White paper

Predictive Analytics in Value-Based Healthcare: Forecasting Risk, Utilization, and Outcomes

Watson Health™

IBM

Delivering effective value-based healthcare requires identifying and mitigating risk by anticipating and preventing adverse events and outcomes. Predictive models have been a part of healthcare practice for decades. However, more advanced analytics have started to take shape to provide better visibility into characterizing a patient's current state and future risk. With the use of big data, it is possible to build models around predicting future events and outcomes, utilization, and overall risk. These predictive models can be:

- incorporated into a clinical workflow to facilitate care management and identify individuals at risk
- used to perform risk adjustment on quality measures to account for patient severity
- employed to understand the treatment pathway with the greatest chance of success

Predictive models have many potential uses, but no single risk score serves as a magic bullet to answer all questions regarding a patient or a population. It is important to understand what question a risk score is addressing to determine the circumstances in which a predictive model should be used.

Predictive model overview

A traditional predictive model is a mathematical formula that uses historical information to predict the chance of an event happening in the future. Most predictive models are built to address a population. The models are built based on a training dataset (a set of patients where the input factors are defined and the resulting output is known). The model is then evaluated to determine its performance using a test dataset (a separate sample of inputs where the result is blinded from the system and only used at the end of the process to determine the prediction's accuracy).

Consider a model predicting the chance of a hospital readmission. The input training dataset might contain thousands of hospitalization events. Each hospital admission would include the information on a patient's medical history, like chronic conditions and prior utilization. The dataset would also contain the "answer" as to whether that particular hospitalization resulted in a readmission within 30 days. Data scientists typically use specialized software (or write their own) to build models that "learn" from the results of a readmission to improve predictive accuracy. The model is then evaluated using the testing dataset, where a similar set of data is provided without the "answer" and the model's prediction is compared to the correct result (whether the patient was truly readmitted or not). The accuracy of a model on a testing dataset is typically what is presented as the true performance of the model.

To indicate the accuracy of a given method, model performance can be reported in a variety of ways, including:

- a classification rate (the number of true predictions over all predictions)
- sensitivity/specificity (rates of correct true positive/negative predictions)
- R2 (the amount of output variability accounted for by the model)
- Mean Absolute Percent Error (MAPE) (the overall percent difference of the predicted value from the true value)
- Area Under the Curve (AUC) (a measurement of performance accuracy allowing the user to see the tradeoffs between true and false positive rates under different conditions)

Predictive models can be designed to apply to individual patients or to a population as a whole. However, the uncertainty of predicting events for individuals or small populations has a higher level of uncertainty or variability associated with it than when predictions are averaged across a larger population. The resulting accuracy of individual models is typically lower when compared to the prediction rate across a larger population.

Predictive models in healthcare practice

Predictive models have been used in healthcare decision-making for decades. Perhaps the most well-known predictive model in healthcare comes from the Framingham Heart Study[1]. The study began in 1958 to track the progression of cardiovascular disease in more than 5,000 subjects. The study is still ongoing and resulted in several predictive models across several diseases. The most commonly used model takes six data points (age, diabetes, smoking status, systolic blood pressure, total cholesterol, and HDL cholesterol) and predicts the probability that someone will develop cardiovascular disease in the next 10 years. The Framingham Study is now in its third generation of participants and has led to a creation of family of risk models, including atrial fibrillation, congestive heart failure, coronary heart disease, diabetes, hypertension, and stroke.

Not all predictive models require a prospective study. The modern age of big data analytics provides access to a wealth of deep clinical and administrative data to study large populations of individuals with significant longitudinal history. Paired with advanced data analysis methods, such as machine learning and cognitive analytics, it is possible to develop new predictive models with a statistical power and scale that is unprecedented. Modern technologies allow for the initial development of such models within hours. The ubiquitous use of electronic medical records and care management platforms allow for the deployment of models, once validated, to be used in tracking and managing patient care.

Event prediction

Several areas of predictive models are in use today. The first major use case is in predicting events and outcomes. In healthcare, that typically relates to identifying the chance of unwanted events occurring such as avoidable healthcare utilization, specific clinical conditions and their progression, and outcomes such as mortality. Examples of avoidable utilizations include a hospital admission or emergency department (ED) visit within the next 30 days or hospital readmissions to the hospital within 30 days of discharge. Disease prediction can identify individuals with a higher than average chance of developing a new chronic condition (e.g., Framingham Risk Model to predict the onset of diabetes in the next 8 years). Disease progression is used to determine how a condition might advance over time (e.g., the chance of that diabetic progressing from chronic kidney disease to end stage renal disease requiring dialysis). Outcomes prediction can assess the chance of any number of future events, from the chance of complication after surgery to mortality within the next year.

The goal of event prediction is to manage a population such that as many mitigatable, undesirable outcomes are avoided as possible. For example, a care team might evaluate the readmission risk for any patient currently admitted. That score could help inform the team to decide whether a patient is kept in the hospital longer, allowed to go home with appropriate follow-up care, or safe to be discharged without any extra assistance. A risk score alone cannot determine the best course of action or treatment. However, a risk score can be a valuable component of the picture of a patient's overall health to assist providers in making care decisions. Predicting these events and then avoiding them can have a profound impact on the overall value of healthcare.

