Comparison of Aberration Detection Algorithms for Biosurveillance Systems

Hong Zhou*[1], Howard Burkom[2], Carla Winston[3] and Umed Ajani[1]

[1]CDC, Atlanta, GA, USA; [2]John Hopkins Applied Physics Laboratory, Washington DC, MD, USA; [3]Veterans Health administration, Palo Alto, CA, USA

Objective

We compared several aberration detection algorithms using a set of syndromic data streams from a large number of treatment facilities in the CDC Biosense 1.0 system. A realistic signal injection strategy was devised to compare different ways of adjusting for total facility visits and background day-of-week effects.

Introduction

Use of robust and broadly applicable statistical alerting methods is essential for a public health Biosurveillance system. We compared several algorithms related to the Early Aberration Reporting System C2 (adaptive control chart) method for practical detection sensitivity and timeliness using a realistic but stochastic signal inject strategy with a variety of data streams. The comparison allowed detail examination of strategies for adjusting daily syndromic counts for day-of-week effects and the total daily volume of facility visits. Adjustment for the total visit volume allows monitoring of surrogate rates instead of just counts, and the use of real data with both syndromic and total visit counts enables this adjustment.

Methods

We used daily syndrome counts aggregated by facility as baseline data. We included sampled facilities with data reported every day from Jan 2010 through May 2011 and with mean daily syndromic counts >=1 over the entire study period. Syndrome groups chosen were gastrointestinal (GI), rash, and upper respiratory infection (URI). Multi-day signals were randomly drawn from a lognormal distribution whose value on the peak day was twice the background standard deviation and such that the signal duration ranged from 4-10 days. Injected counts were applied to both the syndromic and total visit time series for accurate assessment of the total visit adjustment. At a background alert rate of 1%, we calculated (1) sensitivity, the ratio of number of events detected before the distribution peak to the total number of injected signals; and (2) timeliness as the number of days from the start of injection to the first algorithm alert. We compared five algorithms with 28 baseline days and 2-day sliding buffer. Each algorithm was tested with and without stratification by weekday versus weekend/holiday.

1. C2Count: count based C2
2. CSCount: count based CUSUM [1]
3. C2Rate: C2 with count adjusted for total visit [2]
4. C2Ratio: percentage (counts / total visits) based C2
5. CSRate: CUSUM with count adjusted for total visits

Results

For GI, rash, and URI, respectively, the mean observed daily counts per facility were 15.6, 2.8, and 4.4 during the week days, and 2.6, 0.4, and 1.2 during the weekends/holidays. The modified CUSUM algorithm (CSRate) had the highest average sensitivity and timeliness (i.e., shortest time to detection) for all three syndromes regardless of the weekday stratification (Table). With very few exceptions C2Rate, which also adjusted for total visits in the algorithm, performed better than C2Count, CSCount, and C2Ratio. The overall sensitivity and timeliness increased substantially after accounting for weekdays and weekends for all five algorithms and all three studied syndromes.

Conclusions

These analyses on three syndromic data demonstrate that accounting for total facility visits and weekday/weekends improves sensitivity and timeliness in detecting multi-day outbreaks. A modified CUSUM algorithm outperforms many currently used C2 methods. Further analyses using time-series from additional syndromes are needed.

Table: Sensitivity (%) and timeliness (Day) to detect of multi-day signal

Criteria	Wk/end stratified	Syndrome	C2Count	CSCount	C2Rate	C2Ratio	CSRate
Sensitivity	No	GI	26.5	27.0	40.5	34.0	42.3*
		Rash	22.9	24.4	30.9	28.5	34.9*
		URI	34.5	33.3	44.1	27.3	46.1*
	Yes	GI	51.5	50.8	54.1	45.8	56.6*
		Rash	46.0	45.7	48.7	39.8	49.1*
		URI	43.9	43.8	45.9	35.7	48.1*
Timeliness	No	GI	3.25	3.20	3.07	3.09	2.98*
		Rash	2.81	2.76	2.72	2.68	2.62*
		URI	2.89	2.86	2.77	2.93	2.65*
	Yes	GI	2.84	2.82	2.81	2.95	2.71*
		Rash	2.47	2.46	2.43	2.57	2.41*
		URI	2.74	2.69	2.71	2.86	2.62*

* Best among five algorithms

Keywords

BioSense; Aberration Detection; Control Chart

References

1. Fricker RD, Jr., Hegler BL, Dunfee DA. Comparing syndromic surveillance detection methods: EARS' versus a CUSUM-based methodology. Stat Med 2008; 27:3407-29.
2. Tokars JI, Burkom H, Xing J, et al. Enhancing time-series detection algorithms for automated biosurveillance. Emerging Inf Dis 2009;15:533-9.

*Hong Zhou
E-mail: fwd6@cdc.gov