

# The Effectiveness of Syndromic Surveillance for the Early Detection of Waterborne Outbreaks: A Systematic Review

Susanne Hyllestad (✉ [susanne.hyllestad@fhi.no](mailto:susanne.hyllestad@fhi.no))

Norwegian Institute of Public Health

**Ettore Amato**

Norwegian Institute of Public Health

**Karin Nygård**

Norwegian Institute of Public Health

**Vold Line**

Norwegian Institute of Public Health

**Preben Aavitsland**

Norwegian Institute of Public Health

---

## Research Article

**Keywords:** Syndromic surveillance, early detection, gastrointestinal illness, waterborne outbreaks

**Posted Date:** December 23rd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-128132/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Infectious Diseases on July 20th, 2021. See the published version at <https://doi.org/10.1186/s12879-021-06387-y>.

# Abstract

**Background:** Waterborne outbreaks are still a risk in high-income countries, and their early detection is crucial to limit their societal consequences. Although syndromic surveillance is widely used for the purpose of detecting outbreaks days earlier than traditional surveillance systems, evidence of the effectiveness of such systems is lacking. Thus, our objective was to conduct a systematic review of the effectiveness of syndromic surveillance to detect waterborne outbreaks.

**Method:** We searched the Cochrane Library, Medline/PubMed, EMBASE, Scopus, and Web of Science for relevant published articles using a combination of the keywords 'drinking water', 'surveillance', and 'waterborne disease' for the period of 1990 to 2018. The references lists of the identified articles for full-text record assessment were screened, and random searches using the same key words were conducted. We assessed the risk of bias in the included articles using the ROBINS-I tool and PRECEPT for the cumulative body of evidence.

**Results:** From the 1,955 articles identified, we reviewed 52 articles, of which 16 met the eligibility criteria. Ten were retrospective studies, whereas six were simulation studies. There is no clear evidence for syndromic surveillance in terms of the ability to detect waterborne outbreaks (low sensitivity and high specificity). However, one simulation study implied that multiple sources of signals combined with spatial information may increase the timeliness in detecting a waterborne outbreak and reduce false alarms.

**Conclusion:** This review demonstrates that there is no conclusive evidence on the effectiveness of syndromic surveillance for the detection of waterborne outbreaks, thus suggesting the need to focus on primary prevention measures to reduce the risk of waterborne outbreaks. Future studies should investigate methods for combining health and environmental data with an assessment of needed financial and human resources for implementing such surveillance systems.

## Background

Waterborne outbreaks have a particular high risk for public health, as exposure to drinking water that has been contaminated with pathogens could affect a large population in a relatively short period of time (1). The early detection of infectious diseases is crucial to prevent related consequences, such as the loss of life, adverse health events, and societal burdens (2). Moreover, experience has shown that relying only on the passive surveillance of laboratory-confirmed cases is not sufficient for the timely detection of waterborne outbreaks of non-endemic infections and may contribute to late detection and worse overall health impacts (3). Syndromic surveillance (SyS), which aims to identify a threshold number of early symptomatic cases and facilitate the detection of an outbreak days earlier than conventional surveillance, has been implemented worldwide (4). SyS is defined as the real-time (or near real-time) collection, analysis, interpretation and dissemination of health-related data (2), such as indicators of clinical signs and symptoms, as well as proxy measures such as over-the-counter pharmaceutical sales, hospital admission reports or infectious disease surveillance (5–8).

Although SyS has theoretical advantages in detecting waterborne outbreaks, the approach has been questioned in terms of the resources needed to evaluate the signals and distinguish them from "false alarms", i.e. the specificity (9). A number of factors is needed for evaluating public health surveillance system, including resources needed, usefulness, acceptability amongst other, in particular distinguishing an outbreak from a

“false alarm” (10). Some of the technical core assets to evaluate a surveillance system’s ability to detect a true outbreak is timeliness, sensitivity and specificity (11). Useful surveillance systems for detecting a true outbreak is a balance between the timeliness, sensitivity and specificity, where the ideal situation is to have high values of sensitivity and specificity. However, in reality, this would require a less timely detection (9).

SyS systems for detecting waterborne outbreaks were reviewed in 2006 (12), with the recommendation that such surveillance should not be implemented at the expense of traditional surveillance. On the other hand, Berger et al. (12) also suggested that syndromic data sources, such as the over-the-counter sales of anti-diarrheal medications for detection of waterborne outbreaks, should be further evaluated (12). In the aftermath of this aforementioned review, several articles were published in the field of SyS for waterborne illness and early outbreak detection. However, these articles have not yet been reviewed for the purpose of assessing SyS effectiveness, indicating an updated knowledge gap in this field.

With this review, we aim to provide a knowledge update on the use and effectiveness of SyS approaches to detect waterborne outbreaks among populations connected to water supply systems earlier than traditional surveillance. We have specifically examined reported timeliness, sensitivity and specificity using implemented SyS approaches in contexts where health structures in place. An updated evidence for the effectiveness of the application of SyS will contribute to the evaluation and decision-making processes related to the implementation of this approach.

## **Methods**

### **Literature search**

We searched the Cochrane Library (<http://www.thecochranelibrary.com/>), Medline/PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), EMBASE (<https://www.embase.com/login>), Scopus (<http://www.Scopus.com>), and Web of Science (<https://apps.webofknowledge.com>) for relevant published articles using a combination of the keywords ‘drinking water’, ‘surveillance’, and ‘waterborne disease’. A research librarian conducted the search between January to March of 2019, and the search strategy is described in Additional File 1. The publication period was set from 1990 to 2018, and only peer-reviewed publications in English, German, French, Spanish, and Scandinavian languages (Norwegian, Swedish, and Danish) were included in the search. The bibliographies of the eligible articles were screened to identify additional studies. We also searched Google Scholar for articles using the same key words.

