

Review

Cite this article: Ickert C, Cheng J, Reimer D, Greig J, Hexemer A, Kershaw T, Waddell L, Mascarenhas M (2019). Methods for generating hypotheses in human enteric illness outbreak investigations: a scoping review of the evidence. *Epidemiology and Infection* **147**, e280, 1–10. <https://doi.org/10.1017/S0950268819001699>

Received: 1 April 2019

Revised: 17 August 2019

Accepted: 3 September 2019

Key words:

Enteric illness; hypothesis generation methods; outbreaks; scoping review

Author for correspondence:

C. Ickert, E-mail: cickert@ualberta.ca

Methods for generating hypotheses in human enteric illness outbreak investigations: a scoping review of the evidence

C. Ickert¹ , J. Cheng², D. Reimer³, J. Greig³, A. Hexemer², T. Kershaw², L. Waddell³  and M. Mascarenhas³

¹School of Public Health, University of Alberta, Edmonton, Canada; ²Outbreak Management Division, Centre for Food-borne, Environmental and Zoonotic Infectious Diseases, Public Health Agency of Canada, Guelph, Canada and ³Public Health Risk Sciences Division, National Microbiology Laboratory at Guelph, Public Health Agency of Canada, Guelph, Canada

Abstract

Enteric illness outbreaks are complex events, therefore, outbreak investigators use many different hypothesis generation methods depending on the situation. This scoping review was conducted to describe methods used to generate a hypothesis during enteric illness outbreak investigations. The search included five databases and grey literature for articles published between 1 January 2000 and 2 May 2015. Relevance screening and article characterisation were conducted by two independent reviewers using pretested forms. There were 903 outbreaks that described hypothesis generation methods and 33 papers which focused on the evaluation of hypothesis generation methods. Common hypothesis generation methods described are analytic studies (64.8%), descriptive epidemiology (33.7%), food or environmental sampling (32.8%) and facility inspections (27.9%). The least common methods included the use of a single interviewer (0.4%) and investigation of outliers (0.4%). Most studies reported using two or more methods to generate hypotheses (81.2%), with 29.2% of studies reporting using four or more. The use of multiple different hypothesis generation methods both within and between outbreaks highlights the complexity of enteric illness outbreak investigations. Future research should examine the effectiveness of each method and the contexts for which each is most effective in efficiently leading to source identification.

Introduction

Enteric illnesses cause considerable morbidity and mortality worldwide. Waterborne enteric diseases cause 2 million deaths each year, the majority of which occur in children aged 5 and under [1]. Foodborne enteric diseases are responsible for 600 million illnesses and 420 000 deaths annually [2]. These illnesses impact the quality of life of those affected and result in enormous financial consequences for individuals and nations [3]. Although most enteric illnesses are transient, significant chronic sequelae associated with some foodborne pathogens can have long-term public health impacts [4–6].

Enteric illness outbreak investigations seek to identify the source of illnesses to prevent further illness in the population. Timely source identification is a key step towards reducing the incidence of enteric illness worldwide and can lead to change in public health policy or recommendations to prevent future outbreaks, such as changes to food manufacturing processes or regulations. Timely source identification can also lead to public health notices and recalls that may prevent further illnesses in a specific outbreak. Accurate source identification can also provide opportunities to learn more about known and emerging diseases, increase understanding of the impact of current disease prevention practices and improve public confidence in public health agencies responsiveness to disease outbreaks [7].

Outbreak investigations take many forms, depending on the pathogen, context, affected population and suspected route of transmission. Initial cases often alert public health officials that a possible outbreak is occurring. Once an outbreak has been identified a case definition is established to support case finding activities. As cases are identified, information is gathered about the outbreak to generate hypotheses about the potential source(s) and route(s) of exposure. Information can come from a range of sources, including the cases themselves, their friends or family, staff members of businesses and institutions, experts or literature and physical and environmental sampling and inspections. Taken together, this information supports the development of hypotheses about the source of the outbreak.

Hypothesis generation about both the potential source(s) and route(s) of exposure is a key step in outbreak investigations, as it begins the process of narrowing the search for the transmission vehicle. Although some hypothesis generation methods have been described in

© Public Health Agency of Canada 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

summaries of outbreak investigation steps [7, 8], the full range of possible methods used in outbreak investigations or the frequency that they are used is not readily available. We conducted a scoping review to summarise the methods for hypothesis generation used during human enteric illness investigations and to understand the frequency and breadth of methods, as well as to identify knowledge gaps and areas for future research.

Methods

A scoping review protocol was created *a priori* using the framework established by Arksey and O'Malley [9]. A copy of the protocol, including the search strategy, the screening tool and the data characterisation tool can be found in Supplementary Material S1. A full list of the articles identified in this scoping review can be found in Supplementary Material S2. A review team was established and included expertise in synthesis research, food safety, epidemiology and outbreak investigation.

The research question:

What methods have been used, or could be used, in human enteric illness outbreak investigations for hypothesis generation?

Search terms and strategy

A search algorithm (Supplementary Material S1) was constructed using key terms from 30 pre-selected relevant articles and implemented in five databases (PubMed, Scopus, Embase, Cumulative Index to Nursing and Allied Health Literature (CINAHL) and ProQuest Public Health) on 25 May 2015 with a date filter of 1 January 2000–25 May 2015.

The search was evaluated for capture sensitivity by searching reference lists of 12 randomly selected relevant primary methodology papers and 10 of the most recent relevant literature reviews in PubMed (Supplementary Material S1). The grey literature search targeted websites of government and research organisations, and relevant articles from Conference Proceedings (Supplementary Material S1). A total of 202 articles were identified by the grey literature search that were not captured by the search strategy and were added to the literature review (Fig. 1). All citations were exported and de-duplicated in RefWorks (ProQuest, LLC), an online bibliographic management program, before being uploaded into a web-based systematic review management program, DistillerSR™ (Evidence Partners, Ottawa, Canada), for evaluation and characterisation.

Relevance screening of abstracts and full-text citation

Each title and abstract was screened by two independent reviewers using a relevance screening form (Supplementary Material S1). Articles were included if they met the following criteria: (1) used or described methods applicable to enteric illness outbreak investigations to assist in hypothesis generation and source identification; (2) published after 1 January 2000 and (3) were reported in either English or French language. No geographic location was used as an exclusion criterion. The relevance screening form was pretested on 50 citations and resulted in a kappa agreement >0.8, indicating good agreement. Two reviewers screened each citation independently and conflicts were resolved by consensus.

