Part I

Introduction to Biosurveillance

Cambridge University Press
978-0-521-19134-0 - Introduction to Statistical Methods for Biosurveillance: With an Emphasis on Syndromic Surveillance
Ronald D. Fricker
Excerpt
More information

1

Overview

> While the public health philosophy of the 20th Century – emphasizing prevention – is ideal for addressing natural disease outbreaks, it is not sufficient to confront 21st Century threats where adversaries may use biological weapons agents as part of a long-term campaign of aggression and terror. Health care providers and public health officers are among our first lines of defense. Therefore, we are building on the progress of the past three years to further improve the preparedness of our public health and medical systems to address current and future BW [biological warfare] threats and to respond with greater speed and flexibility to multiple or repetitive attacks.
>
> Homeland Security Presidential Directive 21

Bioterrorism is not a new threat in the twenty-first century – thousands of years ago, the plague and other contagious diseases were used in warfare – but today the potential for catastrophic outcomes is greater than it has ever been. To address this threat, the medical and public health communities are putting various measures in place, including systems designed to proactively monitor populations for possible disease outbreaks. The goal is to improve the likelihood that a disease outbreak, whether artificial or natural, is detected as early as possible so that the medical and public health communities can respond as quickly as possible.

The ideal biosurveillance system analyzes population health-related data in near-real time to identify trends not visible to individual physicians and clinicians. As they sift through data, many of these systems use one or more statistical algorithms to look for anomalies and trigger investigation, quantification, localization, and outbreak management. This book is focused on the design, evaluation, and implementation of the statistical algorithms, as well as other statistical tools and methods for effective biosurveillance.

Before discussing the statistical methods, however, this chapter first puts them in the perspective of the systems and the data upon which they are based. It begins by first defining the term “biosurveillance” and various associated terms followed by a brief look at some biosurveillance systems currently in use and concluding with a discussion about what is known about biosurveillance utility and effectiveness.

Chapter Objectives

Upon completion of this chapter, the reader should be able to:

- Define the terms *biosurveillance*, *epidemiologic surveillance*, and *syndromic surveillance*.
- Explain the objectives of biosurveillance: early event detection and situational awareness.
- Describe biosurveillance systems in terms of system functions and components.
- Discuss biosurveillance system utility and effectiveness, including the ongoing research challenges.
- Compare and contrast biosurveillance to traditional public health surveillance and to statistical process control.

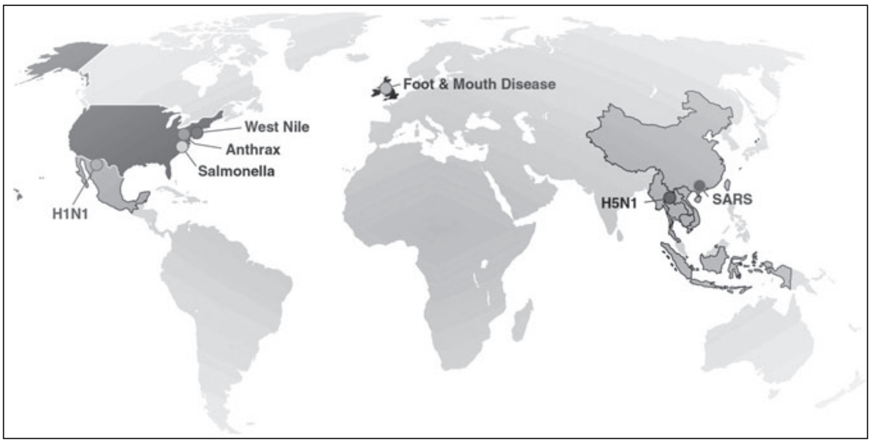

Figure 1.1. A map of select worldwide disease outbreaks in recent decades (GAO, 2010).

1.1 What Is Biosurveillance?

In the *National Biosurveillance Strategy for Human Health*, the US Centers for Disease Control and Prevention (CDC) says, "Biosurveillance in the context of human health is a term for the science and practice of managing health-related data and information for early warning of threats and hazards, early detection of events, and rapid characterization of the event so that effective actions can be taken to mitigate adverse health effects" (CDC, 2010a, p. 11). As discussed in the introduction, early detection has become an important focus as globalization and the threat of bioterrorism make the spread of highly virulent diseases an increasing threat to human health. For example, Figure 1.1 is a map of select worldwide disease outbreaks in recent decades (GAO, 2010).