### **Selection criteria**

We included studies on the early detection of waterborne outbreaks using signals from data sources other than diagnostic data. Descriptive and analytical studies or simulation studies on waterborne outbreaks were included in the review. Studies aiming at demonstrating a general association between gastrointestinal illness and drinking water exposure were excluded from the data synthesis. Publications that were included in the review by Berger et al. (12) were also excluded in addition to studies reporting health surveillance due to temporary emergency settings or as a response to natural disasters.

### **Data extraction and analysis**

The literature search output was uploaded in Rayyan (13), where the publications were screened for removing duplicates and processed for further screening. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (14) were followed in the reporting of this review. Two reviewers independently screened the publications' titles and abstracts against the inclusion criteria using the 'blind-on' function in *Rayyan*. Eligible studies for full-text review were further screened independently by two reviewers, and the following summary information was extracted and analysed from publications fulfilling the aim of the review: region/country, objective of study, study design, study period, outbreak cause, affected population, causative agents in the outbreak, and syndrome/data source for surveillance. A list of excluded studies with reasons for their non-inclusion is presented in Additional File 1.

The protocol of this systematic review was also approved by the National Institute of Health Research with the registration number PROSPERO 2019 CRD42019122332 and is available online ([https://www.crd.york.ac.uk/prospERO/display\\_record.php?RecordID=122332](https://www.crd.york.ac.uk/prospERO/display_record.php?RecordID=122332)).

## **Data synthesis**

The information regarding effectiveness of the SyS in detecting waterborne outbreaks (i.e., timeliness, sensitivity, specificity) reported in the included articles was not suitable for pooling due to heterogeneity; therefore, a meta-analysis was not possible. A narrative summary of the findings of the timeliness of detection is presented as a summary in tabular form. Two researchers were involved in the data synthesis.

## **Risk of bias in the individual studies and cumulative evidence**

We used the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) assessment tool to assess the risk of bias in the individual studies (15). The resulting body of evidence of the cumulative result was assessed by the Project on a Framework for Rating Evidence in Public Health (PRECEPT), which was developed by the European Centre for Disease Prevention and Control (ECDC) in 2012 (16, 17).

# **Results**

## **Descriptive summary of study characteristics**

From the 1,955 articles identified in the literature search, screening of bibliographies (of the 27 articles found eligible for full-text screening in the literature search), and random search, 16 articles were included in the review (Fig. 1). A summary of the study characteristics of the included studies is presented in Additional File 1. Of these included articles, 10 were retrospective studies assessing either historical outbreaks or data of cases of gastrointestinal illness and data signals for early detection of waterborne outbreaks (18–27), and six were simulation studies evaluating the system performance of different SyS systems (28–32). The included studies originated from Sweden (n = 2), France (n = 5), the USA (n = 4), and the United Kingdom (n = 4), with one study assessing data from several European countries suggesting a common surveillance approach (25), covering an overall study period of 1997 to 2013, with multiple agents causing waterborne outbreaks or illness. All of the included studies were published in the period 2010 to 2018—except for one that was published in 2006 (21) (Additional File 1).

Among the excluded articles, the majority were data signal studies, including investigations of water quality data or disturbances in the distribution systems, in combination with other signals from the health sector (33–42). One study used web queries to estimate the burden of disease due to gastrointestinal illness related to pipe breaks (7), while another study assessed the relationship between precipitation and waterborne diseases (43). Common in these studies was the fact that, despite that they demonstrated promising correlations, they did not report on the experienced effectiveness or value of using the same signals in their surveillance system explicitly. The other excluded studies addressed SyS but in the context of describing or reviewing such systems in a general manner (44, 45) (Additional File 1).

### **Data synthesis**

The data extracted from the included articles is synthesized in Table 1. When reported, the sensitivity of the SyS in the retrospective studies was below 50%, with one exception reporting a sensitivity of 89% (19). In the simulation studies, the sensitivity was reported above 70% when using different aberration adjustments.

Table 1  
Synthesis of data from the included articles (n = 16)

Data signal	Reference	Timeliness	Sensitivity/Specificity	Pros	Cons
Over-the-counter (OTC) sales of pharmacy sales	Kirian et al., 2011	NI	Sensitivity: 4–14%, specificity: 97–100%.	It may capture symptoms in the population before a person with gastrointestinal illness seeks health care.	It does not necessarily indicate the buyer's location, their demographic status, or the reason for the purchase. Those who purchase OTC medications for their illness may not be representative of the sick population as a whole. Hoarding behaviour will also affect the outcome.
Reimbursement of prescription drugs	Bounoure et al., 2011	NI	Reported sensitivity of 89% and specificity 89% to distinguish acute gastrointestinal cases based on drug sales.	Prescription drug data can be considered for the development of a detection system of waterborne outbreaks given its ability to describe an epidemic signal. It could support	The algorithm cannot be used directly in other countries because of their different health systems, types and sources of data, and medical practices. The accuracy depends on the
	Mouly et al., 2016	NI	Sensitivity: 6% and 21% for two examined outbreaks.	authorities in slow developing outbreaks.	medical consultation rate in the impacted population.
Calls to health advice line ('telehealth')	Bjelkmar et al., 2017	~ 6 days	NI	It may capture symptoms in the population before seeking health care.	The alarm does not contain information regarding the cases' medical status to validate the cause of the alarm. Moderate and low outbreaks (< 1000 cases)