Potentially relevant articles were procured, confirmed to be in English or French and relevant before broadly being characterised by two independent reviewers using a secondary relevance

screening tool (Supplementary Material S1) to gather information on the outbreak, such as geographic location, type of pathogen, setting (single or general) and implicated source (Supplementary Material S1). This form was pretested on 10 papers to ensure good agreement and clarity within the form.

Data extraction and analysis

The data characterisation and utility tool was used to gather data on the hypothesis generation methods used in the outbreak investigation. The form contained check boxes for 23 known hypothesis generation methods and an option for reviewers to add other methods not captured in the form. Clearly established definitions were used to help data extractors distinguish between instances when a method was used for hypothesis generation or hypothesis testing. Hypothesis generation was defined as the process of developing one or more tentative explanations about the source of the outbreak used to inform further investigation. This was distinguished from hypothesis testing, which was defined as the process of confirming that a specific exposure is or is not the cause of an outbreak. Hypothesis testing is performed on a small number of suspect exposures and may include statistical testing or traceback investigation. Sometimes, when the hypothesis is refuted, additional rounds of hypothesis generation may be initiated. Several methods included in the form could be used for either hypothesis generation or for hypothesis testing in outbreak investigations. For example, analytic studies can be used to examine a wide range of exposures to help generate hypotheses about plausible sources. However, analytic studies can also be used to test a hypothesis when a specific source is suspected. Instances where methods were used to test a hypothesis were not relevant to this review and were not captured on the form. Where more than one outbreak was described in a single paper, multiple forms were completed to capture methods used in different investigations. This form was pretested on five papers to ensure agreement between reviewers was adequate and to improve the clarity of the questions/answers where necessary. Two reviewers independently reviewed each paper and disagreements between reviewers were discussed until a consensus was reached or settled with a third reviewer. Articles with no hypothesis generation methods described or with a known source at the outset of the investigation were excluded at this stage. Papers describing methodology, but not specific outbreak investigations, were identified and are described separately. Descriptive statistics were used to summarise the dataset using Stata 15 (StataCorp, 2017).

Results

In total, there were 10 615 unique citations captured by the search (Fig. 1). Of these, 889 (8.4%) papers were fully characterised and included 903 reported outbreaks (Supplementary Material S2). Of the reported outbreaks, 25 (2.8%) were described in 11 multi-outbreak articles and the remaining 878 (97.2%) were described in single outbreak articles (Fig. 1).

The pathogens associated with the outbreaks included: bacteria ($n = 622$, 68.9%), viruses ($n = 192$, 21.3%), parasites ($n = 64$, 7.1%), bio-toxins ($n = 3$, 0.3%), fungi ($n = 1$, 0.1%) and multiple pathogens ($n = 11$, 1.2%). The pathogen was not identified in 10 (1.1%) outbreaks. In terms of outbreak source, 552 (61.1%) identified food as the source, while 103 (11.4%) identified water, 34 (3.8%) identified direct contact with animals, 25 (2.8%) identified

