1

Introduction: Spatial and Syndromic Surveillance for Public Health

Andrew B. Lawson and Ken Kleinman

1.1 WHAT IS PUBLIC HEALTH SURVEILLANCE?

The Centers for Disease Control and Prevention (CDC) define public health surveillance as:

> the ongoing, systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know. The final link of the surveillance chain is the application of these data to prevention and control. A surveillance system includes a functional capacity for data collection, analysis, and dissemination linked to public health programs.
>
> (Thacker, 1994)

It is clear from this that a broad definition of surveillance is implied and that it relates to a wide range of monitoring methods related to health. From a statistical point of view it is relevant to consider how statistical methods can be developed or employed to best aid the task of surveillance of populations. This will require using all of the relevant data available for analysis. It will certainly include information about *where* the data was recorded as well as *when* it was observed.

1.1.1 Spatial Surveillance

There is thus a need to combine the thinking in two previously mostly distinct fields of statistical research, namely surveillance, which generally constitutes

Spatial and Syndromic Surveillance for Public Health Edited by A.B. Lawson and K. Kleinman

monitoring statistics for evidence of a change, and spatial techniques, which are often used to find or describe the extent of 'clustering' across a map. While both endeavors pre-date the formal study of statistics as a discipline, they have rarely been combined. More often, as in the famous case of John Snow and cholera in London in 1854, note of an increase in a global statistic has been followed by a spatial analysis to determine whether the increase is localized or general. Interest in doing spatial monitoring of data as it accrues has been greatly enhanced by two developments: the perceived need to quickly detect bioterrorism after the terrorist dissemination of *Bacillus anthracis* in October 2001, and the increasing availability of data that contains spatial (geographical location) information.

1.1.2 Syndromic Surveillance

Another result of the burgeoning availability of data has been the recognition of a need and an opportunity. The need is for the ability to group symptoms together in broad groups that combine similar types of complaints – this being necessary to ensure that two cases attributable to the same cause are not considered separately due to variable coding practice on the part of health care providers. Misappropriating from medical nomenclature, these groups of symptoms are loosely designated as 'syndromes'. The opportunity presented by the increasing availability of data is to use, for public health purposes, data that has not often been recognized as useful in this way. Examples include information about school absenteeism and over-the-counter sales of remedies such as anti-diarrheals.

Together, grouping of large numbers of symptoms and data regarding nontraditional sources of information are labeled as 'syndromic surveillance'. The putative advantage of syndromic surveillance is that detection of adverse effects can be made at the earliest possible time, possibly even before disease diagnoses can be confirmed through unmistakable signs or laboratory confirmation.

1.2 THE INCREASED IMPORTANCE OF PUBLIC HEALTH SURVEILLANCE

In addition to the CDC definition, we might consider a dictionary definition of surveillance: 'the close observation of a person or group, especially one under suspicion'. In this light we would define public health surveillance as the monitoring of the health of the public for the onset or outbreak of illness. The illness surveilled may be rare (plague) or recurrent (influenza), natural or intentional (bioterrorism).

Since the intentional release of anthrax in the USA in October 2001, there has been a great deal of interest in establishing systems to detect another such attack as early as possible, should it occur. The need for early detection is motivated by two facts. First, many agents that might be used for such an attack have a prodromal phase that is relatively nonspecific, with symptoms that often resemble those of the common cold. This describes anthrax, botulism, plague, smallpox, and tularemia – all of the CDC class A bioterrorism agents except for viral hemorrhagic fevers (CDC, 2004). If the attack can be detected while most victims are in this phase, they may be helped by specialized care, and future onsets may be prevented by prophylaxis. Second, for contagious diseases, earlier interdiction can slow down or stop the epidemic curve; the latter might be impossible if detection were delayed.

A great deal of resources are being expended on mechanical detection of airborne organisms, such as anthrax spores. We do not discuss such efforts here and merely observe that from the statistical perspective, detection is finished once a spore of anthrax has been positively identified. (Determining the locations in need of treatment or prophylaxis is a separate question). In this book, we focus on individual human beings, in contrast to disease organisms. While it is true that a single definitive diagnosis of anthrax or any of the organisms cited above also ends the statistical interest in detection, the fact of the nonspecific prodrome opens a window for detection of an attack before a definitive diagnosis has been made. To wit, since the prodromal symptoms are so common, one might search for unusual increases in symptoms consistent with the prodrome. Such an increase could be due either to natural variation in the symptom incidence or to an attack with some agent that causes those symptoms in the prodrome.

