Title: Early feedback from a pilot of a cognitive computing system to analyze immunization data

CDC Authors: Stacie Greby, Alexandra Bhatti, Allison Fisher, Yoonjae Kang, Cynthia Knighton, Pamela Srivastava

Abt Authors: Sarah Ball, William Campbell, Alison Thaung

Temple Authors: Marija Stanojevic, Zoran Obradovic

Topic Areas:

3. Programmatic Issues or 4. Health and Risk Communications

Word Count: 299/300 allowed

Background:

The goal of immunization programs is to maintain or improve vaccination coverage to prevent diseases. While quantitative and qualitative data are available to support this goal, qualitative data are not always actionable because of labor intensive analytic methods. Cognitive computing systems can address this challenge by leveraging automation for quantitative and qualitative data analysis.

Objectives:

To describe the development of a cognitive computing system to analyze immunization data.

Methods:

Text data from a variety of formal and informal sources were collected to develop a lexicon, or local language, to describe immunization programs. Formal data included program reports, legal and scientific text searches, and vaccine-related textbooks. Informal data included Sysomos searches of Twitter feeds related to immunization, (e.g., recent vaccine-preventable disease outbreaks). The data were searched with a cognitive computing system using multiple algorithms.

Results:

The formal and informal data effectively created two databases. The formal database is a sample of 64, one entry for each immunization program awardee. The informal database is a sample of over 64 million. Location information is available for up to 60% of the Twitter data, related to the user-profile, not the location of the tweet. Challenges identified by this pilot project included: data structure, classification of misinformation, and language relevance (e.g., MMR can mean measles, mumps, and rubella vaccine or Match Making Rank, a rating system used in online gaming.). Potential solutions include: sorting by type of group tweeting and retweeting the information to help identify sources of misinformation.

Conclusions: The pilot system can quickly search formal and informal immunization data. The formal data may be helpful in identifying program activities associated with changes in vaccination coverage. The informal data may be useful in assessing what information is being shared during an outbreak or other emergencies. The lexicon databases will need regular updating to continue to be relevant.