Homeland Security Presidential Directive 21 (HSPD-21) defines *biosurveillance* as "the process of active data-gathering with appropriate analysis and interpretation of biosphere data that might relate to disease activity and threats to human or animal health – whether infectious, toxic, metabolic, or otherwise, and regardless of intentional or natural origin – in order to achieve early warning of health threats, early detection of health events, and overall situational awareness of disease activity" (US Government, 2007).

As shown in Figure 1.2, *biosphere data* can be divided into information about human, animal, and agricultural populations, and biosurveillance thus consists of health surveillance on each of these populations.

One particular type of biosurveillance is *epidemiologic surveillance*, which HSPD-21 defines as "the process of actively gathering and analyzing data related to human health and disease in a population in order to obtain early warning of human health events, rapid characterization of human disease events, and overall situational awareness of disease activity in the human population." Thus, epidemiologic surveillance addresses that subset of biosurveillance as it applies to human populations.

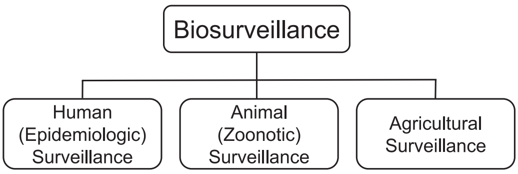

Figure 1.2. A biosurveillance taxonomy consisting of human, animal, and agricultural surveillance.

As shown in Figure 1.3, epidemiologic surveillance is but one element of public health surveillance. Public health surveillance encompasses the surveillance of adverse reactions to medical interventions (particularly drugs and vaccines) and how health services are used, as well as epidemiologic surveillance. Brookmeyer and Stroup (2004, p. 1) quote Thacker (2000) in defining public health surveillance as “the ongoing systematic collection, analysis, interpretation, and dissemination of health data for the purpose of preventing and controlling disease, injury, and other health problems.”

Syndromic surveillance is a specific type of epidemiologic surveillance that Sosin (2003) defines as “the ongoing, systematic collection, analysis, interpretation, and application of real-time (or near-real-time) indicators of diseases and outbreaks that allow for their detection before public health authorities would otherwise note them.” Thus, syndromic surveillance is epidemiologic and public health surveillance restricted to using leading indicators of disease. In particular, syndromic surveillance is based on the notion of a *syndrome*, which is a set of nonspecific prediagnosis medical and other information that may indicate the release of a bioterrorism agent or natural disease outbreak. See, for example, Syndrome Definitions for Diseases Associated with Critical Bioterrorism-associated Agents (CDC, 2003).

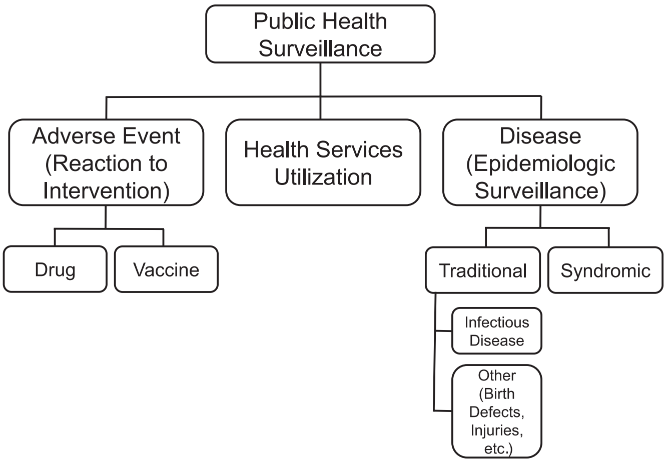

Figure 1.3. A taxonomy of public health showing that epidemiologic surveillance is but one part of a broader set of surveillance activities. (Adapted from Rolka & O’Connor, 2010.)