In this context, we are most interested in detecting attacks while they are ongoing rather than retrospectively. In statistical terms, we might refer to this as 'cluster detection' or 'incident cluster detection', where by 'cluster' we mean the occurrence of extra cases in a short time span. In the literature on surveillance, this is sometimes referred to as 'on-line' surveillance (Chapter 3). Many techniques exist for ongoing monitoring or surveillance of a count; these come from industrial applications – for example, Shewhart control tables and cumulative sum (CUSUM) methods (Chapter 2) – as well as from public health surveillance (Huttwagner, 2003; Sonesson and Bock, 2004).

1.3 GEOGRAPHIC INFORMATION, CLUSTER DETECTION AND SPATIAL SURVEILLANCE

The increased need for cluster detection has coincided with an increasing availability of data, especially data on the location of events. This is often obtained by geocoding the addresses of individual cases. This can be done 'on the fly' as cases are encountered (Beitel *et al.*, 2004) or with static databases that retain the location of all patients eligible for surveillance (Lazarus *et al.*, 2002). In its simplest form, geocoding could imply merely obtaining the zip or postal code, but it may also include finding the exact latitude and longitude of an address using geographical information systems (GIS). In statistical jargon, such data about location is often referred to as 'spatial' data.

The value of spatial data for cluster detection is twofold. First, all attacks are localized at some spatial scale. That is, an attack could conceivably target a neighborhood, but on a city-wide scale this would be a small area. Alternatively, an attack could include a whole metropolitan area, but on a national scale this would be a small region. When surveillance is limited to a single daily count from a neighborhood or city, even sharp increases in relatively small regional counts may be hidden within the natural variation found in the count across a larger area. Spatial surveillance thus promises to increase the power to detect events that occur in small regions, relative to surveillance of the total count only. Secondly, if an incident cluster is identified, public health officials will need to respond. If the data are nonspatial, surveillance can only give vague messages of the sort 'there is an excess of cases in the Boston metropolitan region'; this is unlikely to be of much practical use. In contrast, spatial surveillance would allow more-specific messages, such as 'there are excess cases in zip code 02474'. The job of identifying small regions with extra cases is also referred to as 'cluster detection', where the clustering in this case refers to extra cases in an area on the map.

The coincidence of suddenly increased need and increasingly available spatial data has generated new interest in statistical methods for spatial surveillance, which might be described as the detection of incident clusters in space. The goal of this book is to provide a snapshot of the state of the nascent art of incident spatial cluster detection, provided by statisticians involved in traditional surveillance (of a single statistic), in spatial clustering, and in spatial surveillance.

1.4 SURVEILLANCE AND SCREENING

An idea related to surveillance is that of screening. The use of screening to allow the early detection of disease onset is well established, though possibly controversial, in such areas as cervical or mammarian cancer. These examples of screening involve testing individuals at regular time points to attempt to assess if onset of a condition has occurred or is likely or imminent. Screening could be applied to populations as well as individuals, in that changes in public health might trigger interventions. Such interventions could be designed to redirect health resources towards attempts to improve the health status of the population. However, screening is usually associated with individual assessment or monitoring, while surveillance is usually carried out at an aggregate population level.

Surveillance and screening share an implicit temporal dimension: populations or individuals are assessed (often repeatedly over time) to assess whether changes have occurred which may warrant action. In general, a change is defined as exceeding limits describing the acceptable results of current observation and actions taken if these limits are passed. In screening individuals, the limits may be based on a previously observed known or stable abnormal baseline or on 'normal' standards thought to obtain in healthy persons.

In surveillance, the 'normal' case is rarely known, and most attention is directed to detect passing limits based on observed or expected patterns. These limits may be fixed or may depend on the status of ancillary variables. For example, incidence of influenza-like illness would be expected to vary seasonally, so similar numbers of cases would be more or less alarming at different times of year.

To carry the screening analogy further, the location of the public health incident is as important as the fact that it occurred. A public health report indicating only a disease outbreak is comparable to a garbled mammography report that only indicates a cancer but no suggestion of which breast is affected, let alone a location in which a biopsy would be appropriate. In population-level analysis, statisticians use 'spatial' statistics to discuss location. Further still, mammography uses the spatial information to help identify the existence of the node in the first place.

1.5 OVERVIEW OF PROCESS CONTROL AND MAPPING

Process monitoring is necessary for quality control in a manufacturing context. The subject of statistical process control (SPC) has received the most methodological attention of all surveillance questions. SPC has formed the basis for many disease surveillance systems. In this section we describe some basic SPC methods that could be applied in this context.

1.5.1 Process Control Methodology

A number of methods have been developed for the detection of changes in populations over time. These methods are characterized by the estimation of changepoints in a sequence of disease events or a time series of population rates, or the determination of or application of control limits to the behavior of a system. In this area there are some simple methods available to assist in the assessment of change or 'in control' behavior. Some of these methods are derived from SPC, which was developed for the monitoring of industrial processes over time, and could be applied within a disease surveillance program, with due care. For example, it is well known that the temporal variation in count data can be monitored by using a Poisson control chart (C or U chart), upon which specific limits can be plotted beyond which corrective action should be taken. These charts are based on normal pivotal approximations.

