

# Scan Statistics for Temporal Surveillance for Biologic Terrorism

Sylvan Wallenstein,<sup>1</sup> J. Naus<sup>2</sup>

<sup>1</sup>Mount Sinai School of Medicine, New York, New York; <sup>2</sup>Rutgers, the State University, New Brunswick, New Jersey

**Corresponding Author:** Sylvan Wallenstein, Box 1023, 1 Gustave Levy Place, Mount Sinai School of Medicine, New York, NY 10029. Telephone: 212-241-1526; Fax: 212-860-4630; E-mail: sylvan.wallenstein@mssm.edu.

## Abstract

**Introduction:** *Intentional releases of biologic agents are often designed to maximize casualties before diagnostic detection. To provide earlier warning, syndromic surveillance requires statistical methods that are sensitive to an abrupt increase in syndromes or symptoms associated with such an attack.*

**Objectives:** *This study compared two different statistical methods for detecting a relatively abrupt increase in incidence. The methods were based on the number of observations in a moving time window.*

**Methods:** *One class of surveillance techniques generates a signal based on values of the generalized likelihood ratio test (GLRT). This surveillance method is relatively well-known and requires simulation, but it is flexible and, by construction, has the appropriate type I error. An alternative surveillance method generates a signal based on the p-values for the conventional scan statistic. This test does not require simulation, complicated formulas, or use of specialized software, but it is based on approximations and thus can overstate or understate the probability of interest.*

**Results:** *This study compared statistical methods by using brucellosis data collected by CDC. The methods provided qualitatively similar results.*

**Conclusions:** *Relatively simple modification of existing software should be considered so that when GLRTs are performed, the appropriate function will be maximized. When a health department has data that indicate an unexpected increase in rates but its staff lack experience with existing software for surveillance based on GLRTs, alternative methods that only require computing Poisson probabilities can be used.*

## Introduction

Traditional surveillance systems tend to focus on compulsory reporting of specific diseases. However, in recent years, syndromic surveillance based on emergency department admissions, hospital bed occupancy, pharmaceutical sales, and other correlates of disease has increased to detect possible biologic terrorism attacks (1). This study analyzed methods useful in detecting surges in illness (1), particularly when these increases are abrupt, as might occur during a biologic attack.

This study was based on the assumption that, according to historic data, events occur on the basis of a known pattern of events (e.g., seasonal, specific day of the week, or weather). Methods used to estimate this pattern based on historic data have been addressed by others (1–3) and are not the focus of this paper, although one simple fitting method is illustrated. Although multiple statistical approaches to surveillance have been proposed and compared before 2001 (4,5), interest in these methods has recently increased (6–8).

This study's overall approach scans time, seeking unusual incidence within a short period. The symbol  $t$  represents current time, and  $w$  represents a window of time used for surveillance, usually a limited number of days.  $Y_t(w)$  is the number

of events in the last  $w$  days before and including  $t$ , and  $E_t(w)$  is the expected number of such events, usually based on historic data. The proposed methods result in an alert being generated at time  $t$ , if  $Y_t(w)$  is substantially greater than  $E_t(w)$ . The procedures are designed so that, if the event rates are the same as the historic rates, the probability of generating one or more false signals in a period  $T$  is  $\alpha$ . The total time frame  $T$  is under the investigator's control.

The procedures described in this paper can be contrasted with what are termed quadrat-based tests (9) or cell procedures. In such procedures, time is subdivided into non-overlapping periods of days, weeks, or months, and the data analyst searches for substantial increases in these periods. The Communicable Disease Surveillance Centre (CDSC) in London uses such a system (10) to automatically scan weekly reports to provide early warning of disease outbreaks. CDSC staff compare observed counts of a disease in a given week with historically fitted expected counts. However, equally concerning is a cluster of cases that occurs during a 7-day period that overlaps 2 calendar weeks. In a monitoring system that continuously updates reports, advantages exist, both with power and speed of detection, in using scan-like statistics and

examining the number of cases in a moving time interval instead of just looking at nonoverlapping intervals. This is particularly true for monitoring disease organisms that can be used for a biologic terrorism event, during which an early warning might be critical. If the reported effect of a release of a biologic agent is expected to spread over a 7-day period, then health department staff use a 7-day scanning window rather than a calendar week for monitoring.

