Public Health Reports
Public Health Reports
2017, Vol. 132(Supplement 1) 116S-126S

DOI: 10.1177/0033354917709784
journals.sagepub.com/home/phr
SAGE

A Practitioner-Driven Research Agenda for Syndromic Surveillance

Richard S. Hopkins, MD, MSPH[1], Catherine C. Tong, MPH[2], Howard S. Burkom, PhD[3], Judy E. Akkina, PhD, MPH[4], John Berezowski, DVM, PhD[5], Mika Shigematsu, MD, PhD, MSc[6,7], Patrick D. Finley, MS[8], Ian Painter, PhD, MSc[9,10], Roland Gamache, PhD, MBA[11,12], Victor J. Del Rio Vilas, DVM, PhD, MBA, MSc[13], and Laura C. Streichert, PhD, MPH[2]

Abstract

Syndromic surveillance has expanded since 2001 in both scope and geographic reach and has benefited from research studies adapted from numerous disciplines. The practice of syndromic surveillance continues to evolve rapidly. The International Society for Disease Surveillance solicited input from its global surveillance network on key research questions, with the goal of improving syndromic surveillance practice. A workgroup of syndromic surveillance subject matter experts was convened from February to June 2016 to review and categorize the proposed topics. The workgroup identified 12 topic areas in 4 syndromic surveillance categories: informatics, analytics, systems research, and communications. This article details the context of each topic and its implications for public health. This research agenda can help catalyze the research that public health practitioners identified as most important.

Keywords

research agenda, syndromic surveillance

During the past 15 years, syndromic surveillance has evolved from a set of ad hoc methods used mostly in postdisaster settings[1-4] to a mature technology that runs continuously to detect and monitor a range of health issues. At each stage in their development, syndromic surveillance systems have been guided by results of applied and basic research, either published or presented at annual conferences.[5,6]

Syndromic surveillance has several distinguishing characteristics: (1) the primary purpose of the activity is to identify community patterns of disease rather than individual illness; (2) records typically do not contain personal identifiers; (3) unlike public health case reporting, no selection of records is requested of the sender; and (4) for outbreak and event detection, practitioners prioritize timeliness and sensitivity over positive predictive value. Evidence on the value of syndromic surveillance approaches for outbreak or event detection is limited. Syndromic surveillance systems have, however, also shown value in documenting and characterizing already known outbreaks or events.[7,8]

In many countries, the dominant approach to syndromic surveillance is based on categorizing emergency department (ED) or primary care visits into syndromes and looking for aberrations or patterns in a jurisdiction's daily count of syndromes. In Europe, the development of a consensus in

[1] Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, University of Florida, Gainesville, FL, USA
[2] International Society for Disease Surveillance, Braintree, MA, USA
[3] Applied Physics Laboratory, Johns Hopkins University, Laurel, MD, USA
[4] Center for Epidemiology and Animal Health, Veterinary Services, Animal and Plant Health Inspection Service, US Department of Agriculture, Fort Collins, CO, USA
[5] Veterinary Public Health Institute, University of Bern, Bern, Switzerland
[6] International Biological and Chemical Threat Reduction Program, Sandia National Laboratories, Albuquerque, NM, USA
[7] National Institute of Infectious Diseases, Tokyo, Japan
[8] Department of Operations Research and Computational Analysis, Sandia National Laboratories, Albuquerque, NM, USA
[9] Department of Health Services, School of Public Health, University of Washington, Seattle, WA, USA
[10] Gamache Consulting, Rockville, MD, USA
[11] School of Veterinary Medicine, University of Surrey, Kent, UK
[12] Center for Population Health Information Technology, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[13] School of Veterinary Medicine, University of Surrey, Guildford, UK

Corresponding Author:
Richard S. Hopkins, MD, MSPH, Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, University of Florida, 2004 Mowry Rd, Gainesville, FL 32610, USA.
Email: hopkinsrs@comcast.net

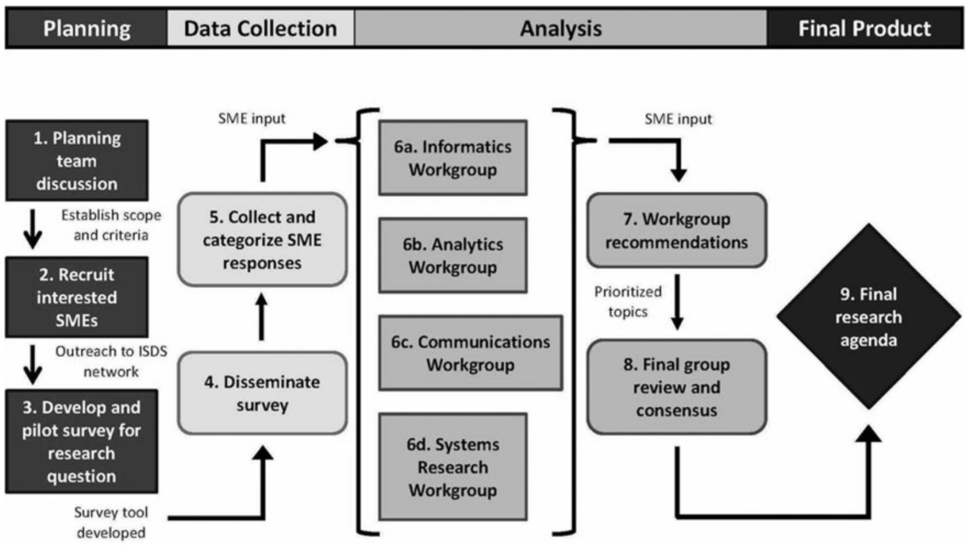

Figure. The syndromic surveillance research agenda development process used by a task force of the International Society for Disease Surveillance (ISDS), 2016. Abbreviation: SME, subject matter expert.

syndromic surveillance practice along these lines was facilitated by the Triple-S project.[9]

Current challenges and opportunities in syndromic surveillance include assessing and incorporating novel data streams[10]; using it to assess and monitor acute population health events after detection; using syndromic surveillance methods for real-time understanding of community patterns of injuries, poisonings, chronic diseases and their complications, and unmet medical care needs; and improving the integration of human, domestic animal, wildlife, and ecologic data in a One Health surveillance view.

