Presenter/Contact: William G. Adams, MD, badams@bu.edu
Co-authors: Kressin, NR, Thomas, S, Thomas, SM, Shanahan, CW, Paasche-Orlow, MK

CTSA: Boston University with the BUMC/BMC Health Disparities Research Program

Title: Extending i2b2 to Support Health Outcome and Disparity Measurement

Availability of high-quality clinical data and i2b2 offers tremendous opportunities for clinical and translational researchers seeking to use health data for intra- and inter-institutional hypothesis generation, cohort identification, and health services research. i2b2 currently excels in the area of cohort identification but lacks functionality related to health care process and outcome assessment. In this presentation, we describe our efforts over the past 12 months to: 1) integrate clinical and administrative data from an academic medical center (BMC), 5 affiliated community health centers, and the largest Medicaid insurance plan in MA; 2) create new i2b2 hierarchies for insurance, service area, and vital signs; and 3) develop a new i2b2 Cell, the "Health Outcome Monitoring and Evaluation (HOME)" Cell. The first two areas will be discussed during the presentation. This report focuses on Home Cell functionality and architecture.

We developed the HOME Cell to serve as a shared resource within the i2b2 community that will be improved over time. A functional prototype is available and will be demonstrated. The Cell functions within i2b2 without any changes to the core i2b2 schema or software. The HOME Cell reuses queries developed within the i2b2 Query Cell and extends functionality through user-specified "constraints" within the HOME Cell. Three core constraints have been developed and tested to date. The Occurrence Constraint models temporal relationships between facts. Users can specify that only persons with a fact that occurred in a certain temporal relationship with another fact should be included (i.e., to assess lipid screening rates in people with diabetes, only include people with a primary care visit that follow a diagnosis of diabetes). The Value Constraint extends the Occurrence Constraint further to specify a numeric range (less than, greater than, or between) on values of a clinical observation and specifies which value(s) to use -- min, max, average, earliest, most recent or all -- (i.e., to assess blood pressure control in people with hypertension, all systolic and diastolic blood pressure measurements after the diagnosis of hypertension should be less than 130 and 90 respectively). The Age Constraint models the relationship between patients' age and facts (i.e., age at visit < 65). The constraints can be grouped as logical conjunctions (AND) of disjunctions (OR), just like in the i2b2 Query Cell. The HOME Cell specification of a clinical process or outcome starts with a base i2b2 query. A group of constraints are applied to the base query to define a denominator patient set and a numerator patient set. Existing i2b2 queries are used to specify various strata for analysis (i.e., African American, uninsured diabetics etc.). A reference interval can be applied to any subset of constraints to report the data within a specified date range by year, month or the entire range.

The architecture of the HOME Cell uses the existing i2b2 execution framework. User-specifications are translated into XML and parsed by the i2b2 server into a set of sequential queries which are executed in the order submitted. The query execution begins with an initial patient list generated as a result of the base i2b2 query. The generated SQL queries representing the specified HOME Cell constraints are applied sequentially to this list using joins. A counter is maintained for every patient in the list to track the number of constraints the patient satisfies. This counter is used to identify subgroups of individuals satisfying all constraints within a numerator and denominator population. Results can be reported as counts, averages, or proportions. The final denominator and numerator populations can then be stratified by any existing i2b2 query to assess exposure-outcome relationships for subgroups of individuals. The SQL 'Group By' function is used during execution to allow results to be reported by year, month, or the entire date range.

Our experience has thus far demonstrated that a vast and varied number of health services queries (disparity outcomes, comparative effectiveness, and quality reporting) can be supported using this model. Within the rapidly expanding i2b2 collaborative network the HOME Cell has the potential to be a powerful tool for translational research.