This study focused on how staff decide that an observed count in a limited window of width  $w$  (measured in days or weeks) is more substantial than expected, taking into account multiple testing during a longer surveillance period  $T$ . Two functions of the observed and expected values were used to judge what constitutes more substantial counts. One function was based on generalized likelihood ratio tests (GLRTs), and the other was based on p-values calculated from the classical, constant-risk scan statistics.

Both of these approaches can be viewed as extensions of the classical scan statistic, the maximum number of observations in an interval of width  $w$ . One of the defects of the classical scan statistic is that it assumes a constant baseline rate (4). This difficulty can be overcome by scanning on the basis of GLRT (11–14). The first procedure discussed in this paper shares a common theoretical background with this surveillance method but differs in that the type I error refers to a period of time (e.g., the time of a limited objective surveillance, or a month, or a year) rather than the instant at which an alert might be generated. The second procedure, based on p-values, does not require simulation and thus can be more easily applied.

For this study, both of these procedures were applied to brucellosis data collected by CDC during 1997–2002. The point of using these example data is not to evaluate brucellosis but to illustrate how such an analysis can be performed.

## Methods

For this study, the authors assumed that the incidence of events follows a Poisson process. In this description of the methods, the notation concerning the process was suppressed, and focus was placed on  $E_t(w)$ , the expected number of events in a window of  $w$  days ending at time  $t$ . The first test requires that the window width  $w$  be fixed before the surveillance; this condition is then removed.

In a biologic terrorism event, the difference between an early signal and an obvious outbreak might be days. A critical period exists,  $d$  days, within which the data analyst should detect the increase. Multiple authors (7,8) have reported that special techniques are needed when only a limited time delay can be tolerated. Therefore, the signal decision should be based on

observations within the past  $d$  days. In this context, the window size is in the range  $w \leq d$ . Alternatively, for increased power, a fixed window of  $w = d$ , or  $w = d - 1$ , can be used.

## G-Surveillance Methods

If the window width,  $w$ , is fixed in advance, G-surveillance used to detect an abrupt increase, on the basis of a fixed type I error for a given period, generates an alert for substantial values of the statistic  $G_t(w)$ ,

$$G_t(w) = Y_t(w) \ln [Y_t(w)/E_t(w)] - [Y_t(w) - E_t(w)]$$

where  $\ln$  is the natural logarithm. (Details of the proof are available from the corresponding author upon request.) An alert will be sounded at time  $t$ , if  $G_t(w)$  is larger than a threshold (i.e., the critical value) obtained through simulation.

The extension to the case where  $w$  is not fixed but is within a certain range (e.g., 1–3 days) follows the same pattern as previously described (9,15,16). G-surveillance with variable window widths will signal an alert at time  $t$  if

$$G_t(u \text{ to } v) = \max_{u \leq w \leq v} G_t(w)$$

is larger than a new critical value.

When data are recorded daily,  $u$  in the previous equation corresponds to the smallest number of days of interest (presumably,  $u = 1$ ), and  $v$  to the largest number of days. When surveillance is continuous,  $u$  should not be set so small that it picks up artifacts of data collection and, in certain contexts, might be  $\geq 24$  hours. If the expected values depend only on past history, the threshold can be obtained before surveillance begins by generating realizations of the complete process. For numerous local health departments to avoid having to develop expertise in simulating the process, this critical value can be computed once a year at a central location and then transmitted to local health departments. In other cases, the expected values depend on current data (e.g., weather conditions), and the user might have to re-do simulations at each time point  $t$ .