Previous collaborative groups have published research agendas and evaluation recommendations for surveillance systems. They have addressed surveillance evaluation in general,[11] evaluation applied to syndromic surveillance for event detection,[12] syndromic surveillance research,[13] public health informatics,[14] and population health informatics research.[15] In 2013, Smith et al updated an earlier broad, high-level surveillance blueprint for the 21st century.[16,17] Our report highlights 12 discrete, global syndromic surveillance research needs as seen by syndromic surveillance practitioners, taking into account advances in public health surveillance since 2010.[13]

Methods

A workgroup of the International Society for Disease Surveillance (ISDS) Research Committee, consisting of national and global subject matter experts in syndromic surveillance and ISDS professional staff members, conducted this project. The workgroup made decisions through iterative review, ranking, and consensus among workgroup leads in 4 topical domains aligned with syndromic surveillance practice: informatics, analytics, systems research, and communications.

To generate topics for consideration, the workgroup developed, piloted, and distributed a web-based SurveyMonkey questionnaire to stakeholders in health surveillance that drew from network lists of the ISDS and partner organizations. This distribution strategy primarily targeted practitioners of syndromic surveillance and researchers focused on syndromic surveillance. An invitation to participate was sent to >3000 email addresses and was also posted on the ISDS website.

Respondents included workers with primary expertise in epidemiology, statistics, data management, and informatics and with a primary focus on practice, research, or both. Respondents were in general interested in the effective use of syndromic surveillance data as the result of their efforts. Respondents were asked to submit their priority research questions, taking into account their own assessment of whether each question met $\geq$1 of the following 4 criteria:

Actionable: Does it inform public health action?
Informed: Does it make use of current knowledge, data, and technology?
Informative: Does it contribute any additional knowledge? Is there any value added?
Feasible: Can it be conducted with reasonable cost and time (eg, does a data source exist?)

We received 90 responses from surveillance professionals working in 15 countries. Three rounds of review and synthesis of the information received were used to create the final research agenda (Figure). Stakeholder responses informed research topics. Workgroup members synthesized, expanded, and combined the responses, using their own specialized knowledge, to generate the topics described in this report. Eighteen topics were identified in the first round of topic generation. Through an iterative prioritization and refinement process, some topics were discarded and others were merged to create an eventual list of 12 topics. Workgroup members produced an extensive set of references to identify the current state of knowledge related to these topics and promising avenues of research.

The recommended topics are grouped first by the 4 domains (analytics, informatics, systems research, and communication) and then by priority order assigned by the workgroup within the domain. The 12 topics presented here were all considered important, but the first topic was the most frequently mentioned by participating ISDS members and was ranked highest by the workgroup (Figure).

Recommendations

Analytics

1. Methods and systems to support the fusion of various types of data. The most common need identified by respondents to our call for topics was the need to improve methods to integrate multiple types of evidence. Health monitors seek corroboration of signals across data sources, which would provide more confidence in investigation decisions than can be provided with single data streams.

Standard statistical data fusion methods (eg, multivariate control charts,[18] multivariate space-time clustering,[19] and dynamic linear models[20]) have been supplemented by Bayesian networks[21] and sophisticated Bayesian statistical models.[22-24] Fusion methods from other domains may require substantial research to be transferrable, practical, and useful in syndromic surveillance. Syndromic surveillance will need to combine human and animal health data with environmental data streams, especially as the need to monitor the effects of climate change, including expanding zoonoses, becomes more urgent.

Even when independent multivariate data streams have the same spatial and temporal resolution and equal epidemiologic relevance, identifying anomalies is a considerable challenge.[25] The background covariance among the streams may change, indications from different sources may be contradictory, and alerts may arise from subtle combinations of effects across sources. These situations require careful validation and visualization to make results understandable to public health surveillance monitors. Spatial resolutions may vary because of privacy concerns, slowly changing legacy systems, or available sensing technology. Monitoring inputs at varying acquisition rates (eg, combining streaming and report-based evidence with sensor data) requires further advanced techniques.

Recent work on multimodal recurrent neural networks appears to have potential in helping to integrate unstructured data (as from social media) with more traditional data. Particle filter and Kalman filter approaches from the physical sciences may help achieve the needed integration for routine syndromic surveillance workflows. Fusion methods should exhibit high throughput and resolution without excessive system resource requirements and should exhibit ease of use, stability, readily interpretable results, and integration into existing syndromic surveillance applications.

2. Methods to adapt syndromic surveillance approaches and systems to changing needs. Syndromic surveillance can be used to monitor newly recognized diseases, chronic diseases and their complications, or environmental conditions and their impact. Additional research is needed on how to design systems that can be easily adapted to changing needs.[26-28]

Syndromic surveillance can supply near–real-time data for monitoring injuries, chronic diseases and their complications, or health inequities. Practitioners will require analytics in addition to aberration detection methods. Incorporating other data streams, such as census data, weather data, or crime information, may increase the value of syndromic surveillance data. Improved functionality requires research that addresses system architecture, interoperability, and security.

Disasters can result in sudden population movements and disruptions to health care infrastructure. Flexible syndromic surveillance methods are needed in such settings to monitor impact and response needs.[29] Cellular telephone services, or at least SMS (short message service) texting, are often consistently available and have been useful for collecting data.[30,31] Further research is needed, however, on how best to improve interoperability and balance functionality with end-user needs in settings with limited infrastructure.[32]

3. Enhanced and adaptive detection algorithms. Syndromic surveillance practice depends on automated statistical algorithms to detect unusual deviations from the expected occurrence of health events. Algorithms have been adapted from classical hypothesis testing, control charts, time-series analysis, and multiple types of regression analysis.[33] Machine-learning and natural-language processing have been incorporated to enrich these algorithms for application to complex data sources.[34]

Implicit or explicit steps underlie each algorithm: preconditioning of input data (normalizing transformations and adjustments for systematic behavior), background or baseline estimation, test statistic formation, selection of the alerting threshold, and communication of alert status. These steps are subjects of ongoing research toward more efficient automated monitoring.[35] Priority research needs that the workgroup identified included (1) methods for data preconditioning to deal with variable late reporting issues[36]; (2) adjustable, dynamic thresholds incorporating past experience and external knowledge; and (3) methods to identify disease clusters without bias resulting from syndromic classification or jurisdictional boundaries.[37]

System users wish to adjust for nonstatistical knowledge and past experience in generating or interpreting alerts. Dynamic and adjustable thresholds are needed to incorporate evidence such as known threats or knowledge of events in neighboring regions. In 2016, most automated systems employed frequentist statistical methods. Bayesian statistical methods could include informative prior probabilities and posteriors reflecting computational experience and borrow strength for realistic thresholds in sparse data situations.[38,39] Improvements in processing power and computational shortcuts enable the application of such methods. Useful development requires epidemiologic expertise. Machine-learning approaches for consideration include Bayesian networks and decision trees, with or without user mediation.

