Predictive Modeling and Population Health Management
Contents

Page 3
Introduction

Page 4
Background

Page 5
Turning Predictions Into Action
Risk Stratification

Page 6
Clinical Judgment and Culture
Provider Attribution and Risk Adjustment

Page 7
Financial Risk

Page 8
Data Sources
Claims Data

Page 9
Clinical Data
Patient-Reported Data

Page 10
Conclusion
Five Ways to Leverage Predictive Modeling

Page 11
Notes
Introduction

As the healthcare industry continues its transition to new care delivery and payment models, an increasing number of healthcare organizations are embracing population health management. By helping people manage their own health so that they need less health care, and by proactively managing the care of chronically ill patients, these organizations seek to achieve the Triple Aim of improving the quality of care, reducing health costs, and enhancing the patient experience.

To manage population health, healthcare systems and group practices must build the requisite infrastructure, including software tools designed for data analysis and workflow automation. The front end of this IT infrastructure is a type of analytic solution known variously as predictive analytics, predictive modeling, or health forecasting. In a population health management context, these algorithmic tools predict which people are likely to get sick or sicker in the near term.

This is crucially important information to provider organizations and health plans that take financial responsibility for care. Ten percent of patients generate roughly 70% of health costs; five percent account for half of health outlays. By identifying which people are high risk or likely to become high risk, care teams can intervene with them to improve their outcomes and lower health costs. Most health plans offer case management, disease management, and health coaching programs to these members. Some healthcare organizations seek to ensure that high-risk patients receive necessary services and day-to-day support from care managers. To improve outcomes and lower costs, these organizations must connect predictive analytics with workflow automation tools that enable care teams to intervene with the right patients at the right time in the right way.

In this paper, we explain what predictive modeling is and what it can and can’t do. We show how healthcare organizations can make the insights of predictive modeling actionable for financial managers and clinical teams. We provide examples of how predictive analytics, population risk stratification, and risk adjustment are applied in practice. We also review the kinds of data required to do predictive modeling and compare the value of each data source for taking action.

To improve outcomes and lower costs, organizations must connect predictive analytics with workflow automation tools that enable care teams to intervene with the right patients at the right time in the right way.
Background

Predictive modeling is a branch of clinical and business intelligence (C&BI) that is used to forecast the future health status of individuals and to classify patients by their current health risk (risk stratification). It can also be used in risk adjustment to compare the aggregate health risks of one physician's or one organization's patients to those of another doctor or healthcare entity. Most important from the viewpoint of healthcare organizations that assume financial risk for care, predictive analytics can be employed to predict health costs for individuals and populations.

Predictive analytics depend on computer algorithms that can recognize patterns in data. The applications draw inferences from the data about the likelihood of patients developing certain conditions or exacerbations of their existing conditions. In some cases, the developers of predictive analytics use large public databases as the basis of their models. Other models are built with data about specific patient populations.

To create a predictive algorithm, developers define a problem, then select and evaluate models to solve it. After selecting the best model and validating it, they test it by applying it to a real-world database. They may also improve the accuracy of the predictive tool by using known outcomes to "train" the algorithm.

Health plans have been doing predictive modeling for years, using paid claims data. Over the past decade, health plans have also used their information on patients' health risks to identify those who might benefit from disease and case management programs. Because of the rapid turnover in their membership, most plans have limited the focus of these programs to the sickest people to ensure a return on investment.

Few healthcare organizations did any health forecasting until the emergence of accountable care organizations and new payment models that put them at financial risk. The majority of healthcare systems still don’t have the enterprise data warehouses or registries required for this approach. Those that do are more likely to use these databases for retrospective review than for predictive modeling. But that’s expected to change in the next few years as more organizations take on financial risk for care.

Another factor that is currently driving the uptake of predictive analytics is the financial penalties that hospitals incur if they readmit too many Medicare patients. Not surprisingly, a number of vendors offer applications designed to predict which patients are most likely to be readmitted. At the same time, studies of predictive analytics related to readmissions have proliferated.