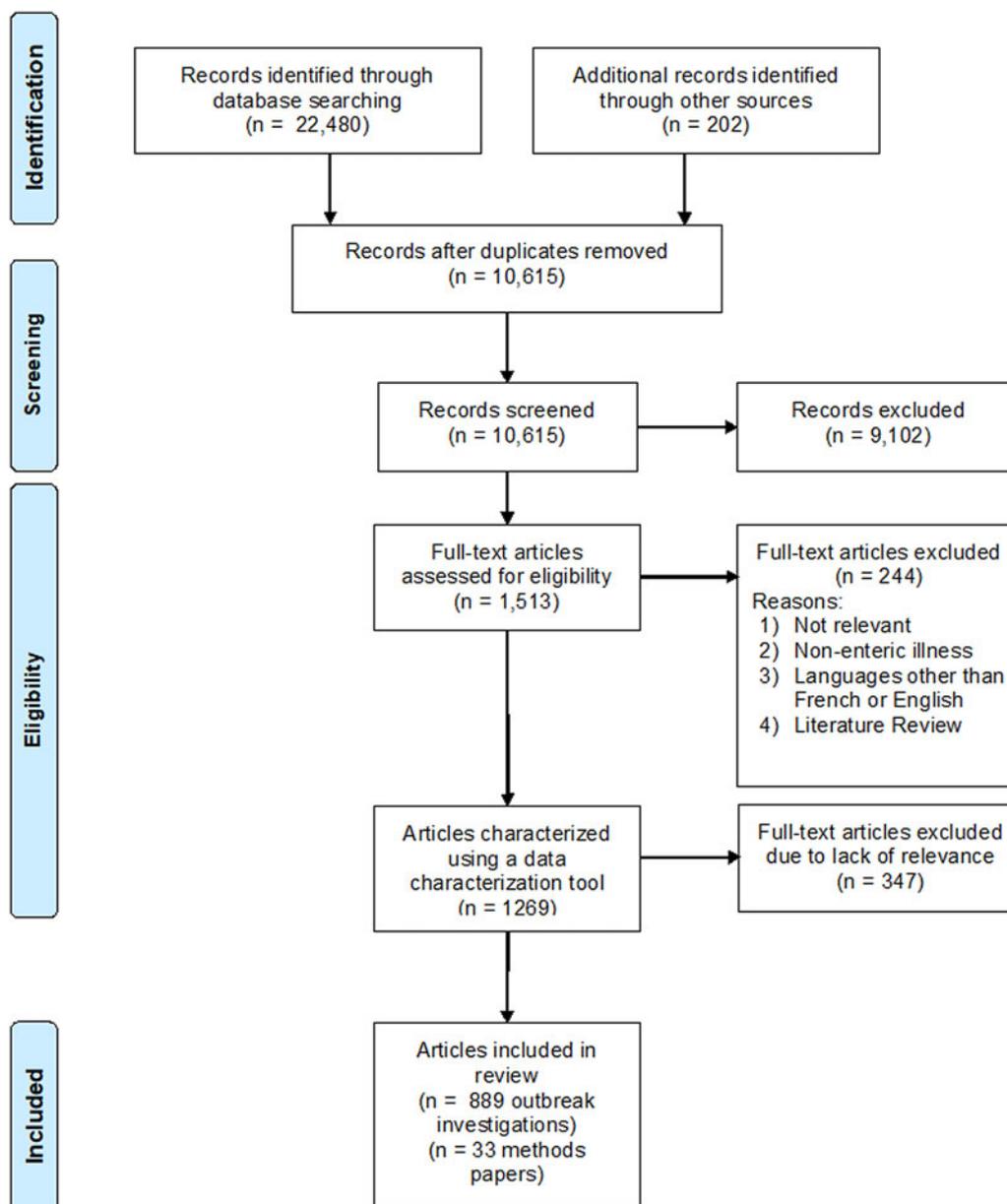


Fig. 1. PRISMA flow chart documenting the literature retrieval and inclusion/exclusion criteria for citations to identify methods of hypothesis generation during human illness investigations.

person-to-person transmission, 25 (2.8%) identified multiple modes of transmission, 20 (2.2%) identified food-handlers, 8 (0.9%) identified soil or environment and 5 (0.6%) reported other modes of transmission as the source. In 131 (14.5%) of the outbreaks, no source was identified.

Hypothesis generation methods used in the enteric illness outbreak investigations are listed and defined in Table 1. The majority ($n = 733$, 81.2%) of investigations employed two or more methods to generate hypotheses; the median number of methods used was three (interquartile range: 2–4). Analytic studies ($n = 585$, 64.8%) were the most commonly reported method category, followed by descriptive epidemiology ($n = 304$, 33.7%), and food or environmental sampling ($n = 296$, 32.8%). Uncommon methods included tracer testing ($n = 1$, 0.1%), anthropological investigation ($n = 1$, 0.1%) and industry consultation ($n = 1$, 0.1%).

Single setting outbreaks

The proportion that each method was used within single setting outbreaks, such as a restaurant, nursing home, or event, is reported in Figure 2. The most commonly reported methods used in single setting outbreaks included analytic studies ($n = 345$, 27.2%), facility inspections ($n = 209$, 16.5%) and food or environmental sampling ($n = 202$, 15.9%). The least common methods used in single setting outbreaks included focus groups ($n = 1$, 0.1%) and tracer testing ($n = 1$, 0.1%). Binomial probability/comparison to population estimates, single interviewer and anthropological investigation were not reported in single setting outbreaks.

General population outbreaks

The proportion that each method was used in general population outbreaks, outbreaks not related to a single event or venue, is

Table 1. Description and frequency of methods used to generate a hypothesis in 903 human enteric illness outbreak investigations identified in scoping review citations

Hypothesis generation method	Definition	<i>n</i> (%) ^a	Setting used <i>n</i>
Questionnaires			
Hypothesis generation questionnaire	Questionnaires designed to capture a large number of exposures to generate hypotheses about possible sources of infection; questions often related to food and water consumption, behavioural habits, travel activities and animal exposures; sometimes referred to as trawling or shot-gun questionnaires.	182 (20.2)	Single: 41 General: 141
Focused questionnaire	Developed for a specific outbreak investigation, often with a shorter, more focused list of exposures than hypothesis-generating questionnaire; types included questionnaires developed based on a specific menu, and questionnaires developed after initial, longer, questionnaires ruled out potential sources.	159 (17.6)	Single: 96 General: 63
Routine questionnaire	Administered as part of initial case (routine) follow-up, often prior to an outbreak being identified or laboratory testing for the pathogen; the questionnaires are usually brief, containing only common risk factors.	133 (14.7)	Single: 43 General: 90
Enhanced surveillance questionnaire	Standardised questionnaire routinely administered as part of an enhanced surveillance initiative for a specific pathogen. Administered to cases following laboratory confirmation for specific pathogens.	13 (1.4)	Single: 3 General: 10
Questionnaire, unspecified	Questionnaires used to identify exposures not described as either focused, routine, enhanced surveillance or hypothesis generating.	127 (14.1)	Single: 42 General: 85
Interviews & focus groups			
In-person interviewing	Face-to-face interviews, sometimes in the cases' home	50 (5.5)	Single: 27 General: 23
Open-ended interviewing	Unstructured, exploratory interview with open-ended questions to collect a detailed exposure history. Questions included food preferences, routines, habits and usual activities.	38 (4.2)	Single: 13 General: 25
Iterative interviewing	Questionnaire items were modified as new cases were interviewed, based on additional information provided by new cases. Exposures reported were amended in the questionnaire for future cases. Previous cases may be re-interviewed with new questions.	16 (1.8)	Single: 4 General: 12
Centralised interviewing	All interviews were conducted by one organisation, with one or more interviewers. Close proximity of interviewers enabled regular discussion of common exposures, which were used to generate hypotheses.	8 (0.9)	Single: 2 General: 6
Focus groups	Multiple interviewers or cases were brought together to discuss exposures to identify commonalities. Group discussion prompted recall of previously forgotten exposures and improved investigators' understanding of plausible sources and transmission routes to support hypothesis generation.	6 (0.7)	Single: 1 General: 5
Single interviewer	All interviews were conducted by the same person, which facilitated hypothesis generation because one can more easily identify commonalities across cases during interviews.	4 (0.4)	Single: 0 General: 4
Food displays	Photographs or physical plates of food used during case interviews to help trigger better recall of exposures from cases.	2 (0.2)	Single: 2 General: 0
Industry consultation	Consultations with independent industry experts to help generate hypotheses about suspected food items of interest or sources of contamination in the food production process.	1 (0.1)	Single: 0 General: 1
Analytic methods			
Analytic Study	An analytic study conducted in the absence of a clearly stated hypothesis. Used to identify significantly different exposures between cases and controls. Types included: case-control, cohort, case-cohort, case-chaos, case-case and case-crossover.	585 (64.8)	Single: 345 General: 240
Interesting descriptive epidemiology	Examination of unique or interesting features of person, place, or time to identify patterns that provided clues about potential sources of the outbreak.	304 (33.7)	Single: 118 General: 186
Investigation of sub-clusters	Investigation of a localised event or non-household setting, such as a restaurant, linked to two or more cases in the outbreak to help identify common exposures.	37 (4.1)	Single: 4 General: 33

(Continued)

Table 1. (Continued.)

