

Increasing Outbreak Detection Power by Data Transformations

Tom Andersson^{*1,2}, Pär Bjelkmar³ and Joanna Tyrcha²

¹Swedish Civil Contingencies Agency, Karlstad, Sweden; ²Stockholm university - Mathematical Statistics, Stockholm, Sweden;

³Swedish Institute for Communicable Disease Control, Stockholm, Sweden

Objective

For the purpose of optimizing baselines for point-source outbreak detection, we carried out a power analysis of the effects of data transformations. More specifically, the aim was to develop statistical criteria for using composite baselines, i.e. ratios and geometric means of data streams. The results were validated by outbreak data on acute gastroenteritis (The Swedish National Telephone Health Service 1177).

Introduction

Syndromic surveillance involves the analysis of time series of health indicators to identify changes in disease patterns. To this end, statistical modeling is used to reduce systematic data variation. Still, there is variation that cannot be accounted for in this approach, e.g. mass gatherings, extreme weather and other high-profile events. To filter sporadic events, data transformation can be applied, e.g. proportion data from correlated data streams (Peter, Najmi and Burkom, 2011; Reis, Kohane and Mandl, 2007). However, we lack systematic criteria for applying data transformations, e.g. ratios versus geometric means. To develop guidelines, we conducted a power analysis and compared the results with empirical findings (Andersson et al, 2013).

Methods

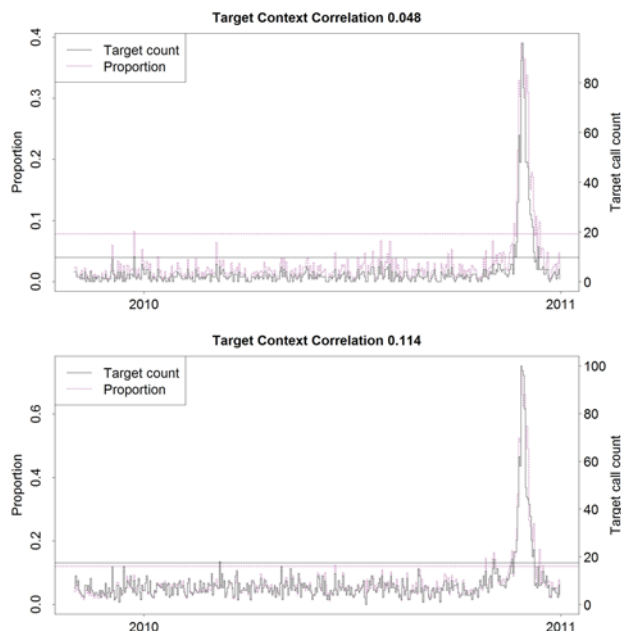
We conducted simulated ANOVAs of detection power. The gamma distribution was used to generate baseline and outbreak distributions: target T , context C and outbreak O . T and C defined three baselines (B): (1) single data ($B=T$), (2) ratio of correlated data ($B=T/C$) and (3) geometric mean of uncorrelated data ($B=\sqrt{T*C}$). For baselines 1 and 2, outbreak values were added pointwise to T , not C , representing a disease anomaly. For baseline 3, outbreak points were applied to both T and C , representing a sporadic effect. Power was defined as the proportion of the data points $BO=B+O$ exceeding a predefined false alarm rate (alpha-level) applied to B . We studied the effects of eight factors in a $3^4 \times 2^4$ design using 10 000 data points for each combination: means and variances of T , C and O , correlation between T and C , and alpha-level. The results were compared with signal-to-noise ratios in triage data.

Results

Baseline 2, compared with 1, increases the detection power in proportion to the correlation between T and C . A correlation $< 0.1-0.2$ gives higher power for baseline 1, validated by empirical findings (cf. figure). The power effect of correlated data is attenuated for larger target distributions. Baseline 3, compared with 1, increases the power in proportion to outbreak size. The effect also increases for larger, asymmetric, target and context distributions.

Conclusions

Data transformations can be used to increase outbreak detection power. Ratios of correlated data have increased sensitivity to moderate point-source outbreaks of specific syndromes. Geometric means of uncorrelated data can be of value to control for sporadic high-impact events that do not represent specific syndromes.



Syndromic data from Swedish Health Care Direct 1177 regarding an outbreak of Cryptosporidiosis in Östersund, Jämtland, Sweden, November 2010. At top, number and proportion of calls on adult diarrhea from Jämtland county. At bottom, number and proportion of calls on diarrhea and general medical inquiries from Östersund. Correlations based on initial 300 days; horizontal lines representing $\alpha = 0.001$. Higher correlation between target (specific syndrome) and context (other calls) increases detection power.

Keywords

outbreak detection; power analysis; data transformation; baseline noise; variation

Acknowledgments

The study is funded by the Swedish Agency for Contingency Planning (MSB).

References

- Andersson T, Bjelkmar P, Hulth A, Lindh J, Stenmark S, Widerström M. Syndromic surveillance for local outbreak detection and awareness: evaluating outbreak signals of acute gastroenteritis in telephone triage, web-based queries and over-the-counter pharmacy sales. *Epidemiology & Infection*. 2013; 15:1-11
- Peter W, Najmi AH, Burkom HS. Reducing false alarms in syndromic surveillance. *Statistics in Medicine*. 2011; 30;30(14):1665-77.
- Reis BY, Kohane IS, Mandl KD. An epidemiological network model for disease outbreak detection. *PLoS Medicine*. 2007;4(6):e210.

*Tom Andersson

E-mail: tom.andersson@msb.se