An exact interval could be constructed for independent Poisson counts in an attempt to utilize SPC methods. However, if the counts were correlated even under the null hypothesis, then some allowance must be made for this correlation in the chart. A further issue, when such methods are to be used within disease monitoring, is the issue of how to incorporate any changes in the background 'at-risk' population which may arise. One possibility, in the temporal domain, is to employ relative risk estimates. For large aggregation scales, time-series methods have been employed which allow temporal dependence (see Chapter 2 in this volume).

In addition, special types of chart (CUSUM charts) have been developed specifically to detect changes in pattern over time (changepoints). These are constructed by cumulative recording of events over time, the accumulation being found to be sensitive to changepoints in the process under consideration. Some recent work in the application of these ideas in medical surveillance and monitoring has been done by Frisén and co-workers (Chapters 3 and 9 in this volume). These methods require special adaptations to be developed to deal with the spatial and spatio-temporal nature of geographical surveillance.

The main issues within temporal surveillance which impact on spatial surveillance and spatio-temporal surveillance can be categorized into three classes: detection of changepoints (mean level, variance), detection of clusters, and the detection of overall process change. Conventional SPC would use control limits to detect shifts in single or multiple parameters where the target parameters are usually constant. However, disease incidence varies naturally in time and so allowance must be made for this variation in any monitoring system, particularly with variation in population at risk. In addition, particular departures from the 'normal' variation are often of greater interest than simple shifts of parameters. Changepoints, where jumps in the incidence occur, could be a focus of interest. Alternatively, clusters of disease may be important. Finally, there may be an overall process change, where various parameters change. Any disease surveillance system is likely to be focused on one or all of these changes. Indeed, it is the multiple focus of such systems that is one of the greatest challenges for the development of statistical methodology.

1.5.2 The Analysis of Maps and Surveillance

In the spatial case, there is a wide range of methods that can be applied to a single map of case events within a fixed time frame/period. Many of the methods applied in disease mapping, clustering or ecological analysis could be applied as a surveillance tool. For example, general clustering tests could be applied or residuals from disease maps fitted in each time period could be examined. Questions which might be appropriate to answer with these methods are such as: Is there evidence of unusual variation in incidence in the map? Is there evidence of 'unusual' clustering on the map? Is there a spatial trend on the map related to, for example, a putative source?

However, when the question relates to a spatio-temporal pattern or change in pattern, then there are few methods currently available which are designed for this purpose. There is a correspondence between the temporal surveillance foci, and features which are important to detect in the spatial domain. First, localized discontinuities in mean level or variance of risk could be of concern (changepoints). Second, spatial clusters of disease could be a focus. Finally, overall process change could also be envisaged spatially.

1.6 THE PURPOSE OF THIS BOOK

We hope that this book may serve a dual purpose. First, we hope that the potential users of spatial surveillance – the public health authorities – will use it as an introduction to the value of spatial data and as guide to analytic methods competing for scarce resources. Second, we hope that the statistical community will use it as a spur to further development of techniques and to resolution of questions unanswered by the chapters which follow.

1.6.1 Statistical Surveillance and Methodological Development in a Public Health Context

The ongoing aim of public health surveillance, since the time of John Snow, has been to identify public health problems as they occur and respond appropriately when they do. Breaking this down into discrete steps, this involves determining from whom to collect data, collecting the data, summarizing the data, evaluating the summarized data, and taking action if the evaluation warrants it. 'Action' can be defined as any additional steps that are not performed on a routine basis, which might include everything from asking for additional data to the vigilante removal of a pump handle or launching some other direct interdiction to prevent further illness.

1.6.2 The Statistician's Role in Surveillance

As statistical analysts, it is important that we remember that our role in public health surveillance is in evaluating the summarized data; this may be the main factor in the decision whether to take action *now*. This is different from academic work and most scientific work in several important ways.

First of all, we do not have the luxury of academic distance from the subject. This means that the case for making the best-supported decision needs to be made in a way that is powerful and easily absorbed by decision-makers. These authorities may lack the time, inclination, centralization, or training to follow complicated arguments or abstract presentations of our evaluations.

One example of simplifying the evaluation is included in Chapter 5, on modeling, where we discuss inverting the p-value as a means of making the unusualness of the result under the null more approachable. A second difference is that the time available to develop analyses is very brief. In contrast, the motivation to complete statistical work quickly mainly derives from fear of competition or career goals. The impact of this difference is that simpler, more easily generalizable methods have a practical advantage – they can be deployed in practical situations. Methods tailored to one situation may fit that data better but be essentially useless when the data changes or when applying the method to a different situation. Similarly, health departments may never use methods so complicated that they require the active participation of a Ph.D. statistician. Finally, in our experience, decision-makers often urgently request results immediately after data becomes available. This is another impetus towards easily generalizable, easily used methods.