Utilization and Risk Prediction

One of the most common forms of risk prediction involves identifying the relative financial risk of an individual within a population of patients in an actuarial, meaningful way. Typically, the healthcare spend of an individual is directly tied to his or her condition, utilization, and overall risk. Utilization models are split into concurrent and prospective prediction. Concurrent models evaluate an individual's risk and utilization for the current year. Prospective models evaluate an individual's risk and utilization for the upcoming year. Naturally, the predictive power lessens the farther you attempt to predict in the future. Still, prospective models offer insight into how an organization should manage a population and where to place resources.

Several cost and utilization models exist in both the public domain and commercial vendor spaces. For example, the Centers for Medicare & Medicaid Services (CMS) utilizes a model called the Hierarchical Condition Category (HCC) to provide prospective utilization calculations for individuals enrolled in a Medicare Advantage plan that are used in determining reimbursement[2]. The HCC score utilizes the patient's demographics and diagnoses for the current year to predict the expected utilization for the following year[3]. HCC scores are normalized such that a score of 1.0 indicates that an individual is expected to cost the average amount for that population. A score of 1.5 indicates that an individual will cost 50% more than average and a score of 3 indicates that an individual will cost three times the average. HCC is a good example of why appropriate documentation and subsequent coding of conditions are important for both risk management and reimbursement. Failure to document a condition on a bill or claim could lead to underestimating the risk of an individual and result in lower reimbursement.

An example of a more complex utilization prediction algorithm is the IBM Explorys Risk Model[4]. Instead of relying on only demographics and diagnoses, the IBM Explorys Risk Model also considers procedures, medications, and prior healthcare costs to predict prospective utilization. The additional input factors increase the overall performance compared to models using diagnoses and demographics alone. The IBM Explorys Risk Model is a good example of how organizations can increase predictive performance of a model by adding additional input features.

Risk Adjustment

One of the largest challenges in accurately reporting quality metrics is taking into account the severity of the population being measured. Risk adjustment in healthcare is a means of using an individual's severity, risk, or burden to normalize a reported outcome or quality metric. A simple example is a provider's patient panel. The number of individuals being cared for indicates the general effort being spent by a provider on delivering care. However, a straight panel count comparison is not effective because not all patients require the same level of care. Instead, a utilization prediction score could be used to adjust the population count. Consider two providers that each care for 1,000 patients. Provider A's normalized population average utilization risk score is 1 while Provider B's is 2. Provider B is expected to spend twice the resources of Provider A and is effectively treating 2,000 patients.

Risk adjustment is most commonly used to adjust utilization rates, including hospital admissions, ED visits, and readmissions. Naturally, the chance of a hospital readmission increases greatly when an individual is sicker. Therefore, adjusted counts provide a more apples-to-apples comparison when looking across different populations, providers, and healthcare systems. For example, healthcare system A might have a heart failure readmission rate of 30% while healthcare system B has a rate of 15%. It might appear as if healthcare system B is doing twice as well as system A. However, if healthcare system A is a tertiary care center, it may be dealing with individuals who are more likely to experience complications. Healthcare system B might be dealing with more traditional treatment cases. After risk adjusting for patient severity, the risk adjusted readmission rate of both healthcare systems might be closer to 20% and their performance is effectively equivalent.

Risk adjustment makes less sense when evaluating process metrics, such as weight screening and smoking cessation counseling. The severity of a patient's condition should not have a direct effect on standard care processes. Additional factors can affect patient compliance, but they should be accounted for in the measure definition, including patients at the end of life, in hospice, etc. More advanced risk adjustment and stratification can take into account patient features such as socioeconomic status, disability status, accessibility of public transportation, and adequate insurance coverage. However, most commonly used risk adjustment methodologies do not currently take these factors into consideration.

Next Generation Analytics: Cognitive and Deep Learning

Predictive model development has seen a shift from relying on traditional statistical techniques to utilizing new techniques and technologies in machine learning, deep learning, and cognitive computing. These newer technologies provide improved predictive power by using significantly more data within a flexible modeling environment that can learn and adapt. Where traditional regression models typically use 10-20 input variables across a fixed regression equation, machine learning models can be built using hundreds or thousands of inputs in a dynamically generated model. Cognitive and other learning technologies expand the types of data available for modeling, including text from multiple sources, wearable devices, and genetic data. Cognitive learning technologies extend predictive power by introducing information context and reasoning to the information being assessed, to link knowledge and insights in the process of model development. Cognitive models build and improve upon a base set of knowledge by continually generating new insights and learnings in an ever-evolving system.

A downside of advanced predictive models is that their complexity can present interpretation challenges to the end user. The advanced models can sometimes be viewed as a black box and can lead to challenges in model adoption due to a lack of understanding. One resolution is to simplify the model to the point that it can be clearly understood (often at the sacrifice of model performance). Alternatively, well-documented development and testing procedures can assist others in understanding how a model was built and how it performs in real-world scenarios.

