
Uri Kartoun, Ph.D.1,2, Vishesh Kumar, MD1,2, Ari Brettman, MD1,2, Sheng Yu, Ph.D.3, Katherine Liao, MD, MPH2,4, Elizabeth Karlson, MD2,4, Ashwin Ananthakrishnan, MBBS, MPH2,5, Zongqi Xia, MD, Ph.D.1,2, Vivian Gainer, M.S.6, Andrew Cagan, B.Sc.6, Shawn Murphy, MD, Ph.D.6, Susanne Churchill, Ph.D.1, Isaac Kohane, MD, Ph.D.2,5, Peter Szolovits, Ph.D.1, Tianxi Cai, Sc.D.3, Stanley Y. Shaw, MD, Ph.D.1,2

1. Center for Systems Biology, Massachusetts General Hospital (MGH), Boston, MA; 2. Harvard Medical School, Boston, MA; 3. Department of Biostatistics, Harvard School of Public Health, Boston, MA; 4. Division of Rheumatology, Brigham and Women’s Hospital (BWH), Boston, MA.; 5. i2b2 National Center for Biomedical Computing, BWH, Boston, MA; 6. Research Computing, MGH, Boston, MA; 7. Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Summary: We develop a risk score for re-admission following an index congestive heart failure (CHF) admission, using codified and narrative variables from 17 years of electronic medical records (EMR). We show a strong correlation between calculated and observed re-admission frequencies throughout the range of risk, in both diabetic and non-diabetic populations.

Introduction/Background: Identifying patients at high risk of re-admission after an index admission for CHF is of major academic, operational and financial interest. We hypothesized that an unbiased method of risk discovery drawn from EMRs may identify patients at high risk for re-admission and provide opportunities for intervention.

Methods: We predicted the likelihood of an index CHF admission being followed by a subsequent admission for any cause within 30 days of discharge, using data available at two time points within the index admission: 1) the first 24 hours ("early"), and 2) at the time of discharge ("discharge"). Our study data included 17 years of inpatient CHF admissions at two urban tertiary care hospitals between 1993 - 2010. We focus on two cohorts: 1) 65,099 type-2 diabetes (T2D) patients, with 5,825 index CHF admissions (with 23.4% 30-day re-admission rate), and 2) a Non-Diabetic Cohort of 43,220 patients (2,203 index CHF admissions and 22.4% 30-day re-admission rate). We extracted 293 EMR variables including demographics, laboratory values and slopes, billing codes, cardiac parameters extracted from narrative electrocardiogram and echocardiographic reports, and medical concepts extracted from physician narrative notes using natural language processing.

Results: Using logistic regression with the adaptive LASSO, we found a strong correlation between predicted and observed risk of re-admission throughout the range of calculated risk for the Diabetic Cohort ($r \geq 0.99$ for both the "early" and "discharge" models). Patients who had a re-admission within 30 days had a significantly predicted risk score vs. patients who were not re-admitted ("early": 28.6% vs. 21.8%; $p = 3.7 \times 10^{-66}$, "discharge": 29.4% vs. 21.5%; $p = 2.7 \times 10^{-77}$). Using a four-fold cross validation scheme yielded C-statistics of 0.65 and 0.67 for the "early" and "discharge" models, respectively. The "early" and "discharge" models had comparable accuracy in assigning patients to the highest and lowest deciles of re-admission risk. Significantly, the "discharge" model successfully re-classified a subset of patients of intermediate risk in the "early" model: calculated and observed re-admission rates for patients re-classified into the highest-risk decile in the "discharge" model were 45.0% and 43.6%, respectively. Applying an analogous approach to the Non-Diabetic Cohort yielded similar results.

Discussion: A generalizable method using unbiased variable selection and model building from EMR data can successfully identify patients at high or low risk of re-admission. "Early" data can identify high and low-risk groups; additional data generated during the admission and available at time of discharge can further re-classify additional individuals into high or low risk groups. This two-phase approach to risk estimation may facilitate intervention for high-risk patients earlier in the index hospital admission.