General review

From outbreak detection to anticipation

De la détection à l’anticipation des épidémies

G. Texier a,*, Y. Buisson b

a Département d’épidémiologie et de santé publique, service de santé des armées, institut de médecine tropicale, BP 60109, allée du Médecin-Colonel-Jamot, Parc du Pharo, 13262 Marseille cedex 07, France
b Institut de la francophonie pour la médecine tropicale, Vientiane, Laos

Received 4 June 2009; accepted 9 June 2010
Available online 25 November 2010

Abstract

Background. – Bioterrorist threats and the emergence of new infectious pathogens force structures and means dedicated to outbreaks detection to evolve. Obtaining early information is becoming a major stake. Development of early warning system and new epidemiological tools open new perspectives in real time management of outbreaks.

Methods. – Through several examples, this overview presents a synthesis and reports the recent multidisciplinary evolutions, which occurred in the field of outbreak detection and information useful for outbreaks anticipation.

Conclusion. – Outbreak detection is a multidisciplinary activity which relies on many protean epidemiologic sensors. Detection should not be perceived as a finality, but as a crucial step of a broader process in outbreak management (detection, alarm, alert, acceptability, implementation of countermeasures or not). That is why the evaluation of a monitoring system should not be limited to only its ability to detect an aberration. Today, progress in biology, modeling, remote sensing and environmental detection allow model development with anticipation aiming.

© 2010 Elsevier Masson SAS. All rights reserved.

Keywords: Outbreak; Detection; Anticipation; Disease surveillance; Early warning

Résumé

Position du problème. – La menace bioterroriste et l’émergence de nouveaux agents infectieux ont obligé à faire évoluer les structures et les moyens modernes dévolus à la détection des épidémies. Obtenir des informations précoces est devenu un enjeu majeur. L’utilisation de systèmes d’alerte et l’apparition de nouveaux outils dans ce champ de l’épidémiologie ouvrent de nouvelles voies pour la prise en compte en temps réel des phénomènes épidémiques.

Méthodes. – À partir de plusieurs exemples, cette revue générale a pour objectif de faire un état des lieux de la détection des épidémies et d’envisager les sources d’information et les outils qui permettraient d’anticiper l’apparition des phénomènes épidémiques avant leur extension dans la population humaine.

Conclusion. – La détection des épidémies est une activité multidisciplinaire qui fait appel à de nombreux senseurs épidémiologiques protéiformes. La détection ne doit pas être perçue comme une finalité, mais comme une étape essentielle d’un processus plus large relevant de la gestion des épidémies (détection, alarme, alerte, acceptabilité, mise en place ou non de contre-mesures). C’est pourquoi l’évaluation d’un système de surveillance ne doit pas se limiter à sa seule capacité à détecter une aberration. Les progrès de la biologie, de la modélisation, de l’imagerie satellitaire et de la détection environnementale rendent aujourd’hui envisageable l’anticipation de ces phénomènes épidémiques.

© 2010 Elsevier Masson SAS. Tous droits réservés.

Mots clés : Épidémie ; Détection ; Anticipation ; Surveillance épidémiologique ;Alerte précoce

1. Introduction

Long endured by humans as divine punishment, disease outbreaks have only begun to be understood since the end of the nineteenth century, within the context of the scientific
3.1. Defining the event under surveillance

The first and most important step is to define cases. Using standardized criteria to define possible cases, probable cases, and confirmed cases, the sensitivity of data collection can be favored to detect the outbreak and the specificity reinforced secondarily to confirm it. Thus, Lapeyssonnie’s aphorism—“Severe diarrhea followed by vomiting that kills adults in a few hours is nearly always cholera”—allows one to deduce the first cases of a cholera outbreak without waiting for the bacteriological identification of *Vibrio cholerae* O1. Case definition, a classical exercise in epidemiology, turns out to be more difficult with an unknown event. This situation, which is predictable in situations of bioterrorism or disease emergence, has justified the use of syndromic surveillance since 2001 [17].

In an attempt to ensure precocity and when the definition of the event is based on biological criteria, the actors involved in surveillance can call on rapid diagnostic tests. Immunochromatographic tests are used to demonstrate antigens or specific antibodies and molecular tests to recognize a preserved region of the pathogenic agent’s genome. Dipstick tests have a number of applications: malaria; streptococcal tonsillitis; bacterial meningitis; plague; cholera; shigelloses; diphtheria; flu; HIV, HBV, and HCV infections; leishmanioses; diarrhea caused by *Giardia*, *Cryptosporidium*, *Clostridium difficile*, adenovirus, rotavirus, etc. [18–20]. Among the molecular tests, the possibility of performing real-time PCR in the field opens a new avenue for rapid diagnosis. Whatever qualities they may possess, they cannot substitute for clinical examination or conventional diagnostic methods, which remain indispensable to making a definite diagnosis and to isolating the pathogenic agent in case of an infectious disease epidemic [21].