4. Methods and processes for monitoring and addressing data quality issues. Basic steps to ensure the quality of streaming data include de-duplication, correct formatting and field placement, and awareness of gaps in receipt of expected data. Insufficient attention to these steps has led to false alarms and can degrade sensitivity to true signals. Dashboards (ie, graphic displays of multiple parameters) have been developed to summarize data completeness,[40] but they have not been uniformly implemented across jurisdictions.

Quality issues result from several tensions inherent in the use of streaming data for prospective analysis. Damage to data quality may be an accepted cost of decisions made to optimize other aspects of the system, such as (1) transmitting data promptly rather than performing accuracy checks, (2) making analytic adjustments despite known data quality issues, accepting these issues in anticipation of data improvements, (3) standardizing data to minimal records to expedite processing rather than waiting for more complete data, and (4) insisting on uniformity in submission from distributed agencies that have individual coding and collection practices rather than adjusting data after ingestion. Choices among these alternatives may affect the determination of whether an apparent event is sufficiently anomalous to warrant further investigation. Quality issues may distort both baseline and current estimates.

Knowing that baselines are immature for lack of historic data, temporarily incomplete, or discontinuous often restricts the choice of analytic methods. Issues may result from problems with connectivity or from changes in data provider participation, coding, and data representation (eg, elimination of free-text fields). These changes may affect all or only some participating distributed data sources.

Analysts have attempted to adjust for late reporting, which always has the strongest effect on the most recent data. If denominator surrogates such as the total number of encounters are available, regression models can adjust current estimates.[36,41-44] Models have also been applied to correct for long-term baseline trends.[45]

The most straightforward solution to data quality problems is assurance of up-to-date, uniform information from providers who remain invested in the surveillance process. Because obtaining such uniform information is not always possible, developing analytic methods to correct or account for data quality issues is a high research priority.

5. Processes to develop and assess syndrome definitions. The workgroup identified 2 key needs related to syndrome definitions: (1) validation of a wider range of definitions and (2) availability of tools to support dynamic, custom, and impromptu syndrome formation.

A syndromic surveillance system tracks counts of records received that have common characteristics indicative of outcomes of interest. These categories, or syndromes, may refer to sets of related symptoms or findings, laboratory tests for certain conditions, web searches containing certain terms, or billing records for a class of remedies. The syndromes used by a syndromic surveillance system depend on the richness of the data, the number of outcomes of concern, and the resources available for investigation and response.

Sensitivity and positive predictive value of a syndrome definition for cases or outbreaks of a particular outcome are strongly influenced by syndrome classification.[46] If the definition is too narrow, signals of interest may be missed; if too broad, true signals may be masked by false alarms.[26]

Most published studies on the accuracy and value of syndromic surveillance data have addressed systems that use chief complaints and have not used diagnosis codes, laboratory test results, assessment by human experts to assist with validation, or text indicating exposures available in chief complaint or triage note fields.[47-49] Few studies have addressed target diseases other than influenza and gastroenteritis. Standardization and validation of syndrome definitions for numerous outcomes of interest are needed for true cross-jurisdictional collaboration.[50]

Natural-language processing may be able to improve the classification of free-text data such as chief complaints, triage notes,[51,52] and data from social media sources.[53] Such studies can be costly, and true outbreak data are rarely available for validation. Several machine-learning approaches have been applied to optimize classification.[54-56] Conclusive validation is challenging, and these studies have yielded mixed results.

Tools are also desired to allow rapid user-generated construction, sharing, and comparison of ad hoc syndrome definitions in dynamic situations, using whatever data fields are available in the syndromic surveillance records. The value of approaches to building these tools should be assessed.

6. Predictive analytic models using surveillance data. Disease forecasting through predictive analytic models promises to improve public health planning and response by providing reliable estimates of the future extent and location of outbreaks. Trusted forecasting has the potential to adjust alert levels and improve the timeliness of outbreak detection. Recent high-profile outbreaks of disease caused by Ebola, Zika, dengue, and influenza viruses have shown the importance of accurate forecasts for effective logistic and countermeasure planning.[57] Increasingly available human and animal health data (including syndromic surveillance data), social media data, environmental measurements, and microbiological and genomic data are used to seed computational disease-propagation models along with relevant data on population demographic characteristics.

Recent prediction efforts focused on influenza incidence in major US cities[58] and dengue in southeast Asia,[59] using data assimilation techniques such as traditional time series, particle filters, and ensemble Kalman filters. For example, Buczak et al applied association rule classification methods to generate 1-month dengue forecasts for the Philippines.[60] Forecasting methods have also been applied to noninfectious diseases—for example, 1-week forecasting of asthma admissions in Greece modeled using artificial neural networks.[61] Improvements in the accuracy and availability of microbiological and genetic data are enabling predictions by merging rapidly advancing technologies.[62] Disease forecasting holds the potential for improving outbreak response by providing early predictions of incidence, hospitalization, and other outcomes that will allow decision makers to marshal resources and personnel in a timely manner.[63] Future improvements in data quality and variety, modeling methods, and validation are needed to improve the reliability and spatial resolution of model-derived forecasts.[64]

Informatics

7. Methods to process, categorize, and code unstructured data in electronic health records. Unstructured narrative text requires categorization for syndromic surveillance. Research is needed to develop automated processes to improve the validity, efficiency, and cost of these methods. English-language vocabulary sets such as the Unified Medical Language System[65] should be mapped to public health concepts for more efficient concept recognition.[66] Flexible processes for categorizing unstructured data will facilitate surveillance in unique situations and for emerging public health issues.