Hypothesis generation method	Definition	<i>n</i> (%) ^a	Setting used <i>n</i>
Binomial Probability/comparison to population estimates	Case exposure frequencies were compared to background rates or population exposure estimates, often using binomial probability calculations, to generate hypotheses about likely sources. Hypotheses were based on a significantly higher level of exposure among cases compared to the baseline population data.	30 (3.3)	Single: 0 General: 30
Investigation of outliers	Examination of one or a subset of cases with unusual exposures or specific food preferences that differed from overall sample. This helped generate new hypotheses or narrow down the number of hypotheses.	4 (0.4)	Single: 2 General: 2
Sampling & inspection			
Food or environmental sampling	Sampling available food items in homes or restaurants, or obtaining environmental swabs of food preparation areas or other plausible sources to identify, through laboratory testing, a source linked to the outbreak.	296 (32.8)	Single: 202 General: 94
Facility inspections	Inspection of a facility to identify possible sources of contamination and foods that might be implicated by such contamination; could involve inspecting food handling and storage practices, food preparation activities, employee hygiene, water sanitation systems, or reviewing policies and procedures.	252 (27.9)	Single: 209 General: 43
Food handler testing	Biological sampling of food handlers working at suspected food establishments. Used to identify, through laboratory testing, a source linked to the outbreak.	23 (2.5)	Single: 19 General: 4
Household inspection	Inspection of a case's home to identify possible sources of contamination and foods that might be implicated by such contamination. Could involve inventories of pantry items for comparison across cases to aid in hypothesis generation of common exposures.	2 (0.2)	Single: 0 General: 2
Other methods			
Review of existing information	Reviewing information sources to generate hypotheses about previously reported exposures to the pathogen or biologically plausible exposures; sources included peer-reviewed scientific or grey literature, published reports, or disease surveillance systems.	86 (9.5)	Single: 14 General: 72
Epidemiology traceback	Traceback to determine whether food consumed by multiple cases commonly converges in the supply chain or to compare the distribution of illnesses to the distribution of a food commodity to see if patterns emerged to help generate hypotheses.	56 (6.2)	Single: 10 General: 46
Menu or recipe analysis	Review of a menu or recipes to verify exposures reported by cases, or to identify specific ingredients within reported meals.	51 (5.7)	Single: 34 General: 17
Purchase records	Records of sales transactions, such as receipts, bank statements, or loyalty card history, used to verify exposure, identify commonalities between cases, or obtain product details. Institutional purchase and delivery records reviewed to generate hypotheses about plausible outbreak sources.	39 (4.3)	Single: 8 General: 31
Anecdotal reports	Unverified reports or suspicions from cases/external sources, such as the public or medical professionals, about the potential source(s) of an outbreak. Obtained directly from individuals, or through online media such as web forums or social media.	37 (4.1)	Single: 24 General: 13
Spatial epidemiology	Spot-mapping or geo-mapping cases to identify potential location-based linkages across cases, such as common grocery stores, activities or neighbourhoods.	6 (0.7)	Single: 2 General: 4
Contact tracing/social network analysis	Identification of all people who came into contact with a case to provide clues regarding plausible sources of illness.	3 (0.3)	Single: 2 General: 1
Anthropological investigation	Team of anthropologists employing ethnographic techniques to understand culturally-specific exposures; helped develop culturally-appropriate questionnaire for hypothesis generation within local language and customs.	1 (0.1)	Single: 0 General: 1
Tracer testing	Fluorescent dyes placed in a water or sanitation system to understand connections and travel time of water or effluent, which helped generate hypotheses about sources of water contamination.	1 (0.1)	Single: 1 General: 0

^aPercentages will not sum to 100% as outbreak investigators could use multiple methods to generate hypotheses.