Data signal	Reference	Timeliness	Sensitivity/Specificity	Pros	Cons
	Cooper et al., 2006	Unlikely to detect local outbreak	NI		are unlikely to be detected. The detection ability varies seasonally. Telehealth may, in general, be driven by media bias.
Emergency care data; medical dispatch, ambulance medical service, emergency department chief complaints	Ziemann et al., 2014	NI	NI	This system could detect changes in local trends and clusters of statistical alarms.	It is not likely to detect local gastrointestinal outbreaks with few, mild, or dispersed cases. The probability of detecting an outbreak increases with the outbreak size. The results cannot be generalized to region-level data or very sparse time series.
Over-the-counter (OTC), web queries, calls to health advice line	Andersson et al., 2014	NI	Calls to health advice line: sensitivity: 40–50%, specificity: 99%, web queries and OTC: no signal.	SyS can serve as an early warning, especially with telephone triage data with sufficient temporal and spatial resolution. It may be suited to detecting widespread rises in syndromes and, rarely, small-scale outbreaks.	The alarm does not contain information on the cases' medical status to validate the cause of the alarm. Moderate and low outbreaks (< 1000 cases) are unlikely to be detected.
Telehealth, in-hours and out-of-hours GP, ED visits	Smith et al., 2010	Peak of calls coincides with outbreak (95% CI) in one area	NI	Multiple syndromic data streams are an advantage.	Telehealth may, in general, be driven by media bias.

Data signal	Reference	Timeliness	Sensitivity/Specificity	Pros	Cons
Combined health, spatial and environmental data	Coly et al., 2017	NI	Detected outbreaks < 100 cases.	Increases sensitivity and timely detection of waterborne outbreaks.	These systems are expensive in terms of resources and shared expertise in incorporating local knowledge regarding both environmental and health data.
	Rambaud et al., 2016	NI	NI	Combining two complementary methods protects against false positives, e.g. confusion of cases stemming from exposure from other types of food or swimming, for example.	Pilot-study and not tested on a larger scale.
Multiple SyS and environmental data	Burkom et al., 2011	NI	Sensitivity: 80%, specificity: 99%	Use of multiple syndromic data streams is an advantage. The number of false alarms is greatly reduced.	Simulation results must generally be improved with real epidemiological data.
	Colón-Gonzales et al., 2018	Unlikely to detect outbreaks < 1000 cases	NI	Framework applicable for other SyS systems.	The detection ability varies seasonally.
	Mouly et al., 2018	NI	Sensitivity: 73%, PPV: 90.5%	Space-time increases the likelihood of detecting outbreaks.	The probability of detecting outbreaks increases with the outbreak size.
	Noufaily et al., 2012	NI	Unknown	Able to reduce the number of false alarms and, in some cases, increase the sensitivity.	Week-to-week variation in outbreaks may prove unhelpful.



Data signal	Reference	Timeliness	Sensitivity/Specificity	Pros	Cons
	Zhou et al., 2014	3.3 to 6.1 days	When reported, the sensitivity ranged from 24 to 77%, and the PPV was 90.5%.	Sensitivity and timeliness increase with stratification.	Study population perhaps not representative.
	Xing et al., 2011	NI	Of the simulated models, the regression method had higher sensitivity (range 6–14% improvement of sensitivity in the surveillance system).	Demonstrates possible improvement in the surveillance system to increase sensitivity.	Simulations based on small number of data points.

Note: NI = not identified, PPV = positive predictive value

Some of the included studies addressed the same surveillance system but with different study purposes. In France, a national surveillance system based on administrative health data from the French National Health Insurance on the reimbursement of prescriptive drugs has been functioning since the late 1990s (19). The system contains information on the medications for gastrointestinal illness, which are reimbursable, prescribed by a general practitioners (GPs) and dispensed in a pharmacies covering approximately 98% of the French population. (24). All the included articles originating from French study data were related to this health administrative database.

In the UK, the SyS at Public Health England (PHE) is based on four National Health Service (NHS) healthcare settings: telehealth, in- and out-of-hours, unscheduled care general practitioner consultations, and emergency department (ED) attendances (30). This system has been examined, together with the of the Health Protection Agency (HPA) and QSurveillance, a national surveillance system set up by the University of Nottingham, and the Egton Medical Information System, which consists of a network of GPs (23). Meanwhile, Noufaily et al. assessed weekly counts of samples sent to the HPA. One of the studies in this review included an older version of the SyS in the UK (21).

In the US, several surveillance systems exist (39, 44, 45), and, in this review, we included publications addressing the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) (28). Additionally, two studies assessed the US Centers for Disease Control and Prevention BioSense surveillance system using emergency department chief complaint data (32) and daily syndrome counts from the outpatients of the U.S. Department of Veteran Affairs' Veteran Health Administration (31). Moreover, both of the two included studies from Sweden addressed data signals from Swedish Health Care Direct 1177 (Vårdguiden 1177), along with signals such as web queries and over-the-counter pharmacy sales in one of the study (18, 22).

### Single data signal SyS system

Five of the included studies addressed a single preclinical data signal for outbreak detection and gastrointestinal illness. Kirian et al. (2011) evaluated the ability of drug sales in predicting endemic and epidemic gastrointestinal disease in the San Francisco area and found no significant correlations between drug

sales and illness case counts, outbreak counts, or the number of outbreak-associated cases and reported a low sensitivity (4–14%) and high specificity (97–100%) in the study (26).

In the UK study included in this review, calls made to the health helpline (NHS Direct) were assessed based on whether the number of calls about diarrhoea exceeded a statistical threshold (21). The authors of the study predicted a 4% chance of detection when assumed that one-twentieth of cryptosporidiosis cases telephoned the helpline, which rose to a 72% chance when assumed nine-tenths of cases telephoned. They concluded that NHS Direct was currently unlikely to detect an event similar to the cryptosporidiosis outbreak used in the study and may be most suited to detecting more widespread increases in symptoms (21).

Bjelkmar et al. (2017) extended on such a system for nurse health calls proposed by Andersson (18). The authors compared phone call patterns to the Swedish Health Care Direct 1177 during the outbreak in Skellefteå in different water distribution areas, suggesting that the systematic monitoring of phone calls made to health services could have limited the outbreak from 18,500 cases to approximately 2,300 cases by detecting the outbreak approximately six days earlier than actually detected (22).