Current syndromic surveillance systems focus on keyword searches of chief complaints. However, the usefulness of this approach is limited by the changing language used by health care providers. For example, novel terms ("tailgate party," "flakka") may indicate a foodborne outbreak or a new illicit substance.[67] Methods for complex parsing of negated concepts are needed to support coding validity.[68] Negated terms such as "denies fever and cough" need to be recognized and handled appropriately. Processes for coding narrative text are needed for languages other than English, to improve comparability of syndromic surveillance data when different languages are in use. Text in various languages may also reflect different conceptualizations of symptoms, syndromes, and diseases.

8. Methods to assess the added value of new data sources and data elements. Adding new data sources or data elements to a syndromic surveillance system has associated costs. These costs are incurred to develop the capability to acquire and work with data from a new source; recruit, implement, and maintain each data provider; and process, monitor, and respond to the new data.[69] Similar tasks are completed when new data elements are added to existing data streams. Costs are also incurred in training staff members when new data are made available. Methods are needed to determine the added value of new sources and elements compared with their cost.

The utility of some novel data streams in detecting outbreaks or supporting situational awareness has been established.[70-81] Syndromic surveillance managers must consider value and cost in deciding whether to include these and other novel data streams. Acceptance of the cost and complexity of incorporating such data types requires quantifying their incremental value.

In resource-limited settings, system limitations on the collection of additional data of sufficient quality[82,83] may affect the ability to conduct surveillance. However, implementation of lean, flexible systems may be easier when there are no fixed costs in the existing infrastructure. Collection of high-quality data has been achieved on a site-specific basis[84] by targeting resources and training, which is more feasible with sentinel sites.

To monitor intervention programs in situations in which sentinel sites are the only source of high-quality data, methods are needed to decide on the most appropriate sentinel site placement. Better methods are also needed to use sentinel site data to estimate population-level trends.

Systems Research

9. Standardized methods to validate and evaluate syndromic surveillance systems. Syndromic surveillance system evaluation should be a scientific discipline with standardized methods and replicable and transferable findings. Documents from the Centers for Disease Control and Prevention have identified key attributes for surveillance system evaluation, including for syndromic surveillance,[11,12] but standard methods for measuring the performance of working systems on those attributes are lacking. Such methods could be used to compare system performance levels with one another and with external standards.

Methods are needed to estimate negative predictive value: how small an event can be reliably detected? Similarly, standardized methods are needed to measure and assess the usefulness or value of a syndromic surveillance system for situational awareness as well as early detection.

For event detection, systems need to be evaluated against standard data sets with known, labeled events and against data sets in which events have been injected into authentic background data. Simulation approaches should be adopted, standardized, and implemented at the population level (eg, by injected counts) and at the individual level (eg, with synthetic records generated by agent-based behavior and transmission models that produce records corresponding to the model outputs).

Research is needed to better understand the representativeness of ED-based syndromic surveillance compared with actual population-level acute illness experience and how it varies by period, geographic region, or demographic group.[85,86] Incomplete understanding of representativeness and sensitivity has limited the use of census data as denominators in estimating rates.

Standard methods are also needed to assess qualitative attributes, such as usefulness and acceptability, and to calculate the costs of designing, building, implementing, and operating systems.

10. Decision support methods for public health surveillance. Technical and organizational challenges converge in decisions about public health surveillance systems. In this situation, rules of thumb and experience are not sufficient to steer strategic and operational decisions. Decision support mechanisms are therefore required to address analytic and organizational complexities arising from the use of syndromic surveillance approaches.

Decision support mechanisms have roles in public health surveillance planning, development, functioning, and evaluation. They can be used to decide on the best sample and data logistics and the best resource structure to maintain overall public health surveillance capability.

Comprehensive decision support mechanisms have been developed in other areas, such as environmental sciences,[87] and are the subject of ongoing research and development for public health surveillance. Given constrained resources, limited evidence, and prevailing risks, such comprehensive methods can inform the best surveillance alternatives in a portfolio fashion and systematically address biases.

In addition to assessing the decision quality contributed by the decision support mechanism via forward-looking metrics that track value delivery, other attributes can be monitored to gauge the acceptability of the decision support mechanism and, ultimately, its institutionalization and alignment with public health surveillance objectives. Desirable qualities to monitor during decision support mechanism development and implementation would include (1) a clear description of the assumptions, uncertainties, constraints, and values that inform the decision support mechanism; (2) an engaging interface with users; (3) a module-based and scalable architecture; (4) flexibility to allow specification of varying contexts; (5) key stakeholder engagement during the planning stage; (6) efficient processes for evidence gathering and functioning of the decision support mechanism; (7) a repository of decisions and associated outcomes to support future analyses and validation of decision support mechanism outputs; and (8) consideration of synergies and interoperability with other current and planned decision support mechanisms, as part of an overarching strategy to promote flexible and adapting architectures.

Syndromic surveillance presents optimization problems because of the need to prioritize objectives and multiple constraints on human and technology resources. Resource allocation efforts have included optimized reporting networks[88] and optimal data provider recruitment.[89] Few formal sampling strategies for detection have been published.[90] A lack of research on meeting needs for efficient use of scarce resources is attributable to limited public health funding and to varying surveillance environments. The combination of evolving public health threats and increasingly complex data sources without concomitant growth in the epidemiologic workforce highlights the need for this research.

Communications

11. Communication methods, processes, and tools to inform decision making. To meet the needs of decision makers, syndromic surveillance communication must go beyond simple dissemination of alerts, which may be misinterpreted. Inclusion of contextual information will help to ensure that decision makers fully understand the meaning of the communication. Decision makers also need information about the properties and performance of the surveillance system. Research is needed to more clearly define the current and emerging information needs of decision makers at different levels of public health. However, engaging decision makers in processes aimed at defining their needs can be a challenge for surveillance practitioners for many reasons, including differences in the concepts and terminology of the 2 groups and differences in understanding surveillance systems and information. If common ground can be found, the benefits will be substantial because it will allow syndromic surveillance practitioners to modify their syndromic surveillance systems to meet the emerging information needs of decision makers.