## P-Surveillance Methods

An alternative method is a fixed-window scan surveillance method, P surveillance, that does not require simulation but instead is based on p-values from the classical scan statistic (17). The traditional fixed window scan statistic,  $S_w$ , is the largest number of cases to be found in any subinterval of length  $w$  (for  $w$ , a known constant) of the surveillance interval  $(0, T)$ . Two recent books (18,19) summarize results on finding the exact probability (20), finding bounds (21), and finding approximations (21,22) for the distribution of  $S_w$ . For the atypical surveillance application, in which the expected

number of events in any interval of width  $w$  is a constant,  $\lambda$ , the approximation (22) is given by

$$\Pr(S_w > k) = (T/w) (k-\lambda) p(k, \lambda) + s(k, \lambda)$$

where  $p(k, \lambda)$  is the Poisson probability of observing exactly  $k$  events,  $p(k, \lambda) = \exp(-\lambda) \lambda^k / k!$ , and  $s(k, \lambda)$  the probability of observing  $k + 1$  or more events.

The limited usefulness of the classic scan statistic in surveillance, because of its assumption of constant baseline risk, has been noted (4). One early method to overcome this limitation involved stretching or contracting time (23), which has the disadvantage that it would not allow surveillance in 24-hour units. G-surveillance is another way to overcome the limitation.

P-surveillance is based on computing a p-value at time  $t$ , focusing on what is happening at that time and ignoring all other information. The same p-value should be used if the baseline risk over the whole period is constant at the local rate at time  $t$ .

Under continuous surveillance, an alert is signaled at time  $t$ , if

$$(T/w) [Y_t(w) - E_t(w)] p[Y_t(w), E_t(w)] + s[Y_t(w), E_t(w)] < \alpha$$

Under this procedure,  $\alpha$  is the probability of generating a false alert in time frame  $T$  (e.g.,  $T = 1$  year) and will usually be set to 0.05 or 0.10. In surveillance applications, loss of precision will be limited if the second term in the last equation is ignored so that an alert will be signaled if

$$(T/w) [Y_t(w) - E_t(w)] [\exp[-E_t(w)] E_t(w)^{Y_t(w)} / Y_t(w)!] < \alpha$$

Thus, P-surveillance in continuous time requires calculating the left side of the previous equation each time an event occurs and deciding if it is less than a prespecified  $\alpha$ .

Conceptually, a different test based on the ratchet scan statistic (24) should be performed when the data are collected daily or weekly instead of continuously. The principle underlying the test would be the same.

Justification for the use of P-surveillance requires 1) demonstrating formally that theoretical (mathematical) reasons exist to assume that P-surveillance has the claimed false-alert rate, and then substantiating it by simulation, and 2) using theory or simulations to demonstrate that P-surveillance had power somewhat comparable to G-surveillance. Work on the first assertion has already been performed (18), and limited numerical work by the authors supports the second assertion.

## Results (Example)

A study of disease characteristics of microbiologic agents with particular potential for biologic terrorism lists brucello-

sis among critical biologic agents reported to the National Notifiable Disease Surveillance System (25). For this paper, weekly national reports of brucellosis are used (for illustration purposes only) as a proxy for the type of daily totals that might arise for certain more common conditions in limited geographic areas.

Provisional (and for years 2001 and 2002, revised) cumulative data can be obtained from *Morbidity and Mortality Weekly Report* (available at <http://www.cdc.gov/mmwr>). The data are revised to adjust for delayed reporting because certain states submit reports in batches and include suspected cases in addition to confirmed ones. In using the provisional cumulative data, distinguishing between negative adjustments caused by removing previous suspected cases and new suspected or confirmed cases is impossible. This study used revised data for 1997–2001 provided by CDC (Table) as a proxy for the analysis possible if the provisional data provided the number of new cases/week.