3.2. Defining the population under surveillance

An increase in the number of cases of a disease is only significant if related to a relevant denominator. The concept of outbreak assumes a reference population, i.e., a more or less large, homogenous, and defined group of individuals. Certain populations are more difficult to monitor because of their openness, the exchanges that they may have with other populations, and the mobility of its members (community gatherings, refugees, travelers, military personnel in operation, the homeless, etc.) [22].

3.3. Who is responsible for detecting outbreaks?

The continual process of surveillance is based on several structures that depend on the population under surveillance and the event being monitored.

At the national level [44], the French Institute for Public Health Surveillance (l’Institut de Veille Sanitaire, InVS) coordinates three large infectious disease surveillance systems: the mandatory declaration applied to 30 ailments, the National Reference Centers (Centres Nationaux de Reference, CNR), and the voluntary professional networks (RENARub for rubella infections during pregnancy, RENACOQ for pertussis,
EPIBAC for invasive meningococcus, pneumococcus, *Haemophilus*, and *Listeria* infections, ONERBA for bacterial resistance to antibiotics, etc.). Since 1984, a network of private practice physicians, the Sentinelles Network (le Réseau Sentinelles), has carried out national surveillance of several health events such as flu syndromes, male urethritis, measles, acute diarrheas, etc.

The infectious diseases that present a risk for public health are subject to mandatory declaration to the Public Health Agency (Agence Régionale de Santé, ARS) by the practitioner or biologist. The 30 diseases listed in decrees no. 2001-910 of 5 October 2001 and no. 2002-1089 of 7 August 2002, with the exception of HIV/AIDS infection, acute hepatitis B, and tetanus, must be declared immediately so as to prevent or limit their dissemination by prophylaxis [23]. Through its epidemiology and public health departments, the Military Health Service carries out similar work to the InVS, monitoring 64 epidemiological events in the military population in France, abroad, or at sea [24]. These actors are involved in several governmental programs (the Flu Pandemic, Smallpox, Piratox, Biotox programs, etc.) [25].

On the European level, The European Network for Surveillance, Warning, and Control of Infectious Diseases has been replaced by the European Center for Prevention and Disease Control (ECDC), a true European CDC inaugurated in Stockholm in 2004. On the global level, the WHO coordinates the Global Outbreak Alert and Response Network (GOARN).

4. Detection

4.1. Estimating the number of expected cases

A major problem at the stage of outbreak detection is the estimation of the number of expected cases. This can be zero for an eradicated (smallpox) or eliminated disease (poliomyelitis in Europe), but most often, the peak of an endemic disease’s incidence must be compared to its baseline or its relaxation level (mean level of the epidemiologic signal). This baseline is most often created by smoothing the data collected previously.

4.2. Choosing the threshold

To respond to the question of when the number of cases observed is higher than the number of cases expected, a threshold must be defined, based on: expert advice (e.g., meningitis in certain countries in the Sahel in Africa); construction of an interval around the result of smoothing the curve of cases (e.g., the WHO’s Cullen method, the Serfling method for flu, etc.) [26,27]; using the well-known laws of data distribution (Poisson, Gauss, etc.) or simulated distribution (permutation techniques, resampling, etc.) [28]. For daily surveillance systems, the threshold level can change from one day to the next. However, the different effects affecting the statistical analysis must be taken into account (seasonal effect, trend effect, week-end bias, etc.).

Seasonal flu is a good model because an outbreak is expected every year during the cold season in temperate countries without its intensity or date of appearance being known beforehand. The outbreak threshold of the GROG (Regional Groups of Flu Observation; Groupes Régionaux D’observation de la Grippe) network [29] is determined using two indicators: the number of acute respiratory infections (ARIs) and the number of short-duration work stoppages (WSs). The mean of the ARIs and WSs during the month of October becomes the reference value. The outbreak threshold is considered to be exceeded when the weekly number of ARIs and WSs goes beyond 120% of October’s mean. The alert is confirmed if several similar flu viruses are isolated in different zones of the same region and if the criteria for exceeding the threshold are all met for several consecutive weeks.