Participatory research approaches[91-93] should be used to engage syndromic surveillance practitioners, communication experts, and information users in a continuous process in which user needs are constantly redefined and used to improve syndromic surveillance information production and communication. Research is needed to identify potential new syndromic surveillance users, explore their needs, and modify syndromic surveillance communication to meet them. Research aimed at overcoming language, ethnic, professional, institutional, and other social barriers will be needed.

Communicating syndromic surveillance alerts directly to the public could be very useful, especially during times of crisis. The risk of misinterpretation is substantial, however, and sociological research is needed to understand the consequences of direct public syndromic surveillance alert communication.

Routinely used communication networks will be familiar to stakeholders. In times of crisis, these networks will need to be quickly expanded to include new, less familiar stakeholders. Communication networks should also be studied to determine their resiliency and responsiveness during times of crisis and to identify the networks needed for different types of crises.

12. Effective strategies for building workforce competency in syndromic surveillance practice. Conducting syndromic surveillance, a highly technical process, has not yet been fully optimized to effectively and efficiently meet user needs.

Box. Top 12 areas of research needed to advance the practice of syndromic surveillance, as identified by a workgroup comprising International Society for Disease Surveillance members, February to June 2016. Priority areas are not ranked by relative importance to public health practice. Topics inherently overlap across syndromic surveillance domains.

Priority Areas in Syndromic Surveillance in Need of Research to Enhance Public Health Practice

Analytics

1. Methods and systems to support the fusion of various types of data.
2. Methods to adapt syndromic surveillance approaches and systems to changing needs.
3. Enhanced and adaptive detection algorithms.
4. Methods and process for monitoring and addressing data quality issues.
5. Processes to develop and assess syndrome definitions.
6. Predictive analytic models using surveillance data.

Informatics

7. Methods to process, categorize, and code unstructured data in electronic health records.
8. Methods to assess the added value of new data sources and data elements.

Systems research

9. Standardized methods to validate and evaluate syndromic surveillance systems.
10. Decision support methods for public health surveillance.

Communications

11. Communication methods, processes, and tools to inform decision making.
12. Effective strategies for building workforce competency in syndromic surveillance practice.

Research is needed to explore the full value for a variety of users of information that could potentially be produced by syndromic surveillance and the most efficient ways to deliver it.

Researchers who specialize in information system application design should evaluate syndromic surveillance applications and optimize them for specific user groups. They could, for example, identify underused system components and determine why they are underused or evaluate keystroke patterns used most frequently, with the aim of reducing keystrokes and improving user performance. Better training materials and methods are also needed. These training materials and methods could include the development of simulation training pilots and stand-alone training modules that help users develop core competencies needed to operate syndromic surveillance systems. They should be developed to meet the needs of practitioners in disciplines including human, animal, and environmental health (Box).

Limitations

This study had several limitations. First, most proposed research topics were received from practitioners and researchers in the United States and other developed countries. Thus, the recommendations may not fully reflect the research needs for syndromic surveillance in resource-limited settings. Some of the recommended research topics (Nos. 2, 3, 8, 11, and 12) are more important than the other research topics in resource-limited settings.

Second, we solicited research topics from syndromic surveillance practitioners and researchers who are active in or members of ISDS. We did not solicit topics from employees in public health and emergency response who are likely to be consumers of syndromic surveillance products but are not affiliated with ISDS. Their perspectives may not be explicitly represented. The utility or value of syndromic surveillance data to decision makers was nonetheless addressed in most research topics. The workgroup assumed that modifications to syndromic surveillance systems that increase the representativeness, accuracy, richness, timeliness, sensitivity, and/or positive predictive value of data or activities that improve the skills of practitioners or the quality of presentations of data could only improve the data's usefulness to decision makers for event characterization and event detection.

Last, we requested research topics related to syndromic surveillance rather than to the early detection of outbreaks, situational awareness, or other purposes, which may have limited the diversity of topics submitted.

Public Health Practice Implications

In the past 15 years, the development of syndromic surveillance methods into an established part of the surveillance portfolio of many public health agencies has been an iterative process involving researchers and practitioners in many disciplines. Our compilation of research topics, initially identified by practitioners and refined by a group of subject matter experts in the field of syndromic surveillance, can be used by academic researchers to identify questions of direct importance to public health practice. It can also be used by funding agencies to help set priorities for surveillance-related research funding. Finally, it can be used by practitioners to guide their interactions with research partners and their funding agencies.

The topics identified in this process overlap with those identified by Uscher-Pines et al in their 2010 syndromic surveillance research agenda. They summarized their top 3 questions as follows: "(1) How should different types of evidence and complementary data systems be integrated (merging data, visualizations)? (2) How can syndromic surveillance best be used in an electronic medical record environment? and (3) What criteria should be used to prioritize alerts?"[13] All 3 questions are still recognizable among the 12 priorities we identified.

Our report places less emphasis on prioritizing alerts and on the alerting function in general than the 2010 research agenda. Since 2010, it has become increasingly clear that public health practitioners value syndromic surveillance methods for day-to-day situational awareness beyond early warning. Among these benefits are characterizing an event (often learned about in other ways) by time, place, and person; following the trajectory of the event over time; and detecting likely individual cases of diseases of concern for further investigation. Practitioners are also increasingly interested in uses of syndromic surveillance methods to understand injury and chronic disease issues in their communities in close to real time, without having to wait for hospital discharge data, complete mortality data files, and population-based survey data to become available. Other benefits include rumor control, corroboration or rule-out of clinical suspicions, and understanding the burden of severe weather and other catastrophic events. Practitioners would like to see the development and validation of methods and tools that will improve their ability to use these data in such ways.

The first priority question from Uscher-Pines et al corresponds to the top-mentioned research priority in the current process: methods and systems to support the fusion of various types of data.[13] It became clear during the deliberations of our expert group that although several existing or emerging methods show great promise for presenting a fusion of data from multiple sources to the end user, a gap between the research (and its results) and syndromic surveillance practice remains. Practitioners are generally interested in research that results in tools that can be put to use in existing syndromic surveillance systems without excessive wait time, that will not require additional processing power, and that will produce readily understandable and actionable displays of surveillance information for practitioners and decision makers.