### **Multiple data signal SyS systems**

For establishing a national SyS system, Andersson et al. (18) evaluated the efficiency of alternate data sources for the early detection of nine investigated outbreaks in Sweden, including telephone triage, web-queries, and over-the-counter (OTC) pharmacy sales. The authors suggested that SyS can serve as an early warning of waterborne outbreaks, especially with telephone triage data with sufficient temporal and spatial resolution (40–50% sensitivity and 99% specificity); however, data was lacking for outbreaks of moderate size (300-1,000 cases) (18).

Smith et al. (23) evaluated the value of SyS in monitoring small waterborne outbreaks using data from a SyS system featuring a direct telephone helpline and QSurveillance national SyS using clinical diagnosis data extracted from the GP clinical information system (23). The authors reported that, for the first time, such a SyS system was helping to monitor a small-scale waterborne outbreak; however, the peaks of calls to the helpline observed may have been influenced by the media as a boil water advisory was issued during the outbreak (23).

Using routine emergency data based on an inventory of sub-national emergency data available in 12 European countries, Ziemann et al. (2013) proposed a framework of definitions for specific symptoms and a SyS system design applying cumulative sum and spatial-temporal cluster analyses for the detection of local gastrointestinal outbreaks in four countries. Based on the suggested system, the authors identified two gastrointestinal outbreaks in two countries, and 1 out of the 147 confirmed outbreaks in the studied countries was detected (25).

### **Combined SyS systems with environmental data**

Two articles included in this review combined water quality data and information on supply zones in the SyS in France. A pilot study was conducted to assess the utility of using a health insurance database for the automated detection of waterborne outbreaks of acute gastroenteritis (27). Overall, 193 clusters were identified, with 10% of the municipalities involved in at least one cluster and less than 2% in several (27). To improve the detection of waterborne outbreaks, Coly et al. (24) developed an integrated approach to detect any study

clusters of acute gastrointestinal infection in geographical areas with a homogeneous exposure to drinking water. They used data from the French SyS system, geographical and population data, and environmental data based and the application of a space-time detection method identified 11 potential waterborne disease outbreaks. The outbreaks identified were not investigated, but the risk factors of exposure were examined (24).

### **Method evaluations via simulations**

Three of the included articles concerned simulations of SyS systems in the US. Burkom et al. (28) studied an integrated approach for the fusion of water quality data (e.g., faecal indicator bacteria, chlorine, pH, conductivity, and turbidity) with health monitoring data (ESSENCE) using probabilistic Bayesian networks. The simulations indicated a sensitivity of 80% and specificity of 99% for the symptoms “nausea/vomit” (28), however, further component simulations and the multidisciplinary development of realistic data scenarios would be needed (28). Xing et al. (32) compared timeliness of the SyS system using five regression models, and found that the sensitivity for ‘nausea and vomiting’ was calculated to approximately 55% (32). The simulations in the study of Xing et al (32) had a number of limitations, including a low number of data points.

Zhou et al. (31) examined the performance of the U.S. Centers for Disease Control and Prevention’s BioSense SyS system by injecting multi-day signals stochastically drawn from lognormal distributions into time series of aggregated daily visit counts for the outpatients at the Department of Veterans Affairs’ Veterans Health Administration (VHA) (31). The authors reported that, with a daily background alert rate of 1% and 2%, the sensitivities and timeliness in the SyS ranged from 24 to 77% and 3.3 to 6.1 days, respectively (31).

In the UK, two published studies presented measures to improve the method performance of national SyS systems. Noufaily et al. (46) reviewed the performance of aberrances among the weekly counts of isolates reported to the Health Protection Agency. By simulating different contrasting scenarios, the authors suggested several improvements related to the treatment of trends, seasonality, and the re-weighting of baselines. They claimed that the suggested system was able to reduce the number of false reports while retaining good power to detect genuine outbreaks. However, no explicit results regarding sensitivity and specificity related to detection of waterborne outbreaks were reported in the article (46).

Colón-Gonzales et al. (30) investigated how the characteristics of different outbreaks affected outbreak detection and the utility of SyS in detecting outbreaks using modelling and probability/statistics for two possible scenarios, including a localized outbreak of cryptosporidiosis. The authors reported that small gastrointestinal outbreaks (e.g., cryptosporidiosis) were unlikely to be detected unless the number of cases was over 1,000, with the detection of waterborne outbreaks varying by season (30). Multiple data streams (e.g., emergency attendance) are an advantage of influenza detection but not for outbreaks of cryptosporidiosis. However, the proposed framework of Colón-Gonzales et al. (2018) could, according to the authors, be applicable for the evaluation of any SyS system (30).

Mouly et al. 2018 (29) evaluated the performance of an algorithm using the French SyS system for waterborne outbreak detection through a simulation-based study using multivariate regression to identify the factors associated with outbreak detection. Almost three-quarters of the simulated outbreak were detected (sensitivity of 73%), and more than 9 out of the 10 detected signals corresponded to a waterborne outbreak (positive predictive value of 90.5%). The probability of was found to increase with the outbreak size (29).

## **Risk of bias and cumulative body of evidence**

The risk of bias of the included studies was overall assessed to be moderate to serious (Additional File 1). Due to the heterogeneity of the articles included, the cumulative body of evidence was partly assessed using the PRECEPT framework. The evidence was graded as high due to the low risk of publication bias.

## **Discussion**

In this systematic review, we identified 10 articles assessing the detection of waterborne outbreaks using different syndromic surveillance systems and six articles simulating a detection using a variation of statistical methods for the system performance improvements. The articles originated from four countries and represented five systems.

