 days.

FIG. 4. Representation of the delivery quantities and routes of all examined seeds (six varieties, including fenugreek). The rectangles contain anonymized names of companies and provide distributors, producers, and retailers (foreign companies are represented in color). The delivery quantities vary between a minimum of 50-g pack and a maximum of 15 tons—indicated by the thickness of the arrows. The most important spots in the network are circled and colored: horticultural farm in Lower Saxony (yellow), French retailer (green), German importer (blue), and Egyptian exporter (red) (created with graphviz, available at www.graphviz.org).

Due to network complexity, size, and number of nodes, it was not possible to visualize all information manually (e.g., on a flipchart or pinboard). The newly developed tools helped to quickly and promptly visualize all connections between suppliers of interest. This permitted easy communication of results to all people involved in the outbreak investigation.

Once the validated data are entered in the database, they may be used as prerequisite for all kinds of analysis, visualization, risk assessment, and epidemiological investigations. The database solution presented here provides interfaces in the form of different export file types, which can easily be used by other software for further analysis.

In general, additional problems are faced in tracing individual lots. Lot numbers or product names may change along the supply chain. Even the composition may change by product processing, conversion, or addition to other distributed products. Thus, a correct attribution of the product and the corresponding lots and amounts of the different raw materials used to generate the product is necessary for exact trace-back and trace-forward. However, the data structure presented here still misses information on the mixture ratios between different ingredients in the product, which should be considered in the future. Furthermore, it would be helpful for risk assessment to have, firstly, additional information on the intended use of the products on each stage of the supply chain (including further processing). Secondly, it would also be beneficial to have a classification of operators into kind of business (e.g., producer, consumer, retail, wholesale, gastronomy). In future data structures, we recommend accounting for this information.

For general applicability, the question arises as to whether network analysis can provide sufficient evidence for identification of an outbreak source in the absence of microbiological confirmation and if the retrieved trade data is not as abundant as in the case of the STEC outbreak of 2011. On a qualitative level, investigators should then try to assess the completeness and possible biases of the available trade data. It should be ensured that retrieval and analysis of trace-back data is continued even after identification of the first common source if there is still a plausible chance that other sources may as well explain the observed outbreak pattern. On a quantitative level, outbreak-related trade data could be analyzed in combination with more complete trading data generated in "peace time." It can be anticipated that statistical significance tests for identification of a single source on the basis of more complete trading data requires additional assumptions (or data) regarding the homogeneity of the pathogen in the food matrix and consideration of survival and growth of the pathogen as features assigned to both nodes and edges of the trading network.

Another lesson learned is that outbreaks do not stop at administrative borders. During the STEC outbreak, a high level of coordination of tracing and other activities was achieved by holding daily telephone conferences. There is a clear need for data sharing of the involved responsible authorities at the regional, national, and international level.

Conclusion

The trading data collecting and analyzing system developed during the STEC outbreak period proved to be effective in summarizing a large amount of complex trade data and identifying a common link of all outbreak clusters for which data were available. Thus, the successful control of the outbreak can be also attributed to this tool, which was developed in real-time. It is applicable to all kinds of food products, and it proved to be manageable in national as well as in international outbreak investigations. It is particularly useful in large outbreak situations, where the network of involved food producers and distributors is highly complex. Without electronic tools, it would be exhausting and error-prone to analyze complex trading networks.

Therefore, based on the concepts developed and experiences earned during the STEC O104:H4 outbreak in Germany in 2011, a generic methodology of trace-back investigations that is appropriate in different disease outbreak situations should be further developed. Appropriate data acquisition systems should be implemented at all responsible levels in the near future. Such data acquisition systems might either be online accessible databases or a two-step system consisting of data acquisition sheets importable into a centralized database, similar to the tool presented here. We additionally recommend using the triad structure for information gathering in any future data acquisition system to reflect general trading relationships of the business operator, which would accelerate information flow.

Acknowledgments

We thank the German Task Force EHEC, the Robert Koch Institute (RKI), the competent authorities of the federal states (especially Lower Saxony, Schleswig-Holstein, Mecklenburg-Western Pomerania, North Rhine-Westphalia, and Hessen), and Federal Office of Consumer Protection and Food Safety (BVL) for collecting and providing the German data, and EFSA for collecting and providing the European data. We thank all involved staff at the Federal Institute for Risk Assessment supporting the outbreak investigation. Work was funded by the Federal Institute for Risk Assessment.

Disclosure Statement

No competing financial interests exist.

References

- Appel B, Böl G-F, Greiner M, Lahrssen-Wiederholt M, Hensel A. EHEC-Ausbruch 2011—Aufklärung des Ausbruchs entlang der Lebensmittelkette. BfR-Wissenschaft 2011;04/2011.
- Askar M, Faber MS, Frank C, Bernard H, Gilsdorf A, Fruth A, Prager R, Hohle M, Suess T, Wadl M, Krause G, Stark K, Werber D. Update on the ongoing outbreak of haemolytic uraemic syndrome due to Shiga toxin–producing *Escherichia coli* (STEC) serotype O104, Germany, May 2011. Euro Surveill 2011;16, pii: 19883.
- Bernard H, Bisping M, Broschewitz B, Bucher M, Fetsch A, Förster D, Frandrup-Kuhr O, Fricke G, Greiner M, Gross S, Hänel C-M, Heusler K, Jähne J, Kenntner N, Kliemant A, Kühn K, Kutzke M, Ladehoff W, Luber P, Mosbach-Schulz O, Müller B, Rampp A, Reinecke A, Rosner B. Ergebnisbericht der Task Force EHEC zur Aufklärung des EHEC O104:H4 Krankheitsausbruchs in Deutschland. J Verbraucherschutz Lebensmittelsicherheit 2011;6:483–495. (In German.)
- Breuer T, Benkel DH, Shapiro RL, Hall WN, Winnett MM, Linn MJ, Neimann J, Barrett TJ, Dietrich S, Downes FP, Toney DM,

Pearson JL, Rolka H, Slutsker L, Griffin PM; Investigation Team. A multistate outbreak of *Escherichia coli* O157:H7 infections linked to alfalfa sprouts grown from contaminated seeds. Emerg Infect Dis 2001;7:977–982.

