R3M: Readmission Risk Monitor

Robert P. Yerex, Ph.D

Updated: July, 2015

UVA Medical Center Accountable Care Organization

**Objectives:** Develop a series of predictive models derived using machine/statistical learning algorithms that target patient readmission probability and timing. Final objective is the deployment of an application that

# Model Development

With the ultimate goal of developing a predictive algorithm targeting 30 day all cause readmission with the ACO patient population, there are three phases planned.

## Deliverables

The deliverable from each phase will be a documented model including:

* R code for training, testing, and scoring
* Description of the data set used
* Discussion of each step taken in refining the model including algorithm selection feature creation/selection
* Diagnostic metrics measuring model accuracy such as; confusion matrices, specificity and sensitivity measures, cost and risk profiles, etc.
* Discussion of potential deployment issues (Phase 3 only)

Each phase builds on the practical knowledge gained from the previous phases in terms of algorithm selection, data transformation, metric development, training and testing.

## Phase 1: Stroke Survival

This phase has been completed and a detailed description can be found in . In summary, using data from the Copenhagen Stroke Study, this model targeted patient reoccurrence of stroke and/or mortality. Given the small size of the well curated data set (< 1,000 records), it was possible to carry out the development of this model on a desktop PC running R. All data transformation and manipulation was done within R, after extraction from a standard csv file.

## Phase 2: Diabetes Patient Readmission

In the current phase of the project, I am building a model to predict which patients, having had an initial inpatient diabetes related encounter will be; readmitted within 30 days, readmitted after 30 days, not readmitted within the time scope of the data set.

The data set consists of some 70,000 patient records from the *Health Facts* [[1]](#footnote-1) database, a national data warehouse of clinical records gathered from US hospitals that use the Cerner Electronic Health Record System. This data set has been used in studies investigating the relationship between clinical indicators and diabetes outcomes/readmissions[1].

Development of this model builds on what I learned from phase 1 while incorporating more complex metrics such as multiple diagnoses. Given the much larger size and complexity of this data set, successful development of an accurate predictive model will require thoughtful development of compound metrics (i.e. for multiple diagnoses tied to multiple procedures) that enable dimensionality reduction while minimizing information loss[2].

## Phase 3: Readmission in ACO Population using (only) Administrative Claims Data

Administrative claims data is intended to be used for operational and financial planning and not for clinical interpretation or patient care optimization. These records lack the granularity or patient diagnostic details found in operational EHR systems such as Epic[[2]](#footnote-2).

In connection with administration of the Affordable Care Act, the Centers for Medicare and Medicaid Services [3] provides claims related data on patients associated with specific Accountable Care Organizations such as *Well Virginia* [4]. The goal of this third and final phase of this project is to use this data to predict the probability of individual patient readmission, and the timing of that readmission. This information can then be used to rank recently discharged ACO patients in order to optimize follow-up care, thus directly reducing costs as well as insuring that Well Virginia meet or exceed specific ACO quality of care guidelines such *ACO #8 – Risk Standardized All Condition Readmission*. A summary description for this particular measure presented in Appendix B.

There is some precedent for using this type of data to study readmission. Vernig, et. al. [5], successfully used survival analysis techniques to analyze Medicare data. This study was focused on examining specific relationships found in the data and not on predicting outcomes.

Other studies that have focused on readmission prediction have used data beyond what is available in the CMS ACO Claims database [6], [7]. In a meta-analysis conducted by the VA on readmission prediction, a number of studies were examined [8]. The nine large population-based or multicenter US studies that relied upon retrospective claims data generally had poor discriminative ability (c-statistics 0.55 – 0.65). Addition of clinical information increased the discriminative ability considerably (c-statistics 0.72 – 0.75). One study further increased discriminative ability through the additional of functional status data (available at discharge (c-statistic 0.82).

Given that my model will only incorporate retrospective claims data, the goal would be to produce a predictive algorithm of sufficient accuracy that the c-statistic surpasses 0.65.

1. Cerner Corporation, Kansas City, MO [↑](#footnote-ref-1)
2. Epic Systems Corporation, Verona, WI [↑](#footnote-ref-2)