Some of these readmission tools appear to be moderately accurate. In addition, a predictive modeling application that calculates the odds of a patient developing a serious chronic condition or having a heart attack has been shown to be effective. Nevertheless, a recent paper on health forecasting points out:

“There is little evidence regarding how or whether forecasting improves healthcare value. This is due to both the modest level of research and what is termed the ‘impactibility’ problem. That is, even if prediction algorithms accurately identify at-risk patients, intervening to achieve desired outcomes is often inhibited by limitations of current disease management approaches or the general state of medical science.”

To put it another way, fairly few organizations are using the insights of predictive modeling effectively to improve chronic disease care. But that is bound to change. While financial forecasting and readmission prevention currently drive the use of predictive modeling in healthcare, the most important use of this approach will be in population health management, because chronic diseases account for 75% of health costs. Indeed, a recent study of Medicare data shows that improving care for patients with complex diseases is the most effective way to “bend the cost curve.” Therefore, the “consumers” of predictive analytics tools will expand from the CFOs to the front line physicians and other care team members.

Most important from the viewpoint of healthcare organizations that assume financial risk for care, predictive analytics can be employed to predict health costs for individuals and populations.
Turning Predictions Into Action

Predictive analytics are worthless unless their insights prompt actions. In the area of population health management, these actions include alerts to providers at the point of care and information that enables care managers to prioritize their patient interventions. On the financial side, healthcare organizations can use predictive modeling to forecast the cost of care delivery so they can evaluate risk contracts.

To be valuable in care management, predictive analytics must be timely. Claims data is necessary to predict the annual costs of caring for a patient population, but information that is three months old will not help clinicians intervene with patients to improve their outcomes. For that, organizations need the latest progress notes, lab results, and medications for a patient—in other words, the clinical data in an EHR. When this clinical data is combined with patient-reported data between visits, the information available for analytics is even more up to date.

Some EHR vendors offer predictive analytics that are capable of doing risk stratification. These analytics modules use claims data, rather than EHR data. Claims data can be used for this task because risk stratification at the broadest level does not require near-real-time data. But clinical information is required to predict health status accurately enough to design cost-effective interventions. Moreover, claims data reflects prior care events and patterns but doesn’t capture recent changes in health behavior; for instance, a heart attack survivor may now be exercising, eating right, and no longer smoking.

Risk Stratification

Predictive modeling forms the basis of risk stratification, which is used to identify the patients who will generate the majority of costs in the near future. Populations can be classified into high, medium and low-risk patients with a fair degree of confidence. Depending on which of those categories a patient is slotted into, they might receive intensive care management; online education and support in managing their own care so that chronic conditions don’t worsen; or just education and encouragement in maintaining a healthy lifestyle.

To prevent people from becoming high risk, it is essential to keep track of and support those who are healthy today but could become sick tomorrow. Of the patients who generate the highest costs in a given year, only 30% had high costs a year earlier.

At a population level, organizations can use risk stratification to decide how best to direct their resources. For the large number of patients who are obese and have high blood pressure, for example, organizations might decide to drill down further to identify patients within this group who have other chronic conditions and unhealthy behaviors that would increase the risk of an acute event.

With this kind of refined data set, organizations can tailor care team alerts and patient interventions by risk cohort. Providers can use care alerts to make sure that the most urgent problems of patients are addressed during office visits. Other clinical staffers can use the insights of predictive analytics to reach out to patients who need to be seen. Care managers can be prompted to intervene with certain patients and can also design campaigns to provide assistance to people with less urgent needs. For example, they might decide that group visits with an endocrinologist would be helpful to patients with diabetes who have not been able to lower their HbA1c levels.
At a population-wide level, this kind of work is very time-consuming and labor-intensive. To make effective use of predictive analytics, healthcare organizations should couple these applications with tools that automate the workflow of care teams. For example, predictive analytics can be applied to electronic registries to give care managers the tools they need to intervene with patients based on their health status. High-risk patients and other patients with care gaps who have not seen their providers recently can be contacted via automated messaging.

Automation can help organizations manage the majority of patients, because most people are healthy or have moderate chronic conditions that don’t require intensive care management. But it is crucially important to identify those patients who are not yet very sick but may move into the high-risk category within the next year. By helping to ensure they follow their providers’ care plans, and by engaging them in managing their own health, care teams can help to reduce the number of these people who become seriously ill.