Looking ahead, exciting opportunities exist to arm public health practitioners with the data and information they need for decision making using near–real-time syndromic surveillance approaches. The steps that we identified included expanding health topics addressed by syndromic surveillance methods; integrating information from nonhuman sources with information from humans; using syndromic and other data to make predictions about the course of outbreaks or other public health events; implementing flexible, inexpensive syndromic surveillance in resource-limited settings; making more effective use of electronic health records in richer countries; strengthening uses of syndromic surveillance to address all types of hazards; making improvements in validated training methods; and generally building and expanding on existing relationships between researchers and practitioners to further research with direct public health relevance.

Acknowledgments

We thank Julia E. Gunn, RN, MPH, and Leah Estberg, DVM, MPVM, PhD, for their contributions to the project. Julia E. Gunn also served as guest editor for this special issue, "Syndromic Surveillance: The Value of Real-time Data for Public Health Action" in *Public Health Reports*.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project was funded by the Centers for Disease Control and Prevention through cooperative agreement U38OT000143 with the Council of State and Territorial Epidemiologists.

References

1. Glass RI, O'Hare P, Conrad JL. Health consequences of the snow disaster in Massachusetts, February 6, 1978. *Am J Public Health*. 1979;69(10):1047-1049.
2. Centers for Disease Control and Prevention. Epidemiologic notes and reports: hurricanes and hospital emergency-room visits—Mississippi, Rhode Island, Connecticut. *MMWR Morb Mortal Wkly Rep*. 1986;34(51-52):765-770.
3. Lee LE, Fonseca V, Brett KM, et al. Active morbidity surveillance after Hurricane Andrew—Florida, 1992. *JAMA*. 1993; 270(5):591-594.
4. Pavlin JA, Mostashari F, Kortepeter MG, et al. Innovative surveillance methods for rapid detection of disease outbreaks and bioterrorism: results of an interagency workshop on health indicator surveillance. *Am J Public Health*. 2003; 93(8):1230-1235.
5. Bioterrorism—syndromic surveillance. Proceedings of the 2002 National Syndromic Surveillance Conference. New York, USA, September 23-24, 2002. *J Urban Health*. 2003; 80(2)(suppl 1):i1-i140.
6. International Society for Disease Surveillance. Past ISDS conferences. http://www.syndromic.org/resources/past-isds-conferences. Accessed April 21, 2017.
7. Centers for Disease Control and Prevention. National Syndromic Surveillance Program: syndromic surveillance data in action. https://www.cdc.gov/nssp/biosense/action.html. Accessed April 21, 2017.
8. Chu A, Savage R, Willison D, et al. The use of syndromic surveillance for decision-making during the H1N1 pandemic: a qualitative study. *BMC Public Health*. 2012;12:929.
9. Triple-S. Guidelines for designing and implementing a syndromic surveillance system. https://pdfs.semanticscholar.org/3

6c9/2e2bb071a1cd4e8ce50a6431c028fd06dd7a.pdf. Accessed April 11, 2017.
10. Third International Conference on Digital Disease Detection. About HealthMap. http://www.healthmap.org/ddd. Accessed June 6, 2016.
11. Centers for Disease Control and Prevention. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm Rep*. 2001;50(RR-13):1-35.
12. Centers for Disease Control and Prevention. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR Recomm Rep*. 2004;53(RR-5):1-11.
13. Uscher-Pines L, Babin SM, Farrell CL, et al. Research priorities for syndromic surveillance systems response: consensus development using nominal group technique. *J Public Health Manag Pract*. 2010;16(6):529-534.
14. Massoudi BL, Goodman KW, Gotham IJ, et al. An informatics agenda for public health: summarized recommendations from the 2011 AMIA PHI Conference. *J Am Med Inform Assoc*. 2012;19(5):688-695.
15. Kharrazi H, Lasser EC, Yasnoff WA, et al. A proposed national research and development agenda for population health informatics: summary recommendations from a National Expert Workshop. *J Am Med Inform Assoc*. 2017;24(1):2-12.
16. Smith PF, Hadler JL, Stanbury M, Rolfs RT, Hopkins RS; CSTE Surveillance Strategy Group. "Blueprint version 2.0": updating public health surveillance for the 21st century. *J Public Health Manag Pract*. 2013;19(3):231-239.
17. Meriwether RA. Blueprint for a national public health surveillance system for the 21st century. *J Public Health Manag Pract*. 1996;2(4):16-23.
18. Fricker RD Jr. Directionally sensitive multivariate statistical process control methods with application to syndromic surveillance. *Adv Dis Serv*. 2007;3(1):1-17.
19. Kulldorff M, Mostashari F, Duczmal L, Katherine Yih W, Kleinman K, Platt R. Multivariate scan statistics for disease surveillance. *Stat Med*. 2007;26(8):1824-1833.
20. Lau EH, Cheng CK, Ip DK, Cowling BJ. Situational awareness of influenza activity based on multiple streams of surveillance data using multivariate dynamic linear model. *Plos One*. 2012; 7(5):e38346.
21. 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. *Stat Med*. 2011;30(5):470-479.
22. Corberán-Vallet A. Prospective surveillance of multivariate spatial disease data. *Stat Methods Med Res*. 2012;21(5): 457-477.
23. Paul M, Held L, Toschke AM. Multivariate modelling of infectious disease surveillance data. *Stat Med*. 2008;27(29): 6250-6267.
24. Schiöler L, Frisén M. Multivariate outbreak detection. *J Appl Stat*. 2012;39(2):223-242.
25. Andersson E, Kühlmann-Berenzon S, Linde A, Schiöler L, Rubinova S, Frisén M. Predictions by early indicators of the time and height of the peaks of yearly influenza outbreaks in Sweden. *Scand J Public Health*. 2008;36(5):475-482.
26. Chapman WW, Dowling JN, Baer A, et al. Developing syndrome definitions based on consensus and current use. *J Am Med Inform Assoc*. 2010;17(5):595-601.
27. Lurio J, Morrison FP, Pichardo M, et al. Using electronic health record alerts to provide public health situational awareness to clinicians. *J Am Med Inform Assoc*. 2010;17(2): 217-219.
28. Gamache R, Stevens KC, Merriwether R, Dixon BE, Grannis S. Development and assessment of a public health alert delivered through a community health information exchange. *Online J Public Health Inform*. 2010;2(2):ojphi.v2i2.3214.
29. Sulzberger AG, Stelter B. A rush to protect patients, then bloody chaos. *New York Times*. http://www.nytimes.com/2011/05/24/us/24tornado.html. Published 2011. Accessed May 17, 2016.
30. Fischer HH, Moore SL, Ginosar D, et al. Care by cell phone: text messaging for chronic disease management. *Am J Manag Care*. 2012;18(2):e42-e47.
31. Chretien JP, Burkom HS, Sedyaningsih ER, et al. Syndromic surveillance: adapting innovations to developing settings. *PLoS Med*. 2008;5(3):e72.
32. Lewis SL, Feighner BH, Loschen WA, et al. SAGES: a suite of freely-available software tools for electronic disease surveillance in resource-limited settings. *PLoS One*. 2011;6(5):e19750.
33. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J R Stat Soc Serv A Stat Soc*. 2012;175(1):49-82.
34. Chanlekha H, Kawazoe A, Collier N. A framework for enhancing spatial and temporal granularity in report-based health surveillance systems. *BMC Med Inform Decis Mak*. 2010;10:1.
35. Blazes DL, Lewis SH. *Disease Surveillance: Technological Contributions to Global Health Security*. New York, NY: CRC Press; 2016.
36. Painter I, Eaton J, Olson D, Revere D, Lober B. Generation of prediction intervals to assess data quality in the distribute system using quantile regression. *JSM Proc*. 2011:5172-5179.
37. Liu Y, Neill DB. Detecting previously unseen outbreaks with novel symptom patterns. *Emerg Health Threats J*. 2011;4:11074.
38. Lawson AB. Spatial and spatio-temporal disease analysis. In: Lawson AB, Kleinman K, eds. *Spatial and Syndromic Surveillance for Public Health*. New York, NY: John Wiley & Sons Ltd; 2005:53-76.
39. Lawson AB. Advanced modeling for surveillance: clustering of relative risk changes. In: Lawson AB, Kleinman K, eds. *Spatial and Syndromic Surveillance for Public Health*. New York, NY: John Wiley & Sons Ltd; 2005:223-243.
40. Cheng CK, Ip DK, Cowling BJ, Ho LM, Leung GM, Lau EH. Digital dashboard design using multiple data streams for disease surveillance with influenza surveillance as an example. *J Med Internet Res*. 2011;13(4):e85.
41. Burkom HS, Elbert Y, Feldman A, Lin J. Role of data aggregation in biosurveillance detection strategies with applications from ESSENCE. *MMWR Suppl*. 2004;53:67-73.