### **Effectiveness of SyS systems in detecting waterborne outbreaks**

The results reported in the included articles are generally modest (sensitivity below 50%) in their ability to detect waterborne outbreaks regardless of data signals. However, the simulation studies included in this review imply that multiple sources of signals combined with spatial information may increase the sensitivity in the SyS system of detecting waterborne outbreaks and reduce false alarms. The effectiveness of a SyS is a balance between sensitivity, specificity and predictive value, and timeliness, implying that high sensitivity may lead to a less timely detection (9). Because surveillance systems vary widely in terms of methodology, scope, and objectives, the characteristics that are important to one system may be less important to another. Efforts to improve certain attributes, such as the ability of a system to detect a health event (sensitivity), may detract from other attributes, such as simplicity or timeliness (10).

The use of over-the-counter pharmacy sales have been reported as not useful to detect waterborne outbreaks (26), while others have reported its usefulness (8). Drug sales data analysis for the outbreak detection of infectious diseases was reviewed in 2014 by Pivette et al (8), with the conclusion that over-the-counter sales appear to be a useful tool in detection trends gastrointestinal disease (8); however, the review may have been prone to publication bias. Only a few studies have shown promising correlations between SyS and signals, such as that origin from contact for health consultations in the health care system (22, 47). Such conflicting reporting of results should not be surprising, since the SyS systems included in the review are context-specific and not directly comparable. Although this review provides an updated overview of published articles assessing the effectiveness of SyS in detecting waterborne outbreaks, the synthesis of the articles was challenging, since they varied greatly in terms of administrative and geographical context, the data signals and algorithms used, and how the results were reported. The effectiveness of SyS system, in general, also largely relies on the methods used to detect aberrations.

The timeliness of surveillance approaches for outbreak detection is the amount of time from exposure to the disease agent to the initiation of a public health intervention (10). Berger et al. scored environmental data in terms of timeliness from a range of typical data used for SyS (12). However, when observing a change in for instance environmental data that may affect public health, disease in the population is likely to have been developed. In drinking water supply systems, there is an obligation of the water supplier to prompt action to mitigate a breach exceedance in the monitoring of microbiological parameters. Often, the mitigating action is

the issuance of a boil water advisory to protect the population from a potentially evolving outbreak (48). The risk of developing a waterborne outbreak is higher when a contamination event goes undetected during day-to-day-operations and routine monitoring, which is a common factor in several waterborne outbreaks (49, 50). In some of the identified articles included in this review, the benefit of combining the surveillance system to geographical supply zones to increase the likelihood of detecting a waterborne outbreak was highlighted. On the other hand, deploying such systems may be challenging, since it will most likely involve two different fields of expertise (health and technical), and the processing of data to inform health decisions must still be accounted for, since local outbreaks are usually short-lived (51).

The association between gastrointestinal illness cases and water quality data, such as turbidity, has been reviewed by de Roos et al. (52) in an attempt to discern the presence of waterborne gastrointestinal illness. However, the utility of turbidity as a proxy for microbiological contamination may be context-specific (52). Several of the excluded articles (Additional File 1) examined the potential of strengthening surveillance by including water quality data. In particular, these included, turbidity, disturbances in the distribution network, and calls to an alarm centre, among others. However, none of the excluded studies reported on analysis linked to real or simulated outbreaks. One of the most reported causes of waterborne outbreaks is heavy rainfall, which represents a future increasing risk (53) and implies a greater call for a risk-based approach to surveillance for water supply systems (51). In general, since there will always be a risk of water contamination going undetected, prioritizing long-term preventive measures and risk-based surveillance should not be underestimated despite promising reporting on SyS systems.

## **Strengths and limitations**

There are several limitations related to our review. First, the rather wide scope of the review resulted in a variety of articles that may have been of interest to the study topic but were excluded due to a lack of eligibility. Still, the included articles were also different from each other in many ways and did not allow for an accurate comparison of the reported results and assessments of the risk of bias. We also only found articles from four countries representing five surveillance systems, which limited the possibility of generalizing the results in terms of effectiveness to detect waterborne outbreaks.

In general, observational and retrospective studies are more prone to bias than randomized controlled studies due to a lack of randomization and blinding, hence jeopardizing their external and internal validity, which also affected the general outcome of the review. The ROBINS-I tool used in this review could only be partly used for the assessment of the risk of bias. We assessed risk of bias on a more-or-less hypothetically manner of the studies since developing the mimic RCT according to Sterne et al. (15), was challenging. Studies examining the detection of waterborne outbreaks based on real investigated outbreaks generally were assessed as having a lower bias due to confounding than those only using data on water quality deviations as a risk factor for waterborne illness. Bias in the selection of participants was, in general, a problem in the observational studies. In this review, all the studies using register data were assessed as having a lower risk of bias have been rated as having a moderate or serious risk of bias.. Moreover, bias in the classification of interventions was different among the studies examining outbreaks as serious due to the risk of differential misclassification (recall bias), while studies using register data with confirmed aetiology had a decreased risk of classification bias; however, there was a lack of evidence on illness attributed to drinking water. Bias in the domains of deviation from the intended interventions, missing data, and the measurement of outcomes were regarded as not applicable to

this review. Bias in the selection of the reported results was assessed as serious in cases in which only one data signal was studied.

A strength of this review is its comprehensive search of published peer-reviewed articles using multiple databases, the screening of bibliographies, and a random search of the topic of SyS systems' effectiveness in detecting waterborne outbreaks. The screening of bibliographies is a 'snow-balling' technique that entails a targeted assessment of the topic.

## **Conclusion**

Waterborne outbreaks still represent a risk in developed countries, and their early detection is crucial for the prevention of societal consequences. SyS systems with different features are widely used for the detection of waterborne outbreaks; however, in this review, we did not find evidence for syndromic surveillance in terms of their ability to effectively detect waterborne outbreaks (low sensitivity and high specificity), especially small and localized outbreaks. There are, on the other hand, promising development towards surveillance systems combining health, geographic, and water quality data; however, such systems must be evaluated in a cost-benefit context. This review demonstrates that there is no conclusive evidence regarding the effectiveness of SyS for the detection of waterborne outbreaks, which also emphasizes the need to focus on primary prevention measures to reduce the risk of waterborne outbreaks and risk-based surveillance. Future studies should include methods for combining health and environmental data with an assessment of the resources required for operating such a system.