Table 1.1. A comparison of the categories of data used in biosurveillance, epidemiologic surveillance, and syndromic surveillance

Category of data	Biosurveillance	Epidemiologic surveillance	Syndromic surveillance
Prediagnosis			
• ER chief complaint	X	X	X
• OTC medicine sales	X	X	X
• EMS call rates	X	X	X
• Absenteeism records	X	X	X
• Lab results	X	X	X
• Other	X	X	X
Medical diagnoses	X	X	
Lab results	X	X	
Water and air monitoring	X		
Zoonotic	X		
Agricultural	X		

Syndromic surveillance differs from traditional epidemiologic surveillance in a number of important ways. For example, whereas syndromic surveillance often uses nonspecific health and health-related data (e.g., daily number of individuals reporting with sore throats to an emergency room), traditional notifiable disease reporting is based on suspected or confirmed cases (e.g., daily number of individuals diagnosed with the flu).[1] In addition, although in traditional public health surveillance, routine surveillance is conducted on specific, well-defined diseases, and non-routine surveillance is generally not initiated without a known or suspected outbreak, syndromic surveillance systems actively search for evidence of possible outbreaks[2] well before there is any suspicion of an outbreak.

As shown in Table 1.1, syndromic surveillance tends to use the least medically specific data. Often it is based on data derived from "chief complaints" of people presenting at hospital emergency rooms (ERs). Chief complaints are broad categories that capture the main reason or reasons an individual goes to the ER. Syndromic surveillance may also be based on over-the-counter (OTC) medicine sales, emergency medical services (EMS) calls, absenteeism records, and other health-related data. Epidemiologic surveillance, in comparison, uses all of these data types as well as other data from actual diagnoses and diagnostic laboratory

[1] For example, during the flu season, the *United States Influenza Sentinel Physicians Surveillance Network*, which consists of individual doctors throughout the United States, reports weekly to the CDC the total number of patients seen and the number of patients with influenza-like illness by age group.

[2] Although discussions throughout the text focus on the detection of disease outbreaks, the more general concept is detection of an adverse event, in which a disease outbreak is but one type of adverse event that may be of interest to the public health community. The term "outbreak" is used simply to make the discussion concrete and should not be taken as a limit on the application of biosurveillance.

results. Biosurveillance adds to these sources of data with information from water and air monitoring stations as well as zoonotic and agricultural data.

For the purposes of this text, biosurveillance is discussed within the context of epidemiologic surveillance and often even within the more specific context of syndromic surveillance. However, because the statistical methods discussed in this book clearly apply to a broad class of public health surveillance problems, the general term "biosurveillance" is used throughout.

1.1.1 Biosurveillance Objectives

Syndromic surveillance has also been defined as ". . . surveillance using health-related data that precede diagnosis and signal a sufficient probability of a case or an outbreak to warrant further public health response" (Fricker & Rolka, 2006; CDC, 2006a). This definition focuses on a number of ideas important to biosurveillance.

- First, biosurveillance is health surveillance, not military, regulatory, or intelligence surveillance. It may use a wide variety of types of data, from case diagnoses to health-related data such as chief complaint counts.
- Second, the data and associated surveillance are generally intended to precede diagnosis or case confirmation in order to give early warning of a possible outbreak. Clearly, after a definitive diagnosis of a bio-agent has been made, the need for detection becomes moot, although tracking the location and spread of a potential outbreak are still important whether an outbreak has been confirmed or not.
- Third, the process must provide a signal of "sufficient probability" to trigger "further public health response." Often the goal is not to provide a definitive determination that an outbreak is occurring but rather to signal that an outbreak *may* be occurring. Such a signal indicates that further investigation is warranted in the form of a more detailed investigation by public health officials.

Biosurveillance systems have two main objectives: to support public health *situational awareness* (SA) and to enhance outbreak *early event detection* (EED). The CDC (2008) defines them as follows:

- **Situational awareness** is the ability to utilize detailed, real-time health data to confirm or refute and to provide an effective response to the existence of an outbreak. It also is used to monitor an outbreak's magnitude, geography, rate of change, and life cycle.
- **Early event detection** is the ability to detect, at the earliest possible time, events that may signal a public health emergency. EED is comprosed of case and suspect case reporting along with statistical analysis of health-related data. Both real-time streaming of data from clinical care facilities as well as batched data with a short time delay are used to support EED efforts.