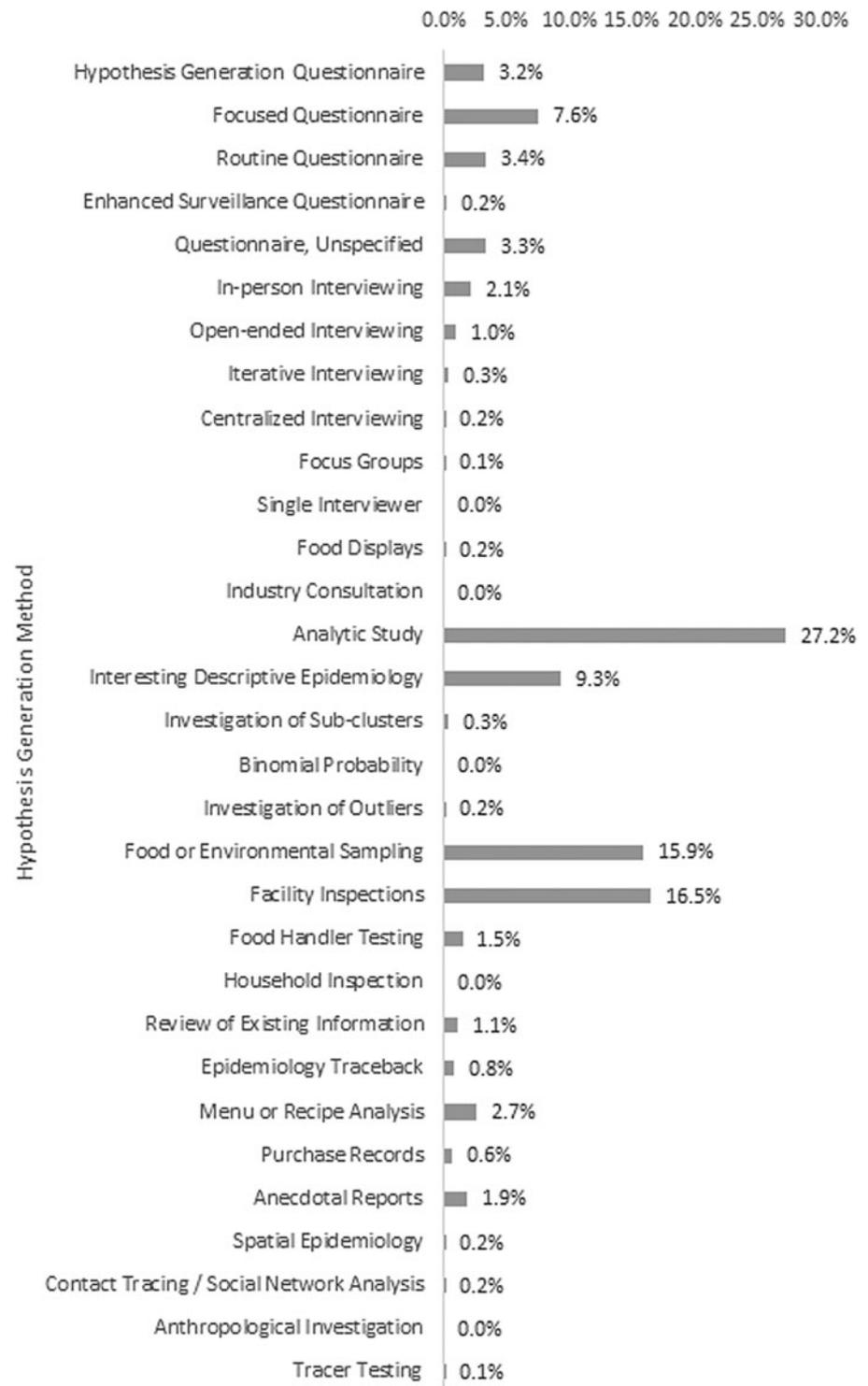


Fig. 2. Hypothesis generation methods used in single setting outbreaks.

reported in Figure 3. The most commonly used methods in general population outbreaks included analytic studies ($n = 240$, 18.7%), interesting descriptive epidemiology ($n = 186$, 14.5%) and hypothesis generation questionnaires ($n = 141$, 11.0%). The least common methods used in general population outbreaks included anthropological investigation ($n = 1$, 0.1%), contact tracing/social network analysis ($n = 1$, 0.1%) and industry consultation ($n = 1$, 0.1%). Tracer testing and food displays were not reported in general population outbreaks.

Hypothesis generation innovation and trends 2000–2015

Trends in method use over the 15-year span were examined in 5-year increments (Supplementary Material S3). Small increases were observed in the use of anecdotal reports, purchase records, binomial probability/population comparison, facility inspections and review of existing information. A decline was observed in the use of analytic studies. Other methods had variable use over the time period or were relatively stable.

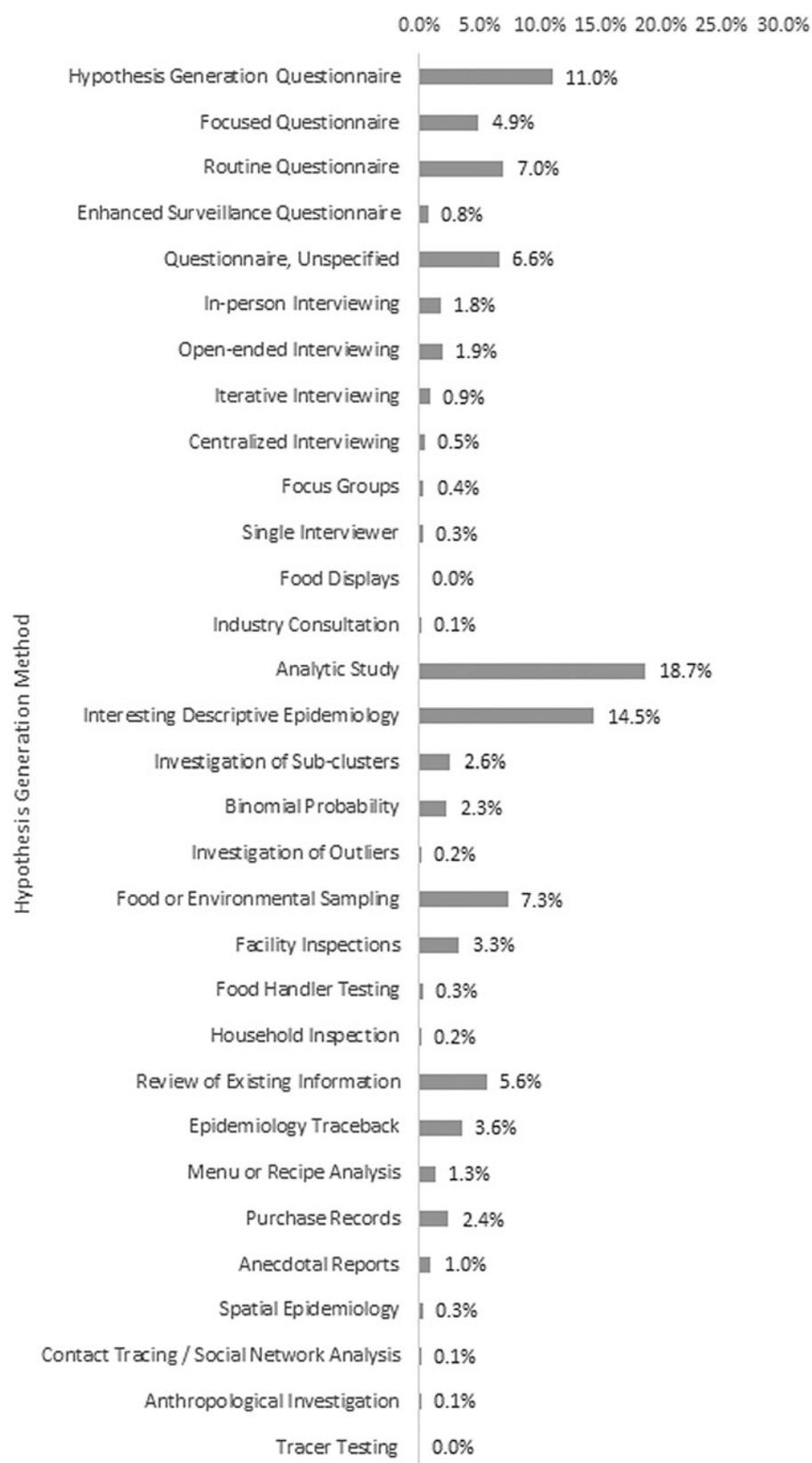


Fig. 3. Hypothesis generation methods used in general population outbreaks.