## **Declarations**

### **Author contributions**

SH and PAA screened the publications identified in the literature search. EA screened the reference lists in the identified articles. SH conducted the random search for other studies. SH and EA assessed the identified studies from the three searches for eligibility. SH drafted the first version of the manuscript. EA, PAA, LV, and KN critically reviewed the drafted manuscript and approved its final version.

### **Acknowledgement**

We thank research librarian, Miriam Bakkeli, from NIPH for developing and conducting the literature search.

### **Ethics approval and consent to participate**

Not applicable

### **Consent for publication**

Not applicable

### **Availability of data and materials**

Additional file 1 with literature search and excluded articles is made available in Additional file 1.

## Competing interests

Not applicable

## Funding

Not applicable

## References

1. Mac Kenzie WR, Hoxie NJ, Proctor ME, Gradus MS, Blair KA, Peterson DE, et al. A massive outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply. *New England journal of medicine*. 1994;331(3):161–7.
2. Triple S Project. Assessment of syndromic surveillance in Europe. *Lancet*. 2011;378.
3. Nygard K, Schimmer B, Sobstad O, Walde A, Tveit I, Langeland N, et al. A large community outbreak of waterborne giardiasis-delayed detection in a non-endemic urban area. *BMC Public Health*. 2006;6:141.
4. Paterson BJ, Durrheim DN. The remarkable adaptability of syndromic surveillance to meet public health needs. *J Epidemiol Glob Health*. 2013;3(1):41–7.
5. Groeneveld GH, Dalhuijsen A, Kara-Zaïtri C, Hamilton B, de Waal MW, van Dissel JT, et al. ICARES: a real-time automated detection tool for clusters of infectious diseases in the Netherlands. *BMC Infectious Diseases*. 2017;17:201.
6. Andersson T, Bjelkmar P, Hulth A, Lindh J, Stenmark S, Widerstrom M. Syndromic surveillance for local outbreak detection and awareness: evaluating outbreak signals of acute gastroenteritis in telephone triage, web-based queries and over-the-counter pharmacy sales. *Epidemiology & Infection*. 2014;142.
7. Shortridge JE, Guikema SD. Public health and pipe breaks in water distribution systems: Analysis with internet search volume as a proxy. *Water Res*. 2014;53:26–34.
8. Pivette M, Mueller JE, Crépey P, Bar-Hen A. Drug sales data analysis for outbreak detection of infectious diseases: a systematic literature review. *BMC Infectious Diseases*. 2014;14(1):604.
9. European Center for Disease Control (ECDC). Data quality monitoring and surveillance system evaluation - a handbook of methods and evaluation. ECDC; [Accessed 15 October 2020]. Retrieved from: <https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/Data-quality-monitoring-surveillance-system-evaluation-Sept-2014.pdf>.
10. Buehler J, Hopkins R, Overhage J, Sosin D, Van Tong D. Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks: Recommendations from the CDC Working Group. *Morbidity and Mortality Weekly Report: Recommendations and Reports*. 2004;53(RR-5):1–13.
11. World Health Organization. Communicable disease surveillance and response systems. [Accessed 25 October 2020]. Retrieved from: [https://www.who.int/csr/resources/publications/surveillance/WHO\\_CDS\\_EPR\\_LYO\\_2006\\_2.pdf](https://www.who.int/csr/resources/publications/surveillance/WHO_CDS_EPR_LYO_2006_2.pdf).
12. Berger M, Shiao R, Weintraub JM. Review of syndromic surveillance: implications for waterborne disease detection. *Journal of Epidemiology and Community Health*. 2006;60(6):543–50.

13. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid AJSR. Rayyan—a web and mobile app for systematic reviews. 2016;5(1):210.
14. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *The BMJ*. 2009;339:b2535.
15. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. 2016;355:i4919.
16. Harder T, Takla A, Eckmanns T, Ellis S, Forland F, James R, et al. PRECEPT: an evidence assessment framework for infectious disease epidemiology, prevention and control. *Eurosurveillance*. 2017;22(40):16–00620.
17. Harder T, Abu Sin M, Bosch-Capblanch X, Bruno C, de Carvalho Gomes H, Duclos P, et al. Towards a framework for evaluating and grading evidence in public health. *Health Policy*. 2015;119(6):732–6.
18. Andersson T, Bjelkmar P, Hulth A, Lindh J, Stenmark S, Widerstrom M. Syndromic surveillance for local outbreak detection and awareness: evaluating outbreak signals of acute gastroenteritis in telephone triage, web-based queries and over-the-counter pharmacy sales. *Epidemiology and Infection*. 2014;142(2):303–13.
19. Bounoure F, Beaudeau P, Mouly D, Skiba M, Lahiani-Skiba M. Syndromic surveillance of acute gastroenteritis based on drug consumption. *Epidemiology & Infection*. 2011;139(9):1388–95.
20. Mouly D, Van Cauteren D, Vincent N, Vaissiere E, Beaudeau P, Ducrot C, et al. Description of two waterborne disease outbreaks in France: a comparative study with data from cohort studies and from health administrative databases. *Epidemiology and Infection*. 2016;144(3):591–601.
21. Cooper DL, Verlander NQ, Smith GE, Charlett A, Gerard E, Willocks L, et al. Can syndromic surveillance data detect local outbreaks of communicable disease? A model using a historical cryptosporidiosis outbreak. *Epidemiology & Infection*. 2006;134(1):13–20.
22. Bjelkmar P, Hansen A, Schönning C, Bergström J, Löfdahl M, Lebbad M, et al. Early outbreak detection by linking health advice line calls to water distribution areas retrospectively demonstrated in a large waterborne outbreak of cryptosporidiosis in Sweden. *BMC Public Health*. 2017;17(1):328.
23. Smith S, Elliot AJ, Mallaghan C, Modha D, Hippisley-Cox J, Large S, et al. Value of syndromic surveillance in monitoring a focal waterborne outbreak due to an unusual *Cryptosporidium* genotype in Northamptonshire, United Kingdom, June–July 2008. *Eurosurveillance*. 2010;15(33):9.
24. Coly S, Vincent N, Vaissiere E, Charras-Garrido M, Gallay A, Ducrot C, et al. Waterborne disease outbreak detection: an integrated approach using health administrative databases. *J Water Health*. 2017.
25. Ziemann A, Rosenkötter N, Garcia-Castrillo Riesgo L, Schrell S, Kauh B, Vergeiner G, et al. A concept for routine emergency-care data-based syndromic surveillance in Europe. *Epidemiology and Infection*. 2014;142(11):2433–46.
26. Kirian ML, Weintraub JM. Is syndromic surveillance of OTC drug sales effective in detecting outbreaks of waterborne GI disease? *Journal American Water Works Association*. 2011;103(8):63–73.
27. Rambaud L, Galey C, Beaudeau P. Automated detection of case clusters of waterborne acute gastroenteritis from health insurance data - pilot study in three French districts. *J Water Health*. 2016;14(2):306–16.
28. Burkom HS, Ramac-Thomas L, Babin S, Holtry R, Mnatsakanyan Z, Yund C. An integrated approach for fusion of environmental and human health data for disease surveillance. *Statistics in Medicine*.