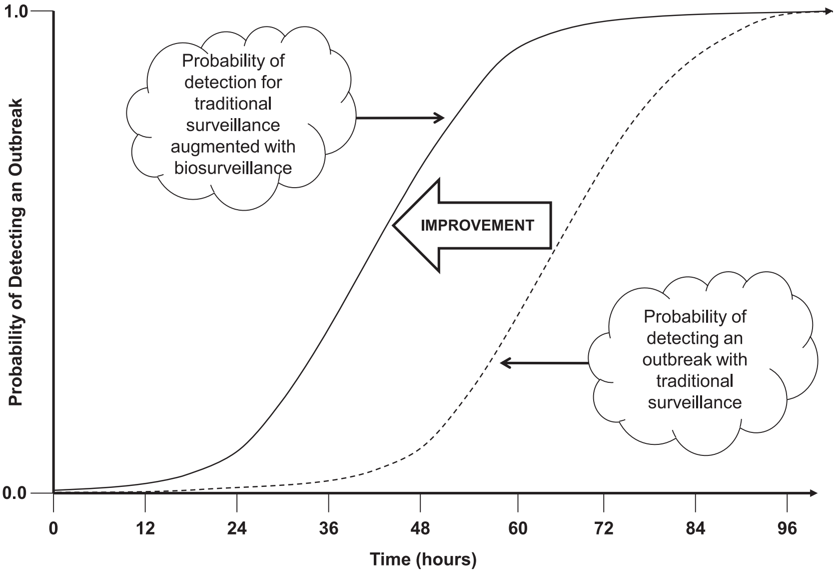

Figure 1.4. An illustration of how biosurveillance is intended to improve the probability of detecting a disease outbreak, whether artificial or natural.

As illustrated in Figure 1.4, biosurveillance systems are supposed to improve the chances the medical and public health communities catch a disease outbreak early. The more biosurveillance improves the probability of detecting an outbreak, the more a biosurveillance system is likely to enhance SA and EED. The goal is "... deployment of surveillance systems that can rapidly detect and monitor the course of an outbreak and thus minimize associated morbidity and mortality" (Bravata *et al.*, 2004).

The vertical axis in Figure 1.4 is in terms of probability specifically because early event detection is a stochastic[3] phenomenon, in which a variety of circumstances drive if and whether either a clinician or a biosurveillance system first detects a particular outbreak. In a given situation, whether a clinician will be faster than a biosurveillance system or vice versa is a function of both the aspects of the specific situation and chance.

When assessing biosurveillance systems, speed of (true positive) detection is one of three dimensions critical for completely characterizing performance. The other two dimensions are the rate of false positives and the probability of successfully detecting an outbreak. In the biosurveillance literature, these dimensions are often generically referred to as timeliness, specificity, and sensitivity.

All three dimensions are necessary, and they trade off. For example, for a given EED methodology, improving the speed of detecting an outbreak generally

[3] The term "stochastic" means the outcome involves an element of chance.

Table 1.2. Bioterrorism diseases, agents, and pathogens and associated symptoms. (Adapted from Grey & Spaeth, 2006, Table 5-2, p. 77)

Symptoms	Disease, agent, or pathogen
Respiratory distress and fever	Anthrax Plague Ricin Staph enterotoxin B
Rash and fever	Smallpox Viral hemorrhagic fevers
Flu-like	Tularemia Brucellosis Q Fever
Neurologic	Botulism Venezuela equine encephalitis Eastern equine encephalitis Western equine encephalitis
Blistering	T-2 mycotoxin Arsenicals Mustards Phosgene oxime

comes at the cost of increasing the rate of false positives. Similarly, increasing the probability of detection usually comes at the expense of the speed of detection. These trade-offs are similar to the Type I and Type II error trade-offs inherent in classical hypothesis testing, although the sequential decision-making aspect of biosurveillance adds an additional level of complexity.

Chapter 6 discusses the appropriate metrics for assessing biosurveillance systems as well as how to evaluate biosurveillance system performance.

The motivation for biosurveillance and syndromic surveillance systems in particular is that some bio-agents have symptoms in their prodromal stages similar to naturally occurring diseases. For example, in the first week or two after exposure to smallpox, individuals tend to have symptoms similar to those of the flu such as fever, malaise, aches, nausea, and vomiting (Zubay, 2005). Table 1.2 lists various bio-agent pathogens by the syndromes corresponding to the most likely symptoms of the pathogens. The idea of syndromic surveillance, then, is that detecting an anomalous increase in one or more syndromes might correspond to a bioterrorism incident. Similarly, geo-spatial displays of one or more syndromes can provide additional situational awareness on the spread of a pathogen.

1.2 Biosurveillance Systems

As HSPD-21 states, “A central element of biosurveillance must be an epidemiologic surveillance system to monitor human disease activity across populations.