42. Craigmile PF, Kim N, Fernandez SA, Bonsu BK. Modeling and detection of respiratory-related outbreak signatures. *BMC Med Inform Decis Mak*. 2007;7:28.
43. Tokars JI, Burkom H, Xing J, et al. Enhancing time-series detection algorithms for automated biosurveillance. *Emerg Infect Dis*. 2009;15(4):533-539.
44. Zhou H, Burkom H, Winston CA, Dey A, Ajani U. Practical comparison of aberration detection algorithms for biosurveillance systems. *J Biomed Inform*. 2015;57:446-455.
45. Levin-Rector A, Wilson EL, Fine AD, Greene SK. Refining historical limits method to improve disease cluster detection, New York City, New York, USA. *Emerg Infect Dis*. 2015; 21(2):265-272.
46. Reis BY, Mandl KD. Syndromic surveillance: the effects of syndrome grouping on model accuracy and outbreak detection. *Ann Emerg Med*. 2004;44(3):235-241.
47. Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang H, Pavlin JA. Code-based syndromic surveillance for influenza-like illness by *International Classification of Diseases*, *Ninth Revision. Emerg Infect Dis*. 2007;13(2):207-216.
48. Ivanov O, Wagner MM, Chapman WW, Olszewski RT. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. *Proc AMIA Symp*. 2002:345-349.
49. Cadieux G, Buckeridge DL, Jacques A, Libman M, Dendukuri N, Tamblyn R. Patient, physician, encounter, and billing characteristics predict the accuracy of syndromic surveillance case definitions. *BMC Public Health*. 2012;12:166.
50. Burkom H, Olson D, Baer A, Paladini M, Revere D, McIntyre A. Analytic evaluation of a standardization effort for the DISTRIBUTE emergency department surveillance project. In: *Proceedings of the Joint Statistical Meetings*. Alexandria, VA: American Statistical Association; 2011:1450-1472.
51. Chapman WW, Christensen LM, Wagner MM, et al. Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artif Intell Med*. 2005;33(1): 31-40.
52. Kawazoe A, Chanlekha H, Shigematsu M, Collier N. Structuring an event ontology for disease outbreak detection. *BMC Bioinformatics*. 2008;9(suppl 3):1-8.
53. Alvaro N, Conway M, Doan S, Lofi C, Overington J, Collier N. Crowdsourcing Twitter annotations to identify first-hand experiences of prescription drug use. *J Biomed Inform*. 2015; 58:280-287.
54. Magruder SF, Lewis SH, Najmi A, Florio E. Progress in understanding and using over-the-counter pharmaceuticals for syndromic surveillance. *MMWR Suppl*. 2004;53:117-122.
55. Wallstrom GL, Hogan WR. Unsupervised clustering of over-the-counter healthcare products into product categories. *J Biomed Inform*. 2007;40(6):642-648.
56. Dórea FC, Muckle CA, Kelton D, et al. Exploratory analysis of methods for automated classification of laboratory test orders into syndromic groups in veterinary medicine. *PLoS One*. 2013;8(3):e57334.
57. Morin CW, Comrie AC, Ernst K. Climate and dengue transmission: evidence and implications. *Environ Health Perspect*. 2013;121(11-12):1264-1272.
58. Shaman J, Karspeck A. Forecasting seasonal outbreaks of influenza. *Proc Natl Acad Sci U S A*. 2012;109(50): 20425-20430.
59. Althouse BM, Ng YY, Cummings DA. Prediction of dengue incidence using search query surveillance. *PLoS Negl Trop Dis*. 2011;5(8):e1258.
60. Buczak AL, Baugher B, Babin SM, et al. Prediction of high incidence of dengue in the Philippines. *PLoS Negl Trop Dis*. 2014;8(4):e2771.
61. Moustris KP, Douros K, Nastos PT, et al. Seven-days-ahead forecasting of childhood asthma admissions using artificial neural networks in Athens, Greece. *Int J Environ Health Res*. 2012;22(2):93-104.
62. Janies D, Witter Z, Gibson C, Kraft T, Senturk IF, Çatalyürek Ü. Syndromic surveillance of infectious diseases meets molecular epidemiology in a workflow and phylogeographic application. *Stud Health Technol Inform*. 2015;216:766-770.
63. Chretien JP, Swedlow D, Eckstrand I, et al. Advancing epidemic prediction and forecasting: a new US government initiative. *Online J Public Health Inform*. 2015;7(1):e13.
64. Chretien JP, George D, Shaman J, Chitale RA, McKenzie FE. Influenza forecasting in human populations: a scoping review. *PLoS One*. 2014;9(4):e94130.
65. US National Library of Medicine. Unified Medical Language System (UMLS). https://www.nlm.nih.gov/research/umls. Accessed May 16, 2016.
66. Travers DA, Haas SW. Evaluation of emergency medical text processor, a system for cleaning chief complaint text data. *Acad Emerg Med*. 2004;11(11):1170-1176.
67. Faigen Z, Deyneka L, Ising A, et al. Cross-disciplinary consultancy to bridge public health technical needs and analytic developers: asyndromic surveillance use case. *Online J Public Health Inform*. 2015;7(3):e228.
68. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*. 2001; 34(5):301-310.
69. Ding Y, Sauerborn R, Xu B, et al. A cost-effectiveness analysis of three components of a syndromic surveillance system for the early warning of epidemics in rural China. *BMC Public Health*. 2015;15:1127.
70. Aslam AA, Tsou MH, Spitzberg BH, et al. The reliability of tweets as a supplementary method of seasonal influenza surveillance. *J Med Internet Res*. 2014;16(11):e250.
71. Velardi P, Stilo G, Tozzi AE, Gesualdo F. Twitter mining for fine-grained syndromic surveillance. *Artif Intell Med*. 2014; 61(3):153-163.
72. Tsou MH, Jung CT, Peddecord M. Situational awareness of health events using social media and the SMART dashboard. *Online J Public Health Inform*. 2016;8(1):e166.
73. Pfeiffer DU, Stevens KB. Spatial and temporal epidemiological analysis in the Big Data era. *Prev Vet Med*. 2015;122(1-2): 213-220.
74. Hickmann KS, Fairchild G, Priedhorsky R, et al. Forecasting the 2013-2014 influenza season using Wikipedia. *PLoS Comput Biol*. 2015;11(5):e1004239.

