Rationale and Design for the Duke Connected Care Predictive Modeling Pilot with a Medicare Shared Savings Program Population

Shelley A. Rusincovitch[1], Ricardo Henao, PhD[2], Michael Gao[3], Lawrence Carin, PhD[2], Ursula Rogers[1,4], N. Benjamin Neely, MS[5], Mary Schilder, MA[4], Daniel Costello, MPA[6], Eugenie Komives, MD[6], Erich S. Huang, MD, PhD[5,6]

[1]Duke Clinical and Translational Science Institute, Durham, NC; [2]Duke University, Durham, NC; [3]Duke Institute for Health Innovation, Durham, NC; [4]Duke Health Technology Solutions, Durham, NC; [5]Duke Office of Research Informatics, Durham, NC; [6]Duke Connected Care, Durham, NC; [7]Duke Department of Biostatistics and Bioinformatics, Durham, NC

Introduction

Predictive modeling offers the opportunity to assess, manage, and intervene on risk in Accountable Care Organizations (ACOs)[1, 2]. Although the deployment of machine learning and other predictive modeling tools has become well-established in many industries, these methods have only recently become more common in healthcare settings.

In this abstract, we discuss the rationale and design for a machine learning pilot undertaken by Duke Connected Care, an ACO participating in the Medicare Shared Savings Program (MSSP)[3], in conjunction with a data science team of informaticists and machine learning experts. The result of the pilot is expected to be twofold: (1) development of a clinically-informed and actionable predictive model that can be assessed for potential implementation within Duke Health, (2) synthesis of experiences arising from an ACO's engagement and collaboration with machine learning expertise.

Methods

The pilot targets inpatient admissions, complimenting earlier work by Duke Connected Care (DCC) in predicting the risk of readmissions. Predicting this initial hospitalization is broadly applicable to the MSSP population, and yields opportunities for intervention and prevention with the goal of achieving higher quality of care with lower cost.

This operational project is undertaken for patient care and its operational scope is carefully framed within appropriate data use parameters. The computational environment that we developed for the pilot is governed by mechanisms that include an honest broker function to manage secure transport of datasets, individual user agreements, and processes for access control and permissions. Links to external knowledge bases include application program interface (API) calls to the National Library of Medicine RxNorm database[4]. API calls never involve protected health information nor other patient attributes.

One of the key criteria used to assess the output of the machine learning model is actionability. Both the quantitative team and the larger cross-disciplinary stakeholder group will evaluate the ability of machine learning to identify predictive factors that contribute to higher risk of admissions, particularly those that care coordinators and providers can act upon to mitigate risk.

Results

Our machine learning methodology has foundations in earlier work with Electronic Health Record (EHR) data and is an extension of Deep Poisson Factor Analysis[5]. We have applied these methods to a new data source: claims data for the MSSP population served by Duke Connected Care. These data are provided by the Centers for Medicare & Medicare Services (CMS) in the ACO Operational System (ACO-OS) Claim and Claim Line Feed (CCLF) file format. These data are transactional and differ notably from the typical structure of research data files available through the CMS Research Data Assistance Center (ResDAC). The ACO-OS CCLF files are received monthly, and the population included in these files varies over time based on complex inclusion criteria and attribution.

A key component for the design of the pilot has been domain modeling to describe the source data and generating high-level concepts applicable across both claims and EHR data sources.

Table 1. A partial list of data domains modeled for feature extraction and consumed by the machine learning methods.

Domain	*Domain Definition*	*ACO-OS CCLF Claims Source*
Admitting diagnosis	Assigned with the information known when the patient is **first admitted** (**prior** to diagnostic testing and evaluation). These data may be symptom-oriented; for example, shortness of breath at admission may be later diagnosed as congestive heart failure through diagnostic evaluation. Aggregated into CCS diagnosis category[1]	Part A Claims Header File (CCLF1)
Discharge diagnoses	Assigned by medical coders after the **conclusion** of the encounter, and incorporating the results of diagnostic testing and provider evaluation. The primary diagnosis assignment is generally weighted by medical coding practices. Aggregated into CCS diagnosis category*.	Part A Claims Header File (CCLF1) and Part A Diagnosis Code File (CCLF5)
Procedures	The discreet **medical interventions** (such as surgical procedures) and execution of **diagnostic testing** (such as laboratory orders) delivered within a healthcare context. Assigned by medical coders based on facility and provider documentation of services rendered. Aggregated into CCS procedure category*.	Part A Procedure Code File (CCLF4)
Dispensed medications	Medication dispensed **directly to a patient** by a pharmacy; this is different than medications prescribed or administered within a healthcare facility. Codified in the highly granular National Drug Code (NDC) terminology, which specifies packaging and other attributes of the drug product. Derived into medication active ingredient and class from RxNorm**.	Part D File (CCLF7)

*Aggregated using the HCUP Clinical Classifications Software, a common-used schema that is used here to aggregate the individual ICD-9 and ICD-10 encounter-based codes into more meaningful, higher-level categories.
**The National Library of Medicine (NLM) RxNorm medication normalization naming platform is part of the open-source, publicly available Unified Medical Language System® (UMLS®). Here, the original NDC codes for each ingredient are mapped against RxNorm to retrieve the medication's active ingredient and class listing from RxNorm.

Discussion

We anticipate that the organizing principles of performance and pragmatism will strongly influence our models' eventual adoption and operationalization. One design consideration is evaluating the tradeoffs in timeliness and accuracy between modeling based on ACO claims versus EHR data, versus the combination of the two. Our current work has been based on Medicare claims data; as a next phase, we plan to combine these claims data with EHR data for the same patient population, and leverage supplemental data sources including geospatial data associated with patient addresses and other auxiliary sources.

Challenges for the pilot have included acquiring and deploying computational resources sufficient to support the high-performance data science within a typical enterprise IT environment, and the effort to assess the claims data structuring and nuances, especially attributes and concepts specific to the MSSP.

ACOs can help health systems achieve better outcomes by efficiently identifying and reducing health risks across populations, and we expect that the pilot model will more quickly and accurately identify high-risk individuals. This improved identification will create efficiency by shifting effort from risk assessment to actual care intervention and patient engagement.

References

1. Berwick DM. Making good on ACOs' promise--the final rule for the Medicare shared savings program. The New England journal of medicine. 2011;365(19):1753-6.
2. McWilliams JM, Hatfield LA, Chernew ME, Landon BE, Schwartz AL. Early Performance of Accountable Care Organizations in Medicare. The New England journal of medicine. 2016;374(24):2357-66.
3. Centers for Medicare & Medicaid Services, Shared Savings Program [cited March 7, 2017]. Available from: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/
4. RxNav APIs [cited March 7, 2017]. Available from: https://rxnav.nlm.nih.gov/APIsOverview.html
5. Henao R, Lu JT, Lucas J, Ferranti J, Carin L. Electronic Health Record Analysis via Deep Poisson Factor Models. JMLR. 2016;17(186):1-32.