Emerging Health Threats Journal 2011, **4**:s13. **doi:** 10.3134/ehtj.10.013

www.eht-journal.org

ABSTRACT

Sample size and spatial cluster detection power

JF Conley

Department of Geology and Geography, West Virginia University, Morgantown, WV, USA
E-mail: Jamison.Conley@mail.wvu.edu

Objective
In syndromic surveillance settings, the use of samples may be unavoidable, as when only a part of the population reports flu-like symptoms to their physician. Taking samples from a complete population weakens the power of spatial cluster detection methods.[1] This research examines the effectiveness of different sampling strategies and sample sizes on the power of cluster detection methods.

Introduction
Prior work demonstrates the extent to which sampling strategies reduce the power to detect clusters.[1] Additionally, the power to detect clusters can vary across space.[2] A third, unexplored, effect is how much the sample size impacts the power of spatial cluster detection methods. This research examines this effect.

Methods
The same six simulated clusters in the Pittsburgh, PA area are used as in [1]. A total of 1000 samples of three different strategies (random, stratified and case–control) were taken. Furthermore, 1000 independent random samples of sizes ranging from 1 to 50% of the complete population were taken. All sample data sets and the complete data sets were analyzed using FleXScan (http://www.niph.go.jp/soshiki/gijutsu/download/flexscan/),[3] which has been proven effective at finding arbitrary shaped clusters.[1,4] The detected clusters were then evaluated with a weighted power statistic that assesses the amount of overlap between the detected cluster and the actual cluster.[4]

Results
Figure 1 shows the relative performance of the different sampling strategies using FleXScan. Case–control sampling is clearly the best method, followed by stratified sampling and random sampling.

Figure 2 shows the effect of sample size on the weighted power of FleXScan when using random samples. As the chart shows, the weighted power increases rapidly up to a sample size of 15–20% of the complete population, and increases slowly thereafter.

Conclusions
These results demonstrate some of the impacts of the use of samples to detect spatial clusters. First, the method of sampling is important, as case–control sampling is more effective than random sampling and stratified sampling. However, in a syndromic surveillance situation, case–control data may be

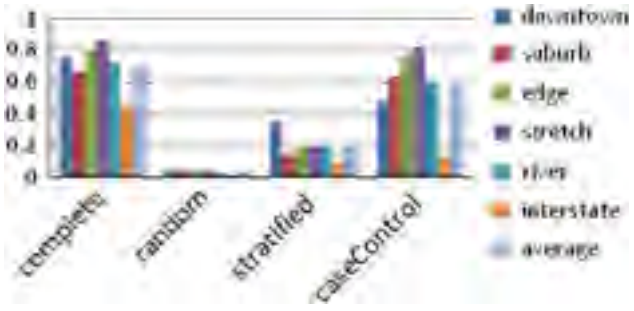

Figure 1 Weighted powers of different sampling strategies for FleXScan across all cluster shapes.

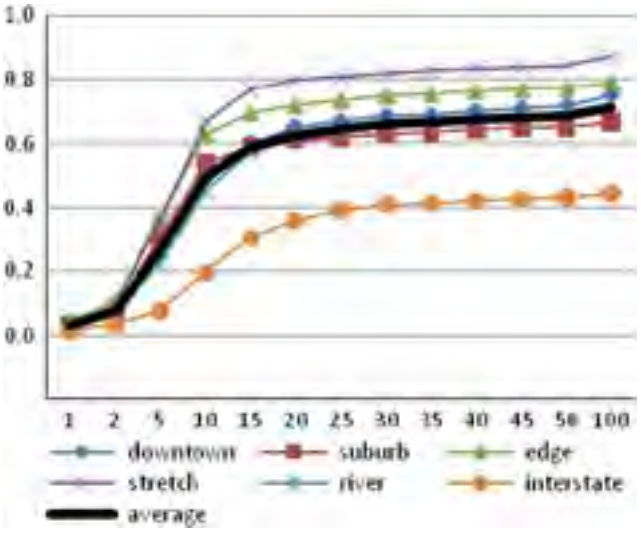

Figure 2 Weighted power versus sample size for FleXScan using random sampling.

unavailable, and the only available data may be considered as a random sample, such as when relying upon patients to report symptoms to their physicians. In this situation, a random sample is almost as effective at accurately detecting the shape of a disease cluster when the sample is more than 15% of the population.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

1 Conley JF. An evaluation of the extended power of cluster detection methods under sampling strategies. *Proc of Geo-Computation 2009*, Sydney Aus. http://www.biodiverse.unsw.edu.au/geocomputation/proceedings/PDF/Conley.pdf Accessed Sep 2 2010.
2 Waller LA, Hill EG, Rudd RA. The geography of power: Statistical performance of tests of clusters and clustering in heterogeneous populations. *Statist Med* 2006;**25**:853–65.
3 Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geog* 2005;**4**:11.
4 Takahashi K, Tango T. An extended power of cluster detection tests. *Statist Med* 2006;**25**:841–52.