Of these 260 weekly baseline counts, all but three are in the range of 0–7. These three cases are all in different years and occur at the end of the year. Careful scrutiny of the counts reveals certain yearly and seasonal patterns; however, to obtain an overall impression of the magnitude, the mean (1.60; standard deviation: 1.45) of the remaining 257 counts was computed (Table).

The following procedure was used to calculate the estimated value per week (Table). The average (or for weeks 49 and 52, the median) number of cases of brucellosis per week during 1997–2001 was calculated. The averages were then smoothed by fitting a spline to the means (or for weeks 49 and 52, the medians) for the first 51 weeks of data. No adjustment was made for a possible secular trend.

G-surveillance (i.e., GLRT-based, scan-type methods) was based on 1) a fixed 3-week window size and 2) on a window that can be either 1, 2, or 3 weeks. Because the model postulated does not involve any factors unknown at the start of the year, percentiles of interest can be computed once before the surveillance period begins. To obtain the percentiles for both statistics, 100,000 realizations of the process were simulated for the period of  $T = 52$  weeks, in which weekly counts were generated on the basis of a Poisson distribution with the expected value (see last column of Table).

	Percentiles of statistics				
	0.50	0.75	0.90	0.95	0.99
G(3 weeks)	1.98	2.82	3.77	4.36	5.94
G(1–3 weeks)	2.67	3.49	4.48	5.19	6.77

**TABLE. Revised brucellosis counts per week and predicted values**

Week							Average	Spline
	1997	1998	1999	2000	2001	2002	1997–2001	
1	1	0	1	1	1	1	0.8	0.81
2	0	0	1	3	0	1	0.8	0.81
3	2	0	1	0	2	2	1.0	0.8
4	0	1	0	0	2	2	0.6	0.8
5	0	1	1	0	2	1	0.8	0.81
6	1	0	2	2	1	1	1.2	0.83
7	0	0	1	1	0	0	0.4	0.88
8	0	0	2	0	1	4	0.6	0.96
9	0	0	0	2	4	3	1.2	1.08
10	3	0	0	1	3	1	1.4	1.18
11	3	0	1	0	5	1	1.8	1.23
12	0	0	1	2	2	1	1.0	1.23
13	1	1	2	1	1	1	1.2	1.2
14	1	1	2	1	2	1	1.4	1.17
15	1	1	0	0	2	2	0.8	1.16
16	1	1	0	3	1	5	1.2	1.2
17	0	1	0	2	0	0	0.6	1.3
18	3	0	0	3	2	0	1.6	1.46
19	0	3	1	3	4	5	2.2	1.62
20	0	1	3	3	1	5	1.6	1.74
21	4	2	0	3	5	9	2.8	1.82
22	1	1	1	1	3	2	1.4	1.87
23	0	0	1	3	2	0	1.2	1.96
24	0	2	1	2	2	—	1.4	2.13
25	5	3	1	0	6	—	3.0	2.34
26	3	0	1	3	6	—	2.6	2.51
27	4	4	5	5	2	—	4.0	2.6
28	1	2	2	3	0	—	1.6	2.59
29	2	2	1	1	4	—	2.0	2.56
30	5	3	1	3	1	—	2.6	2.53
31	4	1	1	4	5	—	3.0	2.46
32	1	2	2	2	6	—	2.6	2.32
33	1	2	3	2	5	—	2.6	2.13
34	2	0	0	1	1	—	0.8	1.93
35	1	1	1	3	3	—	1.8	1.79
36	0	1	1	2	4	—	1.6	1.71
37	3	2	0	1	3	—	1.8	1.65
38	4	2	2	1	1	—	2.0	1.59
39	2	0	2	2	3	—	1.8	1.52
40	1	1	0	1	2	—	1.0	1.47
41	0	0	0	2	2	—	0.8	1.46
42	0	1	0	1	4	—	1.2	1.5
43	4	5	2	2	3	—	3.2	1.54
44	2	0	2	0	2	—	1.2	1.52
45	0	1	2	3	1	—	1.4	1.48
46	3	0	1	0	3	—	1.4	1.46
47	0	1	0	0	1	—	0.4	1.5
48	2	3	2	1	3	—	2.2	1.62
49	2	1	(14)	1	2	—	2.0*	1.74
50	2	1	2	2	3	—	2.0	1.86
51	2	0	0	2	5	—	1.8	1.96
52	(24)	(12)	0	2	7	—	7.0*	7†
<b>Mean</b>	<b>1.5</b>	<b>1.1</b>	<b>1.1</b>	<b>1.7</b>	<b>2.6</b>	<b>—</b>	<b>1.6</b>	<b>1.6</b>
<b>Standard deviation</b>	<b>1.5</b>	<b>1.2</b>	<b>1</b>	<b>1.2</b>	<b>1.7</b>	<b>—</b>	<b>0.8</b>	<b>0.5</b>