On the other hand, while statistical evaluation may be a key datum informing a decision, we should also remember that it is not actually the decision itself. Other data important to decision-makers include financial, political, and organizational feasibility considerations.

1.7 THE CONTENTS OF THIS BOOK

In the initial section (Part I), we provide an introduction and grounding in traditional temporal surveillance. This includes the current chapter, plus an overview and an evaluation of methods used in purely temporal surveillance. The goal of these chapters is to bring a reader unfamiliar with surveillance to a level that subsequent chapters can be more easily digested. This is necessary because those chapters may take as read concepts native to traditional temporal surveillance.

We begin with a discussion of purely temporal surveillance by Yann Le Strat (Chapter 2). Purely temporal surveillance is commonly used in most public health departments, and is an area studied little outside the areas of statistical process control and surveillance. Thus statistical readers may find a review helpful. The chapter includes an introduction to surveillance as well as a survey of typical methods. Methods considered include historical (nonstatistical) limits, process control charts (Shewhart charts, moving average charts, exponentially weighted moving average charts, CUSUM charts), time-series analysis, combinations of process control and time-series methods, integer-valued autoregressive processes, Serfling's method, and log-linear and other parametric models.

We also provide a discussion, by Marianne Frisén and Christian Sonesson (Chapter 3), of optimality in surveillance, and how detection methods might be designed with optimality in mind. This includes a discussion of evaluation metrics for surveillance (including false alarms, delay before alarm, and predictive value of alarms) and optimality criteria (including minimal expected delay, minimax optimality, and average run length). The chapter goes on to discuss the optimality and performance features of several methods described in Chapter 2. It concludes by discussing several features of the public health environment that differentiate it from other applications of surveillance.

In Part II of the book, we provide a summary and some development of statistical approaches currently applied for spatial surveillance. First, Chapter 4 provides an overview of spatial and spatio-temporal health analysis outside of surveillance. This includes a discussion of disease mapping in the cases where individual locations of each case are known and alternatively when cases are aggregated into regions, as well as assessment of maps through residuals and goodness of fit. Finally, spatio-temporal and surveillance issues are introduced in the spatial context.

In Chapter 5, a summary of generalized linear models and generalized linear mixed models, including the use of binomial and Poisson models, is offered. Another purpose of the chapter is to note advantages that are realized through Poisson models (including variable-duration cluster signals) and to compare the surveillance resulting from the various models in an example data set.

In Chapter 6, Peter Rogerson addresses how CUSUM methods can be adapted to spatial surveillance. This includes a discussion of statistical process control that can be tailored for use in spatial applications, followed by a demonstration. Uses include surveillance of multiple local regions as well as of global statistics.

Martin Kulldorff (Chapter 7) discusses how scan statistics can be used in this context, and recent developments in this approach. The chapter mentions tests of spatial randomness, then introduces scan statistics. This is followed by a thorough introduction to the practical application of scan statistics for spatial health surveillance. This includes a discussion of the null and alternative hypotheses for the test, as well as the power and methods for displaying the suggested clusters. Finally, some applications in cancer clustering, infectious disease, other human diseases, veterinary medicine, and plant disease are surveyed.

In Chapter 8, Laura Fosberg and co-workers discuss distance methods for cluster detection and identification. This includes a motivation and summary of distance-based methods, the introduction of a new statistic based on distances, and a simulation-based evaluation of the new statistic. An example of syndromic spatial surveillance using the statistic is provided.

Next Christian Sonesson and Marianne Frisén (Chapter 9) consider multivariate surveillance, what is often described as multiple streams of surveillance data. This topic addresses the common case where either different data sources supply information regarding a single syndrome, or where a single data provider reports on multiple syndromes. The approaches mentioned include a reduction of dimensionality (to one or a few statistics) for each time point, parallel surveillance, vector accumulation methods, and simultaneous solution. They also discuss evaluation in this context.

In Part III, advanced approaches to syndromic and spatial surveillance are considered, including Bayesian models and data mining techniques.

In Chapter 10, Neil and co-workers discuss the use of Bayesian networks and the development of computational algorithms; in Chapter 11 they consider speeding up spatial processing of large data sets. In Chapter 12, David Madigan provides an example of Bayesian modeling of temporal surveillance using hidden Markov models. Finally, in Chapter 13, general issues in the Bayesian analysis of syndromic data and the model-based detection of spatial and spatio-temporal clusters as they evolve in time are discussed.