

# Evaluating the Validity of ED Visit Data for Biosurveillance

Aaron Ratcliffe MS, Clifton Barnett MSIS, Amy Ising MSIS, Anna Waller ScD  
*University of North Carolina, Chapel Hill, Dept. of Emergency Medicine*

## Objective

This study evaluates the validity of a subset of ED data collected in NC DETECT, as well as measures the effectiveness of the data quality processes in place for this surveillance system [1].

## Background

The North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) receives a designated set of data elements electronically available from 110 emergency departments (EDs) (98%) on at least a daily basis via a third party data aggregator. While automated processes monitor for data quality problems such as improper file formats or missing required elements, data corruption can occur at several stages before receipt, and if undetected, data can appear reliable. Hospitals might map to standard codes incorrectly, data aggregators might manipulate text improperly, or updates might be confused with original records. These inaccuracies cause delays and oversights in identifying events of public health importance.

## Methods

This study encompassed one mid-sized, rural community hospital (Hospital L) and one large, urban hospital system (3 EDs, System S) in a full data validity audit. The date range for Hospital L was Feb. 23<sup>rd</sup> – Mar. 7<sup>th</sup>, 2008 and for System S was Mar. 1<sup>st</sup> – 14<sup>th</sup>, 2008. The audit identified 20 data elements for comparison: Initial ED Acuity Assessment; Address (city, state, zip); Arrival Date and Time; Chief Complaint; Date of Birth; ICD-9-CM Final Diagnosis Codes (n=4); ED Disposition; select Initial Vitals (Temperature Read/Route, Blood Pressure); Medical Record Number; Insurance Coverage (financial class); Patient Sex; Transport Mode to ED.

The study gathered electronic data directly from each hospital's information system or from information systems directors within the participating hospitals. Random sampling verified these electronic data against the hospitals' electronic charts. NC DETECT data was mapped to the hospital data to assess accuracy and completeness. Accuracy rates across fields, as well as individual records, were compared to established business rules for data quality. Some data elements required crosswalks to be collected in order to map hospital data to standard codes in NCDETECT. Further investigation determined sources of problems and correlations among inaccurate data elements, when necessary.

## Results

The study examines 9,144 visit records, 1,717 visits from Hospital L and 7,427 visits for System S.

For Hospital L, 11 data fields match with high accuracy and completeness: Acuity (99.1%); Address (97.4%, 99.9%, 96.3%); Arrival Date and Time (99.1%); Chief Complaint (96.1%); Date of Birth (99.9%); ED Disposition (98.1%); Medical Record Number (99.3%); Patient Sex (100%); Transport Mode (100%). Hospital L does not submit First Vitals to NC DETECT. Analysis of Diagnosis Codes and Insurance Coverage awaits receipt of source data and will be complete by September 2008. All records match at least 8 fields; 97.1% at least 10 fields, and 81.6% match all 11 collected fields.

For System S, 16 data elements match with high accuracy and completeness: Acuity (98.6%); Address (97.0%, 99.3%, 98.2%); Arrival Date and Time (99.8%); Chief Complaint (99.5%); Date of Birth (99.6%); ED Disposition (97.7%); First Vitals (99.7%, 89.9%, 99.3%, 99.4%), Medical Record Number (99.8%); Insurance Coverage (89.7%); Patient Sex (99.9%); Transport Mode (99.0%). Initial diagnosis codes (4) currently supplied match at a lower rate (60.6%, 57.1%, 45.8%, 37.3%) due to coder updates in the final codes collected by NC DETECT. Further analysis awaits receipt of correct billing system codes but will be complete by September 2008. 99.5% of all records match at least 13 fields while 87.5% match all 16 relevant fields.

## Conclusions

This study shows high match rates for all accessible data elements for the given time frame. Final diagnosis codes have yet to be audited. Comparison of initial ED diagnosis codes and final billing codes presents opportunity for future research.

The validity of secondary data used in syndromic surveillance systems is difficult to assess without an audit which utilizes source data. An audit is costly in terms of time and money so is rarely done, even though data quality problems can seriously compromise the efficacy of a surveillance system. Methods developed by this data validity audit may be useful to other systems, especially those expanding the scope of data elements collected from EDs.

Beyond data accuracy and completeness, timeliness is also critical [2]. Timeliness is addressed through NC DETECT's ongoing data quality efforts.

## References

- [1] Barnett, C., Ising, A., Travers, D.A., Waller, A. Emergency Department Data Quality Best Practices. *Advances in Disease Surveillance* 2006; 2:193.
- [2] Berenbaum, P., Gibson, M., Blevins, L. Accuracy versus Timeliness for Influenza Detection: A Comparison of Hospital Syndromic Surveillance Data with Discharge Data. *Advances in Disease Surveillance* 2007; 2:93.