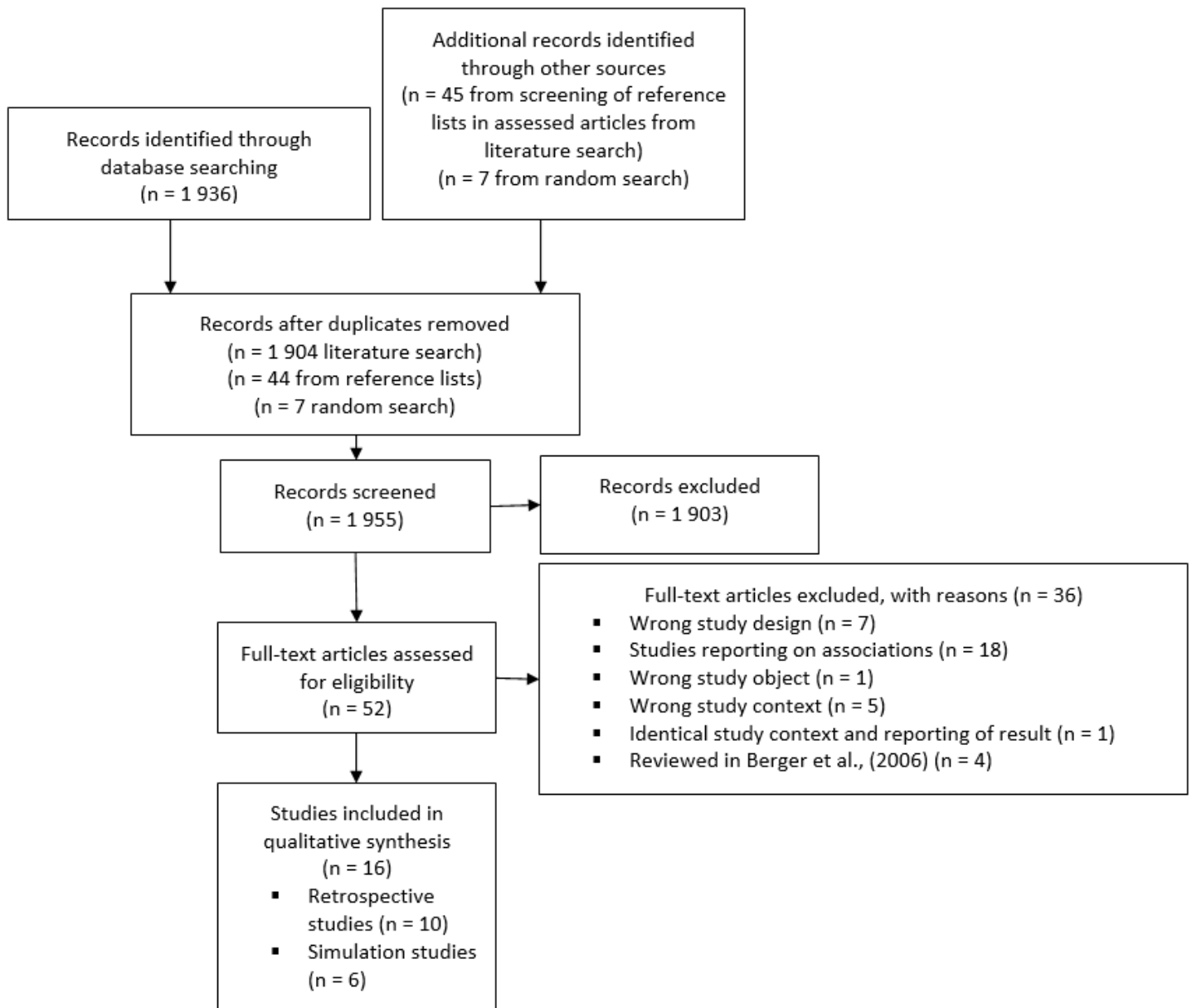


- 2011;30(5):470–9.
29. Mouly D, Gorla S, Mounié M, Beaudeau P, Galey C, Gallay A, et al. Waterborne Disease Outbreak Detection: A Simulation-Based Study. *International Journal of Environmental Research and Public Health*. 2018;15(7):1505.
  30. Colón-González FJ, Lake IR, Morbey RA, Elliot AJ, Pebody R, Smith GE. A methodological framework for the evaluation of syndromic surveillance systems: a case study of England. *BMC Public Health*. 2018;18(1):544.
  31. Zhou H, Burkom H, Winston CA, Dey A, Ajani U. Practical comparison of aberration detection algorithms for biosurveillance systems. *J Biomed Inform*. 2015;57:446–55.
  32. Xing J, Burkom H, Tokars J. Method selection and adaptation for distributed monitoring of infectious diseases for syndromic surveillance. *J Biomed Inform*. 2011;44(6):1093–101.
  33. Beaudeau P, Payment P, Bourderont D, Mansotte F, Boudhabay O, Laubies B, et al. A time series study of anti-diarrheal drug sales and tap-water quality. *International Journal of Environmental Health Research*. 1999;9(4):293–311.
  34. Beaudeau P, Le Tertre A, Zeghnoun A, Zanobetti A, Schwartz J. A time series study of drug sales and turbidity of tap water in Le Havre, France. *Journal of Water & Health*. 2012;10(2):221–35.
  35. Egorov AI, Naumova EN, Tereschenko AA, Kislitsin AA, Kislitsin VA, Ford TE. Daily variations in effluent water turbidity and diarrhoeal illness in a Russian city. *International Journal of Environmental Health Research*. 2003;13(1):81–94.
  36. Drayna P, McLellan SL, Simpson P, Li SH, Gorelick MH. Association between rainfall and pediatric emergency department visits for acute gastrointestinal illness. *Environmental Health Perspectives*. 2010;118(10):1439.
  37. Morris RD, Naumova EN, Levin R, Munasinghe RL. Temporal variation in drinking water turbidity and diagnosed gastroenteritis in Milwaukee. *American Journal of Public Health*. 1996;86(2):237–9.
  38. Tinker SC, Moe CL, Klein M, Flanders WD, Uber J, Amirtharajah A, et al. Drinking water turbidity and emergency department visits for gastrointestinal illness in Atlanta, 1993–2004. *Journal of Exposure Science & Environmental Epidemiology*. 2010;20(1):19–28.
  39. Hsieh JL, Nguyen TQ, Matte T, Ito K. Drinking water turbidity and emergency department visits for gastrointestinal illness in New York City, 2002–2009. *PLoS ONE [Electronic Resource]*. 2015;10(4):e0125071.
  40. Malm A, Axelsson G, Barregard L, Ljungqvist J, Forsberg B, Bergstedt O, et al. The association of drinking water treatment and distribution network disturbances with Health Call Centre contacts for gastrointestinal illness symptoms. *Water Res*. 2013;47(13):4474–84.
