CONTACT INFORMATION

Submitter name:	David Atrubin
Jurisdiction or affiliation:	Florida Dept. of Health
Phone:	
Email:	david.atrubin@flhealth.gov
Co-submitters and affiliations:	

PROBLEM DESCRIPTION

a. A truncated historical dataset is provided from one or more subregions with multiple participating hospitals with enough variety in the patient volume and demographics to make the problem challenging and to generate alerting solutions useful to other regions.

b. Records from the truncated dataset consist of date and time of visit, patient residence zip code, and consistent codes for the different hospitals without naming them. (Other demographic fields would be helpful but not essential and not worth holding up the process.) At least one year of data would be needed; with 2 or more years, the first years' data with all seasons could be used for training, the last year's data for testing.

c. The health department would have to provide an explicit list of in-area zip codes. Depending on their vision for using the method, they might provide more than one list (just the estimated catchment zips for each hospital, or the entire county, or the entire state, or all bordering states, or whatever fits the monitoring concept). Each list then defines "out-of-area" zip codes.

d. The grouping of hospitals for analysis is important. Data could be pooled by individual hospital, hospital group, county, health district,…. Either the health department should specify 1 or more levels of grouping, or that could be left to the developer with certain requirements.

e. The developer's task would be 1) to analyze the out-of-area zip distribution in the provided dataset, stratifying by whatever the data analysis shows to be useful: month, season, weekday/weekend, or whatever seems best. (The health department may want to provide guidelines or advice.), and 2) use this analysis to develop and tune an alerting method to identify groups of ED records representing anomalous sets of "out-of-area" zip codes for possible further investigation.

SOLUTION REQUIREMENTS

Describe the type of solution you are seeking (e.g., anomaly detection, signal validation, data quality characterization):

Using only the patient record fields available in the benchmark dataset, the solution will identify unusually large groups of patient records with out-of-area zip codes. Developers will determine the customary out-of-area patient distribution from the benchmark data and may consider factors related to a convention or other event of interest, weekly patterns, seasonal patterns, and any other factors from the data that will help identify unusual zip-code groups while maintaining a manageable alert rate.

For a facility serving mainly neighborhood patients, a relatively small number of out-of-area patients may be significant, but significance at a larger referral facility may require a much higher number. These sets of records should be identified on a daily basis, and more often on demand with a rapid query to enable the health department to inspect the full records with all fields for possible further investigation. Summaries of the out-of-area zip code locations, the facility identifiers, the dates and times of the interval containing these records, and any other relevant information are needed.

Describe what type of solution would enable you to implement it in your practice setting (e.g., Do you need an algorithm? Do you need code? If you need code, does it have to be written in any particular programming language?).

Queries/code that would allow the user to easily extract and visualize these data for this subset of out of area visitors. Algorithms that would alert the user to statistical anomalies in this subset.

Describe who will use the solution. For example, how many users will there be and what level of skill do the users have? Are the users all within a single jurisdiction/organization?

The solution would benefit epidemiologists responsible for conducting disease/injury surveillance for special events. The users will need data analysis skills, experience in interpreting syndromic surveillance data, and experience in responding to statistical anomalies and ED/UCC visits of PH importance.

Note any other constraints:

Special events differ in terms of the length of the event, the percentage increase in an area's population, the potential for intentional disruption (including terrorism), and possibly the way in which the EDs/UCCs are typically utilized within that community. These all provide additional challenges for finding a solution that has generalizability.
Operational constraints are that user operational and alert burden must be kept to a minimum, and the interesting groups of patient records must be identified using on-demand queries typically requiring less than a minute.

VALIDATION

Does a gold standard exist with which to validate the proposed solutions?

☐ Gold standard exists within the provided data set (e.g., an outbreak signal nested within baseline data)

☐ Gold standard exists in a separate data set, which can be provided to the workgroup (e.g., laboratory data to validate ED data)

☐ Gold standard exists but cannot be furnished

☐ Gold standard does not exist

INPUT DATA

List the minimum data elements that can be provided to address the problem:
Minimum data elements: date and time of visit, patient residence zip code, and consistent codes for the different hospitals without naming them, but not any diagnostic or symptom-related categorical or free-text information. Demographic and any other elements beyond this minimum will be supplied considering the trade-off between surveillance usefulness and any perceived disclosure risk.
Another essential requirement is at least one list of zip-codes to be treated as "in-area" for patient residences.

How much historical data can be provided?

Describe any restrictions for sharing the data:

Note any other relevant data characteristics:

Suspected Rabies Exposure: case defining variables are in **bold.** Answers are: Yes, No, Unknown to case, Not asked /Not answered

OUTPUT DATA

NOTES