There are several major players in the next generation of predictive analytics. Many of these new solutions go beyond traditional predictions and provide solutions to solve real-world problems. Two of the most prominent contributors are Google and IBM.

Google developed its deep learning technology, called DeepMind, which is a combination of machine learning and artificial intelligence[5]. Google successfully trained its AlphaGo system to play the game Go and beat the European champion Fan Hui in 2015[6]. Google is expanding into the healthcare space with DeepMind Health. Initial research projects involve examining retinal scans to identify macular degeneration and other eye-related conditions.

IBM has been involved in advanced analytics and artificial intelligence for decades. One of the first prominent contributions was IBM's Deep Blue system for playing chess, which defeated world champion Garry Kasparov in 1997[7]. In 2011, IBM unveiled its Watson system of cognitive computing on the TV game show Jeopardy![8]. IBM's cognitive computing is also an extension of machine learning and artificial intelligence where the system is able to comprehend, reason, and learn. IBM presents its form of AI as "augmented intelligence," as it is not expected to replace an individual's decisions, only to inform them. Cognitive technologies are at the center of IBM Watson Health to help inform decisions around multiple clinical areas, including oncology treatment, clinical trial matching, and drug discovery[9]. For example, Watson for Oncology examines an individual's DNA sequence from a healthy cell and a tumor cell to identify mutations. The cognitive system then examines the body of scientific literature to identify potential treatment pathways and determine their likelihood of success. The results often include treatments normally administered for a different form of cancer that may have not been traditionally considered[10]. For example, a patient with lung cancer might be presented with a treatment option for stomach cancer because the individual happens to have a mutation found in several forms of stomach cancer. As the body of medical and scientific literature grows, so does the learning and reasoning power of Watson.

Implementing Predictive Models in the Clinical Workflow

Predictive models can be used at different points in the care management process to support benchmarking, drive quality initiatives, and understand individual patient risk. Population-based predictive models in benchmarking allow for a closer, more accurate comparison between populations within and across healthcare systems. For example, understanding the rate of readmissions considering patient risk and severity. Similarly, population-based predictive models in quality management programs allow an organization to assess a population as a whole to determine the overall risk and expected utilization. Organizations can leverage individual-level predictive models to understand a single patient's risk profile, including expected utilization and the probability of adverse events occurring in the near future (e.g., readmission, hospitalization, complication).

The greatest predictive models have value only if the information is reliable and can be acted upon. Actionable insights typically require timely data, analytics, and results that are presented in a form that can be readily used. For example data or analyses that are a week or a month old often provide little value in establishing the risk of an individual currently admitted in the hospital. Similarly, a timely risk score on its own must have the proper context and workflow options to be meaningful. The analytics need to be embedded in the care delivery and management workflow in order to be acted upon. Changes in provider behavior and patient outcomes must be clearly measureable to track performance improvement and the effectiveness of any analytics program.

Conclusions

Healthcare is only beginning to see the potential of predictive analytics. The power of predictive algorithms will increase as more data and technology are available. Modern clinical datasets and analytic technologies present an enormous opportunity for new model development. The predictive power of a model developed on tens of millions of individuals across hundreds of input factors far exceeds traditional methods of training a regression model of ten inputs across a sample of a few thousand individuals. Big data technologies further empower data scientists to quickly build and test models within a few hours instead of days or weeks with traditional techniques.

Predictive algorithms are increasingly becoming a part of clinical healthcare practice. Clinical adoption will grow as models gain in maturity, predictive power, and usable applications. The results of predictive algorithms must be presented in a way that can be clearly understood. The information presented must be actionable to drive an improvement in healthcare practice and made available to a practitioner in a timely manner. It will be necessary to provide true clinical value in order to gain differentiation and adoption. New systems will continue to be developed that help inform provider decisions and drive better outcomes. Predictive models, whether traditional or cognitive, will be an expected component of care to provide the fullest, data-driven approach to a patient's value-based health.

Footnotes

1. Framingham Heart Study. https://www.framinghamheartstudy.org/

2. Center for Medicare and Medicaid Services (CMS). Risk Adjustment. https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html

3. HCC University HCC Calculator. http://www.hccuniversity.com/hcc-university/risk-score-calculator/

4. Rong Yi. "Evaluation of Explorys' Risk Adjustment and Predictive Models." Milliman Client Report. 2014.

5. Google DeepMind. https://deepmind.com/

6. Google DeepMind AlphaGo. https://deepmind.com/alpha-go

7. IBM Deep Blue. http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/

8. IBM's Watson computer takes the Jeopardy! challenge. https://www.ibm.com/midmarket/us/en/article_Smartercomm5_1209.html

9. IBM Watson Health. http://www.ibm.com/watson/health/

10. IBM Watson Health Watson For Oncology. http://www.ibm.com/watson/health/oncology/

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States
of America, January, 2017

HLW03032-USEN-00

Watson Health™