

What Can You Really Do with 35,000 Statistical Alerts a Week Anyways?

Michael Coletta, Hong Zhou

CSELS/DHIS, Centers for Disease Control and Prevention, Atlanta, Georgia, United States

Objective

Find practical ways to sort through statistical noise in syndromic data and make use of alerts most likely to have public health importance.

Introduction

The National Syndromic Surveillance Program's (NSSP) instance of ESSENCE* in the BioSense Platform generates about 35,000 statistical alerts each week. Local ESSENCE instances can generate as many as 5,000 statistical alerts each week. While some states have well-coordinated processes for delegating data and statistical alerts to local public health jurisdictions for review, many do not have adequate resources. By design, statistical alerts should indicate potential clusters that warrant a syndromic surveillance practitioner's time and focus. However, practitioners frequently ignore statistical alerts altogether because of the overwhelming volume of data and alerts. In 2008, staff in the Virginia Department of Health experimented with rules that could be used to rank the statistical output generated in ESSENCE alert lists. Results were shared with Johns Hopkins University Applied Physics Lab (JHU/APL), the developer of ESSENCE, and were early inputs into what is now known as "myAlerts," an ESSENCE function that syndromic surveillance practitioners can use to customize alerting and sort through statistical noise. NSSP-ESSENCE produces a shared alert list by syndrome, county, and age-group strata, which generates an unwieldy but rich data set that can be studied to learn more about the importance of these statistical alerts. Ultimately, guidance can be developed to help syndromic surveillance practitioners set up meaningful ESSENCE myAlerts effective in identifying clusters with public health importance.

Methods

The region/syndrome alert list generated from NSSP's instance of ESSENCE on the BioSense Platform was downloaded and ranked based on five criteria:

1. Observed count causing the alert
2. Expected count generated by ESSENCE
3. Total number of alerts for that syndrome in that county and number of prior alerts during that week for the same syndrome, county, and age group
4. Density of alerts during the prior week
5. Recency of the latest alert

Alerts were then ranked based on:

1. Higher absolute counts (regardless of expected value)
2. Higher partial chi-square, $(\text{Obs-Exp})^2 / \text{Exp}$
3. Higher total alerts for a given county/syndrome
4. Higher number of earlier alerts for same county/syndrome/age group
5. Multiple alerts same day > alerts on consecutive days > alerts separated by days without alerts
6. Alerts present on more recent days



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

The top 20 alerts with the highest scores were then reviewed and if anything unusual was noticed (i.e. problems unrelated to recent data quality problems or onboardings, seasonal trends, etc.) then there was follow-up with the site. The alert list rankings were then evaluated for differences among factors available in the ESSENCE myAlert function. We compared the top 5% of ranked alerts to the remaining 95% to determine if there were significant differences in the following factors:

1. Total number of alerts across six age groups (including all ages) within 8 days of each syndrome and county stratum;
2. Average alert frequency across six age groups (including all ages) within 8 days for each stratum;
3. Average count across the strata;
4. Average expected value across the strata;
5. Average of the difference between the count and expected values for each stratum; and
6. Average Level across the strata.

Results

Preliminary interactions with sites revealed important clusters – some already known and some not. For example, a cluster of healthcare workers exposed to *Neisseria meningitidis*, and kids exposed to a bat at summer camp and presenting for prophylaxis were among the clusters identified. Additionally there were differences seen in the adjustable myAlert parameters when comparing the top 5% to the lower 95% of ranked alerts.

Conclusions

The differences seen and preliminary feedback suggests that this ranking method may be effective in identifying alerts representing true clusters of public health importance. Testing designed to evaluate myAlert parameters based on the differences seen in the top 5% of ranked alerts is underway in sites where more detailed data access is available. More study is needed; however, there are indications that cutoff values for these parameters may be a valuable way for syndromic surveillance practitioners to reduce the review burden and focus on the most important statistical clusters identified by ESSENCE statistical algorithms.

Acknowledgement

Jeremy Smith formerly of the Virginia Department of Health

*ESSENCE stands for the Electronic Surveillance System for the Early Notification of Community-based Epidemics and is designed by Johns Hopkins University Applied Physics Laboratory.

Table 1

Category	Total Alerts	Daily Average Count	Daily Average Expected	Daily Average Difference (Count-Expected)	Daily Average Level	Average Alerts Across 8 Days
Bottom 95%	31265	5	3	3	0.0207	5
Top 5%	1684	27	17	10	0.0036	8



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons AttributionNoncommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.