75. Generous N, Fairchild G, Deshpande A, de Valle SY, Priedhorsky R. Global disease monitoring and forecasting with Wikipedia. *PLoS Comput Biol*. 2014;10(11):e1003892.
76. McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS Comput Biol*. 2014;10(4):e1003581.
77. Moss R, Zarebski A, Dawson P, McCaw JM. Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data. *Influenza Respir Other Viruses*. 2016;10(4):314-323.
78. Thompson LH, Malik MT, Gumel A, Strome T, Mahmud SM. Emergency department and "Google flu trends" data as syndromic surveillance indicators for seasonal influenza. *Epidemiol Infect*. 2014;142(11):2397-2405.
79. Wiwanitkit V. Google Flu for forecasting influenza-like illness. *Am J Emerg Med*. 2014;32(11):1417.
80. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis*. 2009;49(10):1557-1564.
81. Samaras L, García-Barriocanal E, Sicilia MA. Syndromic surveillance models using web data: the case of scarlet fever in the UK. *Inform Health Soc Care*. 2012;37(2):106-124.
82. Ledikwe JH, Grignon J, Lebelonyane R, et al. Improving the quality of health information: a qualitative assessment of data management and reporting systems in Botswana. *Health Res Policy Syst*. 2014;12:7.
83. Rowe AK, Kachur SP, Yoon SS, Lynch M, Slutsker L, Steketee RW. Caution is required when using health facility-based data to evaluate the health impact of malaria control efforts in Africa. *Malar J*. 2009;8:209.
84. Mpimbaza A, Miles M, Sserwanga A, et al. Comparison of routine health management information system versus enhanced inpatient malaria surveillance for estimating the burden of malaria among children admitted to four hospitals in Uganda. *Am J Trop Med Hyg*. 2015;92(1):18-21.
85. Metzger KB, Hajat A, Crawford M, Mostashari F. How many illnesses does one emergency department visit represent? Using a population-based telephone survey to estimate the syndromic multiplier. *MMWR Suppl*. 2004;53:106-111.
86. Hopkins RS, Kite-Powell A, Goodin K, Hamilton JJ. The ratio of emergency department visits for ILI to seroprevalence of 2009 pandemic influenza A (H1N1) virus infection, Florida, 2009. *PLoS Curr*. 2014:6.
87. Convertino M, Valverde LJ Jr. Portfolio decision analysis framework for value-focused ecosystem management. *PLoS One*. 2013;8(6):e65056.
88. Fairchild G, Polgreen PM, Foster E, Rushton G, Segre AM. How many suffice? A computational framework for sizing sentinel surveillance networks. *Int J Health Geogr*. 2013;12:56.
89. Scarpino SV, Dimitrov NB, Meyers LA. Optimizing provider recruitment for influenza surveillance networks. *PLoS Comput Biol*. 2012;8(4):e1002472.
90. Ferguson JM, Langebrake JB, Cannataro VL, et al. Optimal sampling strategies for detecting zoonotic disease epidemics. *PLoS Comput Biol*. 2014;10(6):e1003668.
91. Jost CC, Mariner JC, Roeder PL, Sawitri E, Macgregor-Skinner GJ. Participatory epidemiology in disease surveillance and research. *Rev Sci Tech*. 2007;26(3):537-549.
92. Wójcik OP, Brownstein JS, Chunara R, Johansson MA. Public health for the people: participatory infectious disease surveillance in the digital age. *Emerg Themes Epidemiol*. 2014;11:7.
93. Zinsstag J, Schelling E, Waltner-Toews D, Tanner M. From "one medicine" to "one health" and systemic approaches to health and well-being. *Prev Vet Med*. 2011;101(3-4):148-156.