**Note:** The three values in parentheses are presumptive outliers and were not included directly in descriptive statistics for the year or in the average.

\* The median was used instead of the mean because of presumptive outliers.

† The value was used in generating counts but was not used to fit the spline.

G-surveillance was applied to the 2002 data. The most noteworthy feature (up to week 23) is the observed counts of 5, 5, and 9 for weeks 19, 20, and 21, respectively, contrasted with expected counts of 1.62, 1.74, and 1.82, respectively. For weeks 19–21, the observed 3-week count is 19, and the expected is 5.18. Assuming use of surveillance with a fixed 3-week window, G-surveillance at week 21 is based on the value  $19 \ln(19/5.18) - (19 - 5.18) = 10.88$ , where  $\ln$  is the natural logarithm. Because this statistic exceeds 5.94, the probability of observing such a substantial 3-week excess during a period of 52 weeks was  $<0.01$ .

P-surveillance (i.e., corresponding to the p-value) for a 3-week window starting at an arbitrary day cannot be determined exactly from this data. Using the weekly tabulations, the p-value is less than or equal to that associated with the 19 events in weeks 19–21. The p-value associated with the 19 cases in weeks 19–21 is

$$[(52/3) (19 - 5.18)] [\exp(-5.18) (5.18)^{19}/19!] = 239.2 (0.00001182) = 0.0004$$

Thus, if surveillance were performed for a year, the chance of finding such a substantial excess, relative to the assumed expected values, is approximately 0.0004. This example is extreme, and no formal analysis might be required. Statistical significance at  $p < 0.05$  would be noted if 14 or 15 cases existed in the 3-week period.

## Discussion and Conclusion

Two surveillance procedures associated with a set error rate over a period T are described. G-surveillance as described is a modification of a statistic used by others and implemented in SaTScan™ software (11,26). The procedure in this report differs from that previously implemented in terms of the function maximized, the events to which type I errors refer, and the logistics of implementation. G-surveillance, as described here, can have different properties by setting T to values of 21–30 days somewhat akin to average run lengths proposed for implementation of CUSUM (7) or setting it to 1 year, which would result in substantially fewer false alarms but decreased sensitivity. A comparison with a statistic (e.g., CUSUM) using both data from real outbreaks and simulated data would identify the properties of the proposed statistics under both abrupt increases and gradual increases. For the latter scenario, CUSUM-like statistics might have superior properties over the methods proposed here.

Apparently, the P-surveillance method is new, but it potentially frees the investigator from performing any simulation. However, three caveats exist, as follows:

1. No reason exists to assume that P-surveillance is better than G-surveillance, although reasons might exist to prefer the reverse on the basis of presumed optimal properties of GLRT.
2. The p-values for P-surveillance are approximate, whereas those for the G-surveillance are exact. (The exactness to a given number of decimal places is attributable to performing enough simulations.)
3. G-surveillance is more flexible, its variants have been described extensively, and it has withstood multiple tests over time.