Methodology papers

Of the 10 615 citations screened, 33 (0.3%) methods papers were identified (Supplementary Material S2). These papers focused on

evaluating existing methods or comparing standard *vs.* a novel approach to hypothesis generation (Supplementary Material S4). Of these, the most commonly discussed method was analytic studies ($n = 11$, 33.3%). This included five on the validity of

case-chaos methodology [10–14], two on case-case methodology [15, 16], two on case-control methodology [17, 18], one discussing the validity of case-cohort methodology [19] and one discussing the validity of case-crossover methodology [20].

The use of laboratory methods, including whole genome sequencing, was described in five (15.2%) papers [21–25]. Traceback procedures were explored in five (15.2%) papers, including three on the use of network analysis [26–28], one on the use of food flow information [29] and one examining the use of relational systems to identify sources common to different cases [30]. Four (12.1%) papers described broad outbreak investigation activities, which included the hypothesis generation step, one from the United Kingdom [31], one from Quebec, Canada [32], one from Minnesota [33] and one from the Centers for Disease Control and Prevention (CDC) in the United States [34]. Three (9.1%) papers explored interviewing techniques, two examining the use of computer assisted telephone interviews (CATI) technology [35, 36] and one on when to collect interview-intensive dose-response data [37]. Three (9.1%) papers compared online questionnaires to phone or paper questionnaires [38–40]. Finally, one (3.0%) paper examined the use of mathematical topology methods to generate hypotheses [41] and another (3.0%) paper examined the use of sales record data to generate hypotheses [42].

Discussion

The most commonly reported hypothesis generation methods identified in this scoping review included analytic studies, descriptive epidemiology, food or environmental sampling and facility inspections. Uncommon methods included industry consultation, tracer testing, anthropologic investigations and the use of food displays. Most outbreak investigations employed multiple methods to generate hypotheses and the context of the outbreak was an important determinant for some methods.

The multitude of hypothesis generation methods described and the use of multiple methods by most outbreak investigators point to the complexity of investigating enteric illness outbreaks. Many methods described are complementary with other methods or may be used in sequence as an investigation progresses. For example, routine and enhanced surveillance questionnaires will often be collected before an outbreak is even identified, while hypothesis generating questionnaires are frequently used at the beginning of an outbreak when the focus of the investigation is quite broad. The use of descriptive epidemiology is generally based on questionnaire data and is often one of the first hypothesis generation methods employed in outbreak investigations. Other methods, such as food or environmental sampling, facility inspections and food handler testing may be used in conjunction with questionnaires, particularly if the outbreak occurred in one setting or at an event. Both open-ended and iterative interviewing frequently occur later in investigations when no obvious source has emerged or as new cases are identified.

Investigators consider many factors when choosing a hypothesis generation method. For example, the length of time that has elapsed between case exposure and the identification of outbreak impact investigation tools such as the collection of contaminated food and environmental samples or facility inspections and traceback investigations [43–45]. Cost and feasibility are also important considerations for many hypothesis generation methods. Analytic studies can be expensive and time consuming [46], while food and environmental sampling requires laboratory

resources for testing [47, 48]. Changes in method type used over time, for example increases in the use of anecdotal reports and purchase records, likely reflect the increase in available technology such as online reporting through social media, and availability of online records. The decline in the use of analytic methods may reflect the increased availability of other, less expensive, hypothesis generation methods such as population comparisons or purchase records.

Outbreak setting can impact the choice of hypothesis generation methods. Methods frequently used in single setting outbreaks include tailored menu-based interviewing, facility inspections and food handler testing. These methods are well-suited to these settings because the common connection across cases is obvious and the source is expected to be identified at a single location common to the cases, such as a restaurant or hospital. For outbreaks related to a single event such as weddings or conferences, analytic studies such as a retrospective cohort are well suited to investigating known exposed populations. In contrast, the use of purchase records, such as store loyalty cards or credit card statements, is utilised when the outbreak is among the general population and there appears to be no obvious connection between cases. Similarly, a review of existing information is a method used frequently in outbreaks among the general population when the range of plausible sources of illness is substantially larger than would be present in single event outbreaks. Outbreak setting thus has implications for the feasibility and usefulness of many hypothesis generation methods.

One finding of this scoping review is that hypothesis generation methods are not well reported within outbreak reports. Descriptions of hypothesis generation methods and sequence of events were often limited or entirely omitted from the publications. This incomplete reporting makes it difficult to interpret how frequently some methods are used by outbreak investigation teams compared to what outbreaks are written up and published in detail. Thus, it is likely that some common methods such as routine questionnaires were underreported and are thus underrepresented in this review. Methods that did not contribute to the identification of the source may also not be reported. Thorough reporting of all hypothesis generation methods used by outbreak investigators would allow for a more comprehensive understanding of the range and frequency of methods used to investigate outbreaks.

Most of the methods papers identified in this review focused on analytic studies, laboratory methods, traceback, interviews and questionnaires. No methods papers were identified related to several hypothesis generation methods reported in this review, including focus groups, iterative interviewing, open-ended interviewing, descriptive epidemiology, sub-cluster and outlier investigation, food or environmental sampling, facility inspections, food handler testing, review of existing information, menu or recipe analysis, anecdotal reports and social network analysis. The paucity of methods papers exploring hypothesis generation methods is an important literature gap. The relative merits of different hypothesis generation methods, their validity and reliability and comparable effectiveness across outbreak investigations, are needed to support outbreak investigator decision-making.