- Buchholz U, Bernard H, Werber D, Böhmer MM, Remschmidt C, Wilking H, Deleré Y, an der Heiden M, Adlhoch C, Dreesman J, Ehlers J, Ethelberg S, Faber M, Frank C, Fricke G, Greiner M, Höhle M, Ivarsson S, Jark U, Kirchner M, Koch J, Krause G, Luber P, Rosner B, Stark K, Kühne M. German outbreak of *Escherichia coli* O104:H4 associated with sprouts. N Engl J Med 2011;365:1763–1770.
- Butts CT. network: A package for managing relational data in R. J Stat Softw 2008;241–36.
- Como-Sabetti K, Reagan S, Allaire S, Parrott K, Simonds CM, Hrabowy S, *et al.* Outbreaks of *Escherichia coli* O157:H7 infection associated with eating alfalfa sprouts—Michigan and Virginia, June–July 1997 JAMA 1997;278:809–810.
- European Commission. Regulation (EC) No. 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Official J Eur Commun 2002;L31:1–24.
- [EFSA] European Food Safety Authority. Shiga toxin–producing *E. coli* (STEC) O104:H4 outbreaks in Europe: Taking stock. EFSA J 2011a;9:2390.
- [EFSA] European Food Safety Authority. Tracing seeds, in particular fenugreek (Trigonella foenum-graecum) seeds, in relation to the Shiga toxin–producing *E. coli* (STEC) O104:H4 2011 outbreaks in Germany and France. Technical Report of EFSA. 5 July 2011b. Available at: www.efsa.europa.eu/en/ supporting/pub/176e.htm, accessed November 30, 2012.
- Federal Institute for Risk Assessment/Bundesinstitut für Risikobewertung (BfR). EHEC pathogen not yet typed: tomatoes, cucumbers and salads should nonetheless continue not to be consumed raw. Opinion No. 016/2011 of BfR of 31 May 2011. 2011a. Available at: www.bfr.bund.de/cm/349/ehec_ pathogen_not_yet_typed_tomatoes_cucumbers_and_salads_ should_nonetheless_continue_not_to_be_consumed_raw.pdf, accessed November 30, 2012.
- Federal Institute for Risk Assessment/Bundesinstitut für Risikobewertung (BfR). Preliminary results of the EHEC/ HUS Study. Joint Opinion No. 014/2011 of BfR and RKI of 25 May 2011. 2011b. Available at: www.bfr.bund.de/cm/ 349/preliminary_results_of_the_ehec_hus_study.pdf, accessed November 30, 2012.

- Ferguson DD, Scheftel J, Cronquist A, Smith K, Woo-Ming A, Anderson E, Knutsen J, De AK, Gershman K. Temporally distinct *Escherichia coli* O157 outbreaks associated with alfalfa sprouts linked to a common seed source—Colorado and Minnesota, 2003. Epidemiol Infect 2005;133:439–447.
- Frank C, Faber MS, Askar M, Bernard H, Fruth A, Gilsdorf A, Hohle M, Karch H, Krause G, Prager R, Spode A, Stark K, Werber D; HUS Investigation Team. Large and ongoing outbreak of haemolytic uraemic syndrome, Germany, May 2011. Euro Surveill 2011;16: pii: 19878.
- LAVES and NLGA. EHEC/HUS-Ausbruch 2011: Gemeinsamer Abschlussbericht von LAVES und NLGA für Niedersachsen [EHEC/HUS outbreak 2011: Final report of LAVES and NLGA for Lower Saxony]. 2011. Available at: www.frucht portal.de/aktuelles/lesen/38370/links, accessed November 30, 2012. (In German.)
- Michino H, Araki K, Minami S, Takaya S, Sakai N, Miyazaki M, Ono A, Yanagawa H. Massive outbreak of *Escherichia coli* 0157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. Am J Epidemiol 1999;150:787–796.
- [RKI] Robert Koch Institute. Final presentation and evaluation of the epidemiological findings in the EHEC O104:H4 outbreak, Germany 2011. September 9, 2011. Available at: www.rki.de/ EN/Home/EHEC_final_report.html, accessed November 30, 2012.
- STEC Workshop Reporting Group. Experiences from the Shiga toxin–producing *Escherichia coli* O104:H4 outbreak in Germany and research needs in the field, Berlin, 28–29 November 2011. Euro Surveill 2012;17. Available at: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20091, accessed November 30, 2012.
- Taormina PJ, Beuchat LR, Slutsker L. Infections associated with eating seed sprouts: An international concern. Emerg Infect Dis 1999;5:626–634.

Address correspondence to: Armin A. Weiser, PhD Federal Institute for Risk Assessment (BfR) Department of Biological Safety Unit 43: Epidemiology and Zoonoses Max-Dohrn-Str. 8-10 D-10589 Berlin, Germany

E-mail: armin.weiser@bfr.bund.de