The choice of a threshold should also be adapted to the surveillance context so as to respond as best possible to the operational objectives, in particular when a vaccination campaign can be set up to stop the outbreak. In the example of meningococcus meningitis in the Sahel in Africa, the WHO recommended abandoning the threshold of 15 cases per 100,000 inhabitants per week despite its excellent specificity. This threshold was not sufficiently sensitive to detect all the outbreaks of meningitis and left only a 3-week delay between exceeding the threshold and the outbreak’s peak, insufficient for effective mass vaccination [30]. The new recommendation proposed two levels, defined in relation to the population size and the outbreak risk:

- an alarm threshold to launch an investigation and prepare for an outbreak;
- an outbreak threshold to confirm the outbreak and reinforce the control measures.

4.3. Detection methods

Even though expert advice remains pertinent in detecting an outbreak, the current tendency (because of the need for early warning and continued service) is to use statistical tools to provide automated data analysis [31]. Most authors working in real-time epidemiologic surveillance agree on the need for associating several techniques and algorithms. These techniques are used to create an expectation, a threshold, or both. More than 120 original techniques have been developed, but some are more often used for their statistical properties and their capacity for automatization:

- control charts stem from statistical methods used to follow quality in the industrial sector. They are used to detect any piece that does not conform to the norms defined by the manufacturer (beyond a certain tolerance interval). Most are based on the measurement of the mean and standard deviation. These graphs can be applied to the detection of outbreaks. The most widely used techniques today are the cumulated sum (CUSUM) and the exponentially weighted moving average (EWMA) [32,33];
- regression models are statistical tools that describe the dependence between a variable to explain Y (e.g., the number
of cases of malaria) and several explanatory variables Xi (e.g., the level of endemicity in villages). From a collection of qualitative or quantitative observations, the model is constructed based on historical data and then used with the current Xi data to provide an expected Y. This expected Y is then compared to the observed data. The most widely used models in the detection of outbreaks are multiple linear regressions, Poisson regressions, and logistical regressions with or without a mixed component [33–36];

- time series analysis models take into account the temporal relation that may exist between the data. Two broad categories exist. The first consider that the data have a time function \( Y = f(t) \), such as the Serfling method [27], developed in 1963, although it remains valid for detecting flu outbreaks. The second, called autoregressive models, seek to determine each value of the series in relation to the values that precede it \( (Y_t = f(Y_{t-1}, Y_{t-2}, \ldots)) \). The value Y to explain at time t depending on the values of Y taken the previous days \( (j - 1, j - 2, \ldots) \). Thus, it can be considered that the day's number of flu cases is not only related to the number the day before but also the number on the preceding days. Autoregressive integrated moving average (ARIMA) models are the most widely used [37,38]. They are also capable of forecasting at a given horizon;

- scan statistics are standards in outbreak detection [39]. They integrate the space and time dimension into their analysis. Basically, groups of cases are detected on a map by comparing the number of cases observed in a circle (of a varying radius) with what could be expected if the event followed a Poisson distribution law. The technique was improved in a cylinder-shaped model whose height represents time and the circular base the geographic zone of interest. Several variants have been developed, but they are all limited by the need to obtain georeferenced data (GPS coordinates, zip codes, etc.);

- bayesian networks are calculators of conditional probabilities. They take into account both the a priori knowledge (experts, previous publications, etc.) and the knowledge obtained a posteriori by data collection. The cause-effect relations between the variables are not deterministic (detection or nondetection of an outbreak) but probabilistic (percentage of chances that an outbreak is present). Thus, the observation of a cause or of several causes does not systematically lead to an effect, but only modifies the probability of observing one. The current tendency is to use dynamic bayesian networks [40] or hidden Markov chain models [41].

5. Managing detection

The result of the analysis during which an epidemiologic signal is identified (by statistical analysis, expert advice, etc.) as higher than what was expected is called an alarm. The objective of this alarm is to raise awareness (the situation awareness concept described in cognitive psychology) [42] of the actors in charge of surveillance and possibly set off conservative measures while awaiting the next stage. The human expertise stage that follows the alarm is designed to confirm the reality of the outbreak (by eliminating false-positives) to set off an alert. This stage is based on other data sources than those of the surveillance system (health surveillance data, epidemiologic survey, complementary analyses, diagnosis-assistance tools in syndromic surveillance, etc.). This is part of the situation diagnosis (Fig. 1) described by Chaudet et al. [43]. Confirmation of the outbreak therefore alerts the actors and health authorities [44]. These authorities must evaluate the risk and the impact of the outbreak before deciding whether to set up countermeasures.

6. Anticipate outbreaks

Early-warning systems classically use computerization and automation of surveillance processes to increase their reactivity. However, how early they react also depends on the choice of epidemiological captors or sensors (humans, animals, automated machines, software, etc.) and the data to be collected. This choice can directly influence the surveillance system’s capacity to anticipate outbreaks.

