Exploring Predictive Models of ACS Outcomes with a Longitudinal, Multi-institutional Patient Database

Wynona Black[1], Eric Horvitz[1,2], Meliha Yetisgen-Yildiz[1] & John H. Gennari[1]
[1]Biomedical & Health Informatics, University of Washington, Seattle, WA;
[2]Microsoft Research, Redmond, WA

Abstract

Acute Coronary Syndrome (ACS) affects millions of patients. The increased use of EMRs data and machine learning could yield better predictions of patient outcomes and aid in decision support in the treatment of ACS. In our study, we aim to develop predictive models of ACS outcomes and explore challenges to these models such as inter-institutional variability and changes over time.

Introduction

Acute Coronary Syndrome (ACS) affects millions of patients per year in the US, with significant morbidity and mortality. Its treatment is complex, and the AHA has helped write numerous healthcare guidelines to help decision-making and to reduce variability in practice. However, because decisions must be made quickly and with incomplete information variability persists in both practice and health outcomes.

As EMR use increases in the US (especially with the passage of the HITECH Act), data warehouses of de-identified patient data are increasingly available for research. There is an opportunity to analyze this EMR patient data to better understand the correlations between practice choices and outcomes. Such analyses could help with improved clinical guideline and decision-support development efforts and could uncover practices that lead to better health care outcomes.

Methods

We will use a 9-year longitudinal EMR-derived database provided by an EMR vender to investigate the relationships among clinical practice, patient demographics, institutional characteristics and the outcomes of patients with ACS. The Acute Coronary Syndrome Patient Database includes 23 million encounters for 6 million patients in 128 healthcare institutions across the US. There are 88 million medication orders, 648 million laboratory test results, 35 million diagnoses, and almost 4 million procedures. Of the 6 million patients, 58% are female (n=3,694,642) and 73% are Caucasian (n=4,664,543). The patients averaged 4.1 encounters (range: 1-564, standard deviation: 7.7) over the nine year time period. Forty–five percent (n=58) of the health care institutions have at least 200 beds, are mostly in urban locations (86%, n=110) and are split between non-teaching (56%, n=72) and teaching (35.1%, n=45; unknown 8.5%, n=11).

We plan to use machine learning algorithms and features derived from clinical decisions and healthcare events to predict outcomes of interest such as mortality and readmission. We have the unique opportunity to build general and site-specific models and investigate the similarities and differences between those models. For example, we will use all of the data to build a general model of mortality then we can apply our general model to a specific site to measure its predictive performance for that site. When building models to predict an outcome, there is a balance between over-fitting and the generalizability of the model to other data sources or populations. The model created at one institution, and by extension, for a specific population, may have diminished predictive performance in data derived from another institution. We will investigate institutional characteristics, such as size, location, teaching or non-teaching, and their effect on model performance. Furthermore, as our database spans 9 years, we will investigate changes in features over time.

Conclusion

Statistical analyses of large databases of EMR data promise to provide guidance that can lead to enhanced outcomes prediction, improved guidelines, and corresponding improved decision support tools for the clinician. In our study, we aim to develop predictive models for key observational variables and outcomes of interest, focusing on such questions as sensitivity of outcomes to the specific hospital providing care. Ultimately, we hope our analyses can contribute to guideline authoring, improved decision making, and patient outcomes for acute coronary syndrome.