# NC DETECT

## CONTACT INFORMATION


| | |
|---|---|
| **Submitter name:** | Amy Ising |
| **Jurisdiction or affiliation:** | Carolina Center for Health Informatics, NC DETECT, UNC Chapel Hill Department of Emergency Medicine |
| **Phone:** | 919-966-8853 |
| **Email:** | ising@ad.unc.edu |
| **Co-submitters and affiliations:** | Lana Deyneka & Jenna Waggoner, North Carolina Division of Public Health<br>Anna Waller, CCHI, NC DETECT |
| **Date / Version Information:** | Version 4; March 11, 2013 |


## PROBLEM DESCRIPTION

**Summarize the problem:**
A syndrome cannot be created to identify every possible cluster of potential public health significance. A method is needed to identify clusters without pre-classification into syndromes. This could include clusters of signs or symptoms, clusters of place names (e.g. mentioning a specific restaurant), clusters of events (e.g. mentioning a specific fair, concert, etc.).

PLEASE SEE EXAMPLE ON PAGE 3.

How this issue is currently being addressed:

North Carolina has implemented a partial solution that uses Time of Arrival Analysis (TOA) developed by the Johns Hopkins Applied Physics Lab to detect clusters of interest based solely on visit counts (no syndromes) (Li et al, 2013). TOA detects clusters based on arrival date and time. The user can then scan the line listing of these time-based clusters to see if the line listing data reveal any commonalities among the visits that might be of public health significance. This can be time-consuming depending on the number of clusters detected by TOA. In addition, this method will not detect clusters of events that occur over a longer time period. For example, if 5 people come into an emergency department complaining about "Larry's hot dog shack" but these visits are scattered over a 24-hour period, TOA will not group these into a cluster.

## SOLUTION REQUIREMENTS

**Describe the type of solution you are seeking (e.g., anomaly detection, signal validation, data quality characterization):**
The tool would be a language parser that works in concert with a cluster detection tool. (Acquiring the Time of Arrival analysis algorithm would be a separate process and is outside the scope of this problem).

We are looking for a method to find a group of ED visits worthy of follow-up based on a link identified in the chief complaint strings. The priority is to identify clusters of public health significance that would not be detected in currently existing syndromes, e.g. gastrointestinal, fever/rash, meningo-encephalitis, etc. We want to be able to identify clusters that we would never think of actively monitoring (e.g. Larry's hot dog shack).

**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?).**
The primary requirement is that the solution needs to run in a Windows environment. The solution can be a newly developed tool or a compilation of commercial off the shelf tools. Any newly developed tool should be made available at no cost to interested parties. Preference is Java-based or something that is interoperable with Java.

The solution could work with or without the TOA algorithm. If the solution is designed to work with TOA, then any agency implementing this would need the TOA code from JHUI APL. Absolutely needs to run in a Windows environment.

**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 intention would be to embed the solution in NC DETECT (and any other system with an interest in this tool). NC DETECT serves 200+ active registered users across North Carolina. Our intention would be to run this tool on the emergency department data that are received and processed twice daily. For that to happen, the tool's performance could not negatively impact our data processing times in a significant way and the tool could not require systems and computer processing power not typically seen in health department environments.

**Note any other constraints/ assumptions:**
Tools developed to solve this problem using data from NC DETECT should be open source or made available at no cost.

## VALIDATION

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

X Gold standard exists within the provided data set (e.g., an outbreak signal nested within baseline data) – we would inject sample clusters into the dataset
- The data providers can inject clusters into the dataset and can work with the solution provider to fine tune these as needed

☐ 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:**
- Arrival date and time
- Masked facility ID
- Age group
- Chief complaint that has been processed by select modules of EMT-P (to standardize abbreviations and misspellings)
- 100 Injected test clusters

**How much historical data can be provided?**

To be determined after consultation with the data provider. A minimum of 100,000 records will be provided.

**Describe any restrictions for sharing the data:**

Anyone accessing the data will need to sign a Data Use Agreement with the North Carolina Division of Public Health (Communicable Disease Branch)

For more information: https://ncdetect.org/data-reports/data-requests

Each researcher requesting access to the data will need to sign the Data Use Agreement (DUA). The DUA specifies the recipients of the data, the data elements received and the purpose of the research. Anytime a researcher is added to the group or a new dataset is requested, the DUA is amended. The data can be used only for the specific research specified in the DUA.

**Note any other relevant data characteristics:**

While triage notes cannot be provided for this problem, we would hope that the solution could be applied to triage notes as well as chief complaints.

## OUTPUT DATA

## NOTES

NC DETECT Problem Example: A syndrome cannot be created to identify every possible cluster of potential public health significance. A method is needed to identify clusters without pre-classification into syndromes. We would still like a way to identify clusters like the one highlighted below without having a pre-defined syndrome that specifically looks for "Larry's Hot Dog shack."

| Arrival Date & Time | Location | Age Group | Chief Complaint | Existing NC DETECT syndromes this record will be binned into |
|---|---|---|---|---|
| 1/5/2013 10am | X | 0-4 | Fever. Cough, congested, diarrhea | GI ALL, GI Severe, Respiratory, Resp ALL |
| 1/5/2013 10:30 am | X | 25-44 | Bodyaches fever cough | ILI, Respiratory, Resp ALL |
| 1/5/2013 11 am | X | 5-9 | FEVER,VOMITTING,CHILLS,SORE THROAT,HEADACHE,DIARRH | GI ALL, GI Severe, ILI, MenEnc |
| 1/5/2013 11 am | X | 65+ | Multiple medical complaints | |
| 1/5/2013 11:15 am | X | 45-64 | Carbon dioxide exposure | Carbon Monoxide Poison (keyword) |
| 1/5/2013 11:16 am | X | 45-64 | CO | Carbon Monoxide Poison (keyword) |
| 1/5/2013 12:00 pm | X | 45-64 | Possible carbon monoxide poisoning | Carbon Monoxide Poison (keyword) |
| 1/5/2013 1:00 pm | X | 10-14 | Headache, fever | MenEnc |
| 1/5/2013 1:05 pm | X | 25-44 | sick ate at Larry's Hot Dog shack | |
| 1/5/2013 1:10 pm | X | 25-44 | Not feeling well since Larry's hot dog shac | |
| 1/5/2013 1:15 pm | X | 25-44 | Sick person (larry's hot dog) | |
| 1/5/2013 1:20 pm | X | 45-64 | Multiple complaints (larry's hot dog shacl) | |
| 1/5/2013 1:22 pm | X | 19-24 | Bath salts | Bath salts |
| 1/5/2013 1:24 pm | X | 19-24 | OD – bath salt | Bath salts |
| 1/5/2013 3:00 pm | X | 5-9 | SOB | Resp ALL |
| 1/5/2013 3:16 pm | Y | 5-9 | SHOB, wheezing | Resp ALL |
| 1/5/2013 4:00 pm | Y | 5-9 | Asthma exacerbation | Asthma (keyword) |
| 1/5/2013 5:00 pm | Y | 10-14 | Asthma exacerbation | Asthma (Keyword) |
| 1/5/2013 6:00 pm | Y | 25-44 | Sickle cell crisis | |
| 1/5/2013 7:00 pm | Y | 65+ | Sick person | |
| 1/5/2013 8:00 pm | Y | 65+ | Done fell out | |
| 1/5/2013 8:00 pm | Y | 10-14 | Fever, rash | Fever Rash Illness |
| 1/5/2013 8:05 pm | Y | 65+ | fall | |
| 1/5/2013 8:10 pm | Y | 25-44 | MVC | MVC |