The frequencies of hypothesis generation methods reported in this scoping review may differ from their frequencies in practice as most outbreaks identified had successfully identified the source of the outbreak. Only 15% failed to identify the source of the outbreak, which is a much lower proportion than expected in practice [49, 50]. This suggests that investigations where the source is not

identified are less likely to be published and/or are published with few details, so they did not fulfil the inclusion criteria. This under-reporting makes it impossible to accurately assess individual hypothesis generation methods' relative impact on investigation success based solely on published literature. Increased reporting of outbreak investigations where the source is not identified would improve our understanding of effective *vs.* ineffective hypothesis generation method use. Alternatively, organisations with access to administrative data on a full complement of outbreaks could analyse the relationship between the hypothesis generation methods used and associated outcomes of all outbreak investigations. For instance, Murphree *et al.* [49] compared the success of analytic studies to other methods in identifying a food vehicle across all outbreaks in the United States Foodborne Diseases Active Surveillance Network (FoodNet) catchment area. Analytic studies had a 47% success rate compared to all other methods with a 14% success rate [49], suggesting that analytic studies, where feasible, are more likely to lead to the identification of the source. However, given that analytic studies are not always feasible or appropriate, additional information on the relative success of other methods would help outbreak investigators choose appropriate methods to optimise the likelihood of successfully identifying the source. It would be valuable if outbreak investigators reported brief evaluations of their hypothesis generation methods to improve our understanding of the strengths and limitations of each method.

This review employed a comprehensive search strategy to identify enteric outbreak investigations and articles on hypothesis generation methods for outbreaks or other foodborne illness investigations. It is possible that despite our efforts some outbreak reports with hypothesis generation information were missed, as outbreaks are often not reported in the peer-reviewed literature and thus are not indexed in searchable bibliographic databases. To circumvent this shortfall, we performed a comprehensive grey literature search, however, it is possible some relevant reports were missed. It is also possible that there is some language bias, as the search was conducted in English and only papers reported in English or French were included in the review. This may have resulted in a failure of the search to identify relevant non-English papers. The effect of this on our results and conclusions is unknown. Lastly, because some methods identified in this review could be used for either hypothesis generation or hypothesis testing, we may have misclassified some uses of those methods as hypothesis generation when the investigators actually used the method for hypothesis testing. We relied on author reporting to understand when hypothesis generation was taking place, but incomplete or inadequate reporting may have resulted in misclassification that overestimated the extent to which some methods, such as analytic studies, are used to generate hypotheses.

This review demonstrated the range of hypothesis generation methods used in enteric illness outbreak investigations in humans. Most outbreaks were investigated using a combination of methods, highlighting the complexity of outbreak investigations and the requirement to have a suite of hypothesis generation approaches to choose from, as a single approach may not be appropriate in all situations. Research is needed to comprehensively understand the effectiveness of each hypothesis generation method in identifying the source of the outbreak, improving investigators' ability to choose the most suitable hypothesis generation methods to enable successful source identification.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268819001699>.

Acknowledgements. The Public Health Agency of Canada library for their help in the procurement of publications. The Public Health Agency of Canada Centre for Food-borne, Environmental and Zoonotic Infectious Diseases, Outbreak Management Division contributors: Jennifer Cutler, Kristyn Franklin, Ashley Kerr, Vanessa Morton, Florence Tanguay, Joanne Tataryn, Kashmeera Meghnath, Mihaela Gheorghe, Shiona Glass-Kaastra.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of interest. None.

References

1. **World Health Organization** (2018) *Waterborne Disease Related to Unsafe Water and Sanitation*. Geneva, Switzerland: World Health Organization.
2. **World Health Organization** (2015) *WHO Estimates of the Global Burden of Foodborne Diseases: Foodborne Disease Burden Epidemiology Reference Group 2007–2015*. Geneva, Switzerland: World Health Organization.
3. **McLinden T et al.** (2014) Component costs of foodborne illness: a scoping review. *BMC Public Health* **14**, 509.
4. **Keithlin J et al.** (2014) Systematic review and meta-analysis of the proportion of campylobacter cases that develop chronic sequelae. *BMC Public Health* **14**, 1203.
5. **Keithlin J et al.** (2015) Systematic review and meta-analysis of the proportion of non-typhoidal salmonella cases that develop chronic sequelae. *Epidemiology and Infection* **143**, 1333–1351.
6. **Scallan E et al.** (2015) An assessment of the human health impact of seven leading foodborne pathogens in the United States using disability adjusted life years. *Epidemiology and Infection* **143**, 1–10.
7. **Reingold A** (1998) Outbreak investigations – a perspective. *Emerging Infectious Diseases* **4**, 21–27.
8. **World Health Organization** (2008) *Foodborne Disease Outbreaks: Guidelines for Investigation and Control*. Geneva, Switzerland: World Health Organization, NLM classification: WC 260.
9. **Arksey H and O'Malley L** (2005) Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* **8**, 19–32.
10. **Doerken S et al.** (2014) Assessment of the 'Case-chaos' design as an adjunct to the case-control design. *American Journal of Epidemiology* **179**, 775–780.
11. **Höhle M** (2013) Re: 'The "Case-chaos study" as an adjunct or alternative to conventional case-control study methodology'. *American Journal of Epidemiology* **177**, 1021–1022.
12. **Gillespie IA et al.** (2012) The 'case-chaos study' as an adjunct or alternative to conventional case-control study methodology. *American Journal of Epidemiology* **176**, 497.
13. **Pulliam JRC and Dushoff J** (2014) Re: 'The "Case-chaos study" I adjunct or alternative to conventional case-control study methodology'. *American Journal of Epidemiology* **179**, 261–262.
14. **Edelstein M, Wallensten A and Kuhlmann-Berenzon S** (2014) Is case-chaos methodology an appropriate alternative to conventional case-control studies for investigating outbreaks? *American Journal of Epidemiology* **180**, 406–411.
15. **Gillespie IA et al.** (2002) A case–case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: a toll for generating hypotheses. *Emerging Infectious Diseases* **8**, 937–942.
16. **Wilson N et al.** (2008) Case-case analysis of enteric diseases with routine surveillance data: potential use and example results. *Epidemiologic Perspectives & Innovations* **5**, 6.
17. **Gu W et al.** (2015) Use of random forest to estimate population attributable fractions from a case-control study of salmonella enterica serotype enteritidis infections. *Epidemiology and Infection* **143**, 2786–2794.
18. **Jervis R et al.** (2015) Another nail in the coffin? moving away from population-based case-control studies during outbreak investigations.