Nevertheless, the p-value computed by P-surveillance, using either the method described here for continuous surveillance or the ratchet scan for daily or weekly surveillance, should give an overall indication of the likelihood of observing a given excess over expected values in a certain time window, taking into account that the surveillance is performed for a specified period.

#### Acknowledgments

The two authors are listed in order of who presented the poster session on which this paper is based. The authors thank the reviewers; Martin Kulldorff, Ph.D., for his helpful comments; Sam Groseclose, M.D., David Williamson Ph.D., and Willie J. Anderson of CDC for providing the brucellosis data; and Elana Silver, M.S., for her comments on an early draft.

#### References

1. Mandl KD, Overhage JM, Wagner MM, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. *J Am Med Inform Assoc* 2004;11:141–50.
2. Lazarus R, Kleinman K, Dashevsky I, et al. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerg Infect Dis* 2002;8:753–60.
3. Williamson GD, Hudson GW. A monitoring system for detecting aberrations in public health surveillance reports. *Stat Med* 1999;18:3283–98.
4. Stroup DF, Williamson GD, Herndon JL. Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Stat Med* 1989;8:323–9.
5. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting *Salmonella* outbreaks. *Emerg Infect Dis* 1997;3:395–400.
6. Rogerson P. Monitoring point patterns for the development of space-time clusters. *J R Stat Soc A* 2001;164:87–96.
7. Sonesson C, Bock D. A review and discussion of prospective statistical surveillance in public health. *J R Stat Soc A* 2003;166:5–12.
8. Frisén M. Statistical surveillance, optimality and methods. *Int Stat Rev* 2003;71:403–34.
9. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Stat Med* 1995;14:799–810.
10. Farrington CP, Andrews NJ, Beale AA, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. *J R Stat Soc A* 1996;154:61–74.
11. Kulldorff M. A spatial scan statistic. *Communications in Statistics—Theory and Methods* 1997;26:1481–96.
12. Kulldorff M, Feuer E, Miller B, Freedman L. Breast cancer in Northeastern United States: a geographical analysis. *Am J Epidemiol* 1997;146:161–70.
13. Kulldorff M, Athas WF, Feuer E, Miller B, Key C. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos. *Am J Public Health* 1998;88:1377–80.
14. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *J R Stat Soc A* 2001;164:61–72.
15. Loader C. Large deviation approximations to the distribution of scan statistics. *Advances in Applied Probability* 1991;23:751–71.
16. Nagarwalla N. A scan statistic with a variable window. *Stat Med* 1966;15:845–50.
17. Naus J. The distribution of the size of the maximum cluster of points on a line. *J Am Stat Assoc* 1965;60:532–38.
18. Glaz J, Naus J, Wallenstein S. Scan statistics. New York, NY: Springer Verlag, 2001.
19. Glaz J, Balakrishnan N, eds. Scan statistics and applications. Boston, MA: Birkhäuser, 1999.
20. Naus J. Some probabilities, expectations, and variances for the size of the largest clusters and smallest intervals. *J Am Stat Assoc* 1966;61:1191–9.
21. Glaz J. Approximations and bounds for the distribution of the scan statistic. *J Am Stat Assoc* 1989;84:560–9.
22. Wallenstein S, Neff N. An approximation for the distribution of the scan statistic. *Stat Med* 1987;6:197–207.
23. Weinstock M. A generalized scan statistic for the detection of clusters. *Int J Epidemiol* 1981;10:289–93.
24. Wallenstein S, Weinberg C, Gould M. Testing for a pulse in seasonal event data. *Biometrics* 1989;45:817–30.
25. Chang M, Glynn MK, Groseclose SL. Endemic, notifiable bioterrorism-related diseases, United States, 1992–1999. *Emerg Infect Dis* 2003;9:556–64.
26. National Cancer Institute. SaTScan version 3.0: software for the space and space-time scan statistics. Bethesda, MD: National Cancer Institute, 2003.