6.1. The detection continuum

A collection of data sources can be used to detect outbreaks. These sources can be distributed on a temporal axis depending on the time that separates them from the confirmed diagnosis. Along this continuum, each source can correspond to a detection process (Fig. 2). The most reliable sources, but the most delayed, are those that use humans: the healthcare structures encountering the first cases, physician practitioners, hospital emergency units, and medical analysis laboratories. With the risk of increasing false alarms, the earlier the data are collected on this continuum, the more sensitive and reactive the system is. Use of direct indicators on the health of the population under surveillance can also be envisaged (e.g., the number of hits on medical Internet sites, the number of calls to emergency services or poison centers, the use of thermal-image scanners in airports, etc. [45] or environmental sources.
6.2. Environmental surveillance

The environment (in its Latin sense “that which surrounds”) is the interface between the outbreak source and the population under surveillance. Monitoring the environment allows a certain level of anticipation by displacing the cursor before the medical observations of the population under surveillance. This requires completing the surveillances system by connecting it to data sensors on fauna and the physical environment (water, air, and climate) as well as associating it with a broadened capacity for health surveillance.

6.2.1. Veterinary surveillance and sentinel animals

Surveillance of zoonoses transmissible to humans illustrates these changes well. In France, West Nile fever, an arbovirus caused by a flavivirus, was under particular surveillance in the southwest region of the country. It is a complementary system associating detection of human cases of arbovirus by the national reference center, the detection of equine encephalitis cases by veterinary services, and the early detection of viral circulation in avifauna by searching for seroconversions in sentinel birds and identification of strains causing death in wild birds. In the United States, the most sensitive method for early detection of West Nile viral activity was surveillance of wild birds, before surveillance of sentinel birds, surveillance of mosquito vectors, the appearance of equine cases, and finally, the appearance of human cases [46].

The use of sentinel chicken flocks, relatively low cost in veterinary surveillance [47], allows one to identify the circulation of several viruses: West Nile (France), Western equine encephalitis and St. Louis encephalitis (California), and Murray Valley and Kunjin (Australia). Similarly, surveillance of wild monkeys can provide early detection of Ebola virus in equatorial Africa and amaril virus in Brazil.

6.2.2. Surveillance of the physical environment

6.2.2.1. The water environment. In 1993, a large outbreak of Cryptosporidium gastroenteritis hit Milwaukee [48], the largest city in Wisconsin in the United States, affecting 400,000 inhabitants and causing more than 60 deaths. Drinking water, the source of this outbreak, was nonetheless in accordance with the legislation relative to the microbiological indicators for potability.

In France, the distribution of drinking water is monitored using more than 60 parameters, 54 that stem from the legislation of the Health Ministry and in application of the European directives, requiring use of highly varied techniques: microbiological, genetic, physical and chemical dosages, the search for radioactive elements, acoustic, video inspection, flow measurements, etc. Other surveillance systems employ biosentinels, either fish that are highly sensitive to alterations in the environment (trout, bluegill) or invertebrates to measure the accumulation of heavy metals (Mussel Watch network). The US Army and the company IAC have developed a system called the Intelligent Aquatic Biomonitoring System (IABS) associating an aquarium containing eight fish (bluegill) and a computerized system capable of monitoring drinking water in real time. This system makes it possible to set off a sound, electronic, or telephone alarm while automatically closing the network’s valves. The IABS has been tested in several reservoirs feeding New York City [49].

Fig. 2. Outbreak detection continuum.
6.2.2.2. The air environment. Continual surveillance of the air can be carried out with optronic systems (combining optical and electronic systems). There are two classes: passive and active systems. The first receive rays emitted by the observed source, focusing the optical flow on a detector and converting it to an electrical signal, e.g., spectroradiometers or spectrophotometers used to detect nebulized or aerosolized chemical substances in the air. In the active systems, light rays are used (generally a laser emission) to illuminate a scene or a target. LIDAR (light detection and ranging) and its applications (ADEDIS and bioLIDAR) can remotely detect propagation of biological agents, pollens, pollutants, etc., by analyzing a laser light backscattered by atmospheric particles.

Other sensors such as biocollectors can be used to evaluate aerobiocntamination (e.g., aerosols of legionella) of a site [50]. Particles (viruses, bacteria, etc.) are separated from the air by impaction on a solid or liquid medium or by filtration (Fig. 3).