  41. Hunter PR, Chalmers RM, Hughes S, Syed Q. Self-Reported Diarrhea in a Control Group: A Strong Association with Reporting of Low-Pressure Events in Tap Water. *Clinical Infectious Diseases*. 2005;40(4):e32-e4.
  42. Derby M, McNally J, Ranger-Moore J, Hulette L, Villar R, Hysong T, et al. Poison Control Center–Based Syndromic Surveillance for Foodborne Illness. *MMWR*. 2005;2005(54).
  43. Curriero FC, Patz JA, Rose JB, Lele S. The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. *American Journal of Public Health*. 2001;91(8):1194–

9.

44. Conway M, Dowling JN, Chapman WW. Using chief complaints for syndromic surveillance: a review of chief complaint based classifiers in North America. *J Biomed Inform.* 2013;46(4):734–43.
45. Buckeridge DL. Outbreak detection through automated surveillance: A review of the determinants of detection. *Journal of Biomedical Informatics.* 2007;40(4):370–9.
46. Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. An improved algorithm for outbreak detection in multiple surveillance systems. 2013;32(7):1206–22.
47. Smith S, Elliot AJ, Mallaghan C, Modha D, Hippisley-Cox J, Large S, et al. Value of syndromic surveillance in monitoring a focal waterborne outbreak due to an unusual *Cryptosporidium* genotype in Northamptonshire, United Kingdom, June - July 2008. *Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin.* 2010;15(33):19643.
48. World Health Organization. Guidelines for drinking-water quality. 4th ed. [Accessed 1 November 2020]. Retrieved from:  
[http://www.who.int/water\\_sanitation\\_health/publications/2011/dwq\\_chapters/en/index.html](http://www.who.int/water_sanitation_health/publications/2011/dwq_chapters/en/index.html).
49. Rizak S, Hrudey S. Evidence of Water Quality Monitoring Limitations for Outbreak Detection. *Environmental health.* 2007;7(1):11–21.
50. Rizak S, Hrudey S. Achieving safe drinking water – risk management based on experience and reality. *Environmental Reviews.* 2007;15(NA):169–74.
51. Hyllestad S, Iversen A, MacDonald E, Amato E, Borge B, Bøe A, et al. Large waterborne *Campylobacter* outbreak: use of multiple approaches to investigate contamination of the drinking water supply system, Norway, June 2019. *Eurosurveillance.* 2020;25(35):2000011.
52. De Roos AJ, Gurian PL, Robinson LF, Rai A, Zakeri I, Kondo MC. Review of Epidemiological Studies of Drinking-Water Turbidity in Relation to Acute Gastrointestinal Illness. *Environmental Health Perspectives.* 2017;125(8):086003.
53. Hedlund C, Blomstedt Y, Schumann B. Association of climatic factors with infectious diseases in the Arctic and subarctic region – a systematic review. *Global Health Action.* 2014;7:10.3402/gha.v7.24161.

## Figures



**Figure 1**

Flow of studies identified and screened in the review

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)