- Integrated Foodborne Outbreak Response and Management Conference*, vol. 41.
19. **de Waroux OLP, Maguire H and Moren A** (2012) The case-cohort design in outbreak investigations. *Eurosurveillance* **17**, 1.
 20. **Haeghebaert S, Duche L and Desenclos JC** (2003) The use of the case-crossover design in a continuous common source food-borne outbreak. *Epidemiology and Infection* **131**, 809–813.
 21. **Gilpin B** (2015) Campylobacteriosis: outbreak investigation using MLPA-based binary typing (MBiT). *Integrated Foodborne Outbreak Response and Management Conference*, vol. 58.
 22. **Gerner-Smidt P** (2015) Transforming public health microbiology in the United States with Whole Genome Sequencing (WGS) – PulseNet and beyond. *International Conference on Emerging Infectious Diseases*, vol. 13.
 23. **Jackson B et al.** (2015) Linking Listeriosis to recalled foods: use of Whole-genome Sequencing to enhance public health investigations. *International Conference on Emerging Infectious Diseases*, vol. 302.
 24. **McDonnell J et al.** (2013) Retrospective analysis of whole genome sequencing compared to prospective typing data in further informing the epidemiological investigation of an outbreak of *Shigella sonnei* in the UK. *Epidemiology and Infection* **141**, 2568–2575.
 25. **De Lappe N and Cormican M** (2015) Case study of the use of pulsed field gel electrophoresis in the detection of a food-borne outbreak. *Methods in Molecular Biology* **1301**, 35–40.
 26. **Doerr D et al.** (2012) Accelerating investigation of food-borne disease outbreaks using pro-active geospatial modeling of food supply chains. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems 2012*, 44–47. doi: 10.1145/2452516.2452525.
 27. **Manitz J et al.** (2014) Origin detection during food-borne disease outbreaks – a case study of the 2011 EHEC/HUS outbreak in Germany. *PLoS Currents* **6**, 10.
 28. **Yan S et al.** (2012) A case study of CPNS intelligence: provenance reasoning over tracing cross contamination in food supply chain. *Proceedings – 32nd IEEE International Conference on Distributed Computing Systems Workshops, ICDCSW 2012*. pp. 330–335. doi: 10.1109/ICDCSW.2012.67.
 29. **Huber U and Luber P** (2014) Follow the food flow – investigational tracing for outbreak clarification. *European Scientific Conference on Applied Infectious Disease Epidemiology*, vol. 82.
 30. **Weiser AA et al.** (2013) 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. *Foodborne Pathogens and Disease* **10**, 263–269.
 31. **Harker KS et al.** (2013) National outbreaks of salmonella infection in the UK, 2000–2011. *Epidemiology and Infection* **142**, 601–607.
 32. **Gaulin C et al.** (2014) Challenges in listeriosis cluster and outbreak investigations, province of Quebec, 1997–2011. *Foodborne Pathogens and Disease* **11**, 1–7.
 33. **Rounds JM et al.** (2012) Dynamics of *Escherichia coli* O157:H7 outbreak detection and investigation, Minnesota 2000–2008. *Epidemiology and Infection* **140**, 1430–1438.
 34. **Sobel J et al.** (2002) Investigation of multistate foodborne disease outbreaks. *Public Health Reports* **117**, 8–19.
 35. **Fox LM et al.** (2005) Emergency survey methods in acute cryptosporidiosis outbreak. *Emerging Infectious Diseases* **11**, 729–731.
 36. **Kirk M et al.** (2006) Computer-assisted telephone interview techniques. *Emerging Infectious Diseases* **12**, 697–698.
 37. **Jones RC et al.** (2004) Intensive investigation of bacterial foodborne disease outbreaks: proposed guidelines and tools for the collection of dose-response data by local health departments. *Journal of Food Protection* **67**, 616–623.
 38. **Oh JY et al.** (2010) Comparison of survey methods in norovirus outbreak investigation, Oregon, USA, 2009. *Emerging Infectious Diseases* **16**, 1773–1776.
 39. **Parry AE et al.** (2011) Online questionnaires for outbreak investigations. *Epidemiology* **22**, 875–876.
 40. **Rosner J and Morrison K** (2015) Illness and the internet: an electronic system for online enteric disease food history. *Integrated Foodborne Outbreak Response and Management Conference*, vol. 72.
 41. **Buscema M et al.** (2009) Outbreaks source: a new mathematical approach to identify their possible location. *Physica A: Statistical Mechanics and its Applications* **388**, 4736–4762.
 42. **Kaufman J et al.** (2014) A likelihood-based approach to identifying contaminated food products using sales data: performance and challenges. *PLOS Computational Biology* **10**, e1003692.
 43. **Gutiérrez Garitano I et al.** (2011) Shigellosis outbreak linked to canteen-food consumption in a public institution: a matched case-control study. *Epidemiology and Infection* **139**, 1956–1964.
 44. **Sloan-Gardner TS et al.** (2014) An outbreak of gastroenteritis linked to a buffet lunch served at a Canberra restaurant. *Communicable Diseases Intelligence Quarterly Report* **38**, E273–E278.
 45. **Lienemann T et al.** (2011) Iceberg lettuce as suggested source of a nationwide outbreak caused by two salmonella serotypes, newport and reading, in Finland in 2008. *Journal of Food Protection* **74**, 1035–1040.
 46. **Schmid D et al.** (2011) Foodborne gastroenteritis outbreak in an Austrian healthcare facility caused by asymptomatic, norovirus-excreting kitchen staff. *Journal of Hospital Infection* **77**, 237–241.
 47. **Rakesh PS et al.** (2014) Investigating a community-wide outbreak of hepatitis A in India. *Journal of Global Infectious Diseases* **6**, 59–64.
 48. **Lee VJ, Ong AES and Auw M** (2009) An outbreak of salmonella gastrointestinal illness in a military camp. *Annals of the Academy of Medicine Singapore* **38**, 207–211.
 49. **Murphree R et al.** (2012) Characteristics of foodborne disease outbreak investigations conducted by foodborne diseases active surveillance network (FoodNet) sites, 2003–2008. *Clinical Infectious Diseases* **54**, S498–S503.
 50. **Centers for Disease Control and Prevention** (2017) *Year One Cumulative Metrics Data: Salmonella, Shiga Toxin-Producing Escherichia Coli, Listeria, Shigella, and Campylobacter [SSL(SC)]*. Atlanta, GA, USA: National Center for Emerging and Zoonotic Infectious Diseases.