6.2.2.3. The climate and land environment. Remote sensing, providing images in different spectral bands of the Earth, has opened a new direction by providing useful environmental and meteorological data for predicting outbreaks (the concept of the WHO’s Early Warning System). In the malaria domain, this type of anticipation can reach several months by identifying the favorable conditions for survival and vectorial pullulation, from 1 to 2 months for the index related to the vegetation cover such as the Normalized Difference Vegetation Index (NDVI) [50–53], up to 6 months in the study of an oceanic climate phenomenon such as El Niño Southern Oscillation (ENSO) [54].

6.2.3. Medical intelligence or healthwatch

A health watch reflects the physician’s point of view on environmental surveillance. It uses the results of veterinary and environmental surveillance extended to health data concerning the populations not affected by the surveillance system. These data are collected from different media (newspapers, radio, Internet, health documents, etc.) or rumors. The healthwatch improves the sensitivity of the system in making the surveillance actors aware of the surrounding situation (situation awareness).

In outbreak detection, two major networks carry out active health surveillance on the world level:

- the Program for Monitoring Infectious Diseases (ProMED-mail) [55,56] is a system using the Internet to disseminate information on outbreaks and emerging diseases in real time. The data sources are several (articles in the press, official reports, communiqués, correspondence, etc.). The information is analyzed and commented by a group of experts and put online;
- the Global Public Health Information Network (GPHIN) [57] is a semiautomatic expert system developed in Canada for the WHO. It provides continuous monitoring of the international media on the Internet in English, French, Russian, Chinese, Spanish, etc. This system is the source of nearly 40% of the alerts for approximately 200 outbreaks managed every year by the WHO;
- the Google trends site (http://www.google.com/trends) monitors rumors of outbreaks. Google constantly analyzes Internet user searches and then displays, for a given key word, a statistic over several years taking into account the IP address; which allows one to determine the geographic origin of the Internet user; based on the idea that outbreaks often give rise to an explosion of searches (Fig. 4). A site on flu trends (http://www.google.org/flutrends) has even been started.

6.3. The contributions of modeling to anticipating outbreaks

Simulation of outbreak or pre-outbreak situations allows one to test the effectiveness of detection and control methods [33,58]. However, one can also improve the detection and prediction of outbreaks by constructing models using variables of interest such as environmental data. Briefly, two different approaches have been described: deterministic models and

Fig. 3. Biocollectors (Coriolis delta and MS) and the principle of impaction in Cyclone-type fluid medium.
stochastic models. Deterministic modeling is applied to phenomena that are sufficiently well known to be able to acceptably predict them from available data. They model the mean behavior of a system, attempting to reduce the variability, i.e., the random part. These models always predict the same results for a given set of explanatory variables. Conversely, stochastic models (meteorological models, bayesian models, regression models, etc.) simulate the variability that is observable around the mean state. Thus, for the same initial state of the population and its environment, several developments in terms of an outbreak can be envisioned. They provide both the forecasting and variability of these exit variables [59]. In this sense, these models seem much closer to biological reality because they are based on probabilities that an event will occur.

A good example is provided by the stochastic model of outbreak propagation around the world, which was developed based on a database on passenger flow between different airports in the world and the populations of the zones served by air transportation [60]. It thus becomes possible to obtain confidence intervals in the prediction of outbreaks. These predictive models, however, are limited by the high number of hypotheses that must be postulated.

7. Conclusion

In 2010, detection and anticipation of outbreaks is a multidisciplinary activity (epidemiology, biology, statistics, bioinformatics, health and environmental surveillance, veterinary medicine, engineering, ergonomics and human factors, etc.) calling on proteiform epidemiological sensors (healthcare personnel, sentinel chicken flocks, optronic detectors, satellites, Internet, etc.). This activity has an increasingly significant medical, medical-legal (forensic epidemiology), media, and socioeconomic impact. The Mexican Ministry of Finance estimates that the A/H1N1 flu will cost 0.3–0.5% of the nation’s gross domestic product. Such an impact makes production of false-positives less and less tolerable. Yet emitting alarms that are not confirmed by an alert is inherent to any detection activity aiming for early warning. This is why detection should not be considered a finality in itself but an essential link in the chain of outbreak management. Studies in knowledge engineering and human factors reinforce this approach by providing cognitive feedback loops returning to the detection step to validate the pursuit of the outbreak management process [61]. In public health, the most frequent attitude of statisticians remains evaluation of a surveillance system based only on its capacity to detect an aberration. For methodological reasons (e.g., the correlation between evaluation tools and statistical tools used for detection) and in agreement with the published guidelines, it appears that the most relevant level to evaluate the efficacy of a surveillance system is the alert level reached after the process of diagnosing the situation.

Conflict of interest statement

The authors have not declared any conflict of interest.

Appendix A. Supplementary data

A French version of this article is available at doi:10.1016/j.respe.2010.06.169.

References