Predictive modeling is also required to identify those who are already high risk and to prioritize those who need help right away. The sickest patients, such as those with metastatic cancer, HIV, and end-stage renal disease, will automatically go into care management programs. Risk stratification can help identify others who could likely benefit from care management, based on rules such as their number of diagnoses, types and numbers of medications, and prior hospital admissions. But in a large population, thousands of patients might fit these definitions—far too many for care managers to handle personally with limited resources. So it may be necessary to use criteria such as prior costs from claims data and clinical risk status to further prioritize which patients need immediate attention.

Another key point is that there is a direct correlation between comorbidities and health risk. In 2010, for example, Medicare beneficiaries with multiple conditions accounted for nearly all readmissions. Average annual spending for Medicare patients with six or more conditions was $32,658, versus $12,174 for people with four or five conditions; $5,698 for those with two or three diseases; and $2,025 for people with one or no conditions. So a good predictive analytics tool must factor in those comorbidities.

Clinical Judgment and Culture

When it comes to individual patients, predictive algorithms cannot predict who will be hospitalized or who will need to visit the ER with a high degree of accuracy. The findings of predictive tools must be combined with clinical judgment to produce the best results in most cases.

For example, predictive modeling might indicate that an elderly patient who leaves the hospital with several conditions and is on multiple medications is a prime candidate for readmission. But one patient who has those risk factors might receive good home health care and be cognitively alert, whereas another with the same risk factors might have no support at home and might have little ability to understand discharge instructions. A physician who knows those two patients will be able to tell which of them is at higher risk of readmission.

To be of any use in improving the quality of care, predictive modeling tools must be accepted by clinicians. That requires some cultural change on the part of physicians who don’t want to take advice from a computer. Here again, the role of clinical judgment is paramount: If doctors believe that their judgment is being overridden by a computer algorithm, they’ll rebel; but if they view predictive analytics as a kind of clinical decision support, they’ll be more likely to use this tool, much as they use drug interaction checks in e-prescribing applications to avoid medication errors.

Provider Attribution and Risk Adjustment

A prerequisite of population health management is the correct matching of patients to their primary providers. Accurate provider attribution is required for both risk stratification and risk adjustment, which is used in comparing the performance of organizations and individual providers. Asaf Bitton, MD, a researcher at Harvard Medical School’s Center for Primary Care, explains that attribution takes effort but can be done properly:

To make effective use of predictive analytics, healthcare organizations should couple these applications with tools that automate the workflow of care teams.
Attribution happens with about 60-90 percent fidelity, so some patients fall through the cracks. It is a key starting point for knowing generally who your clinicians care for, and getting to near-100 percent attribution within your EHR is an important milestone at the outset of your journey toward population management.\textsuperscript{16}

Risk adjustment enables payers and provider organizations to compare the performance of clinicians, practices, or hospitals fairly by differentiating between the characteristics of the patients they serve.

The most common type of risk adjustor is based on the severity of the health conditions in a particular population. The ACG Predictive Model from Johns Hopkins University, for example, is widely used in provider profiling. Based on diagnosis and pharmacy data, it describes the differences between providers’ case mixes.\textsuperscript{15} Verisk Health offers another commercial risk adjustor that’s closely related to the one used by Medicare and the new state health insurance exchanges. Its DxCG risk adjustor uses Diagnostic Cost Groups and RxGroups. Like the ACG predictive model, DxCG depends on claims data.

The difference between risk adjustment and the broader kind of predictive modeling lies in the data inputs. Whereas both approaches use diagnostic codes, for example, risk adjustment excludes prior costs and utilization of services, which might reward inefficient providers. Predictive analytics, on the other hand, embraces prior costs and a wide range of other variables that might play a role in future health outcomes and utilization of resources.

Financial Risk

Beyond measuring the efficiency of individual providers, healthcare organizations that aspire to take financial risk must be able to project the costs that their patient population is likely to generate. Predictive analytics can be a big help to these organizations, but they must also recognize the limitations of these tools.

Take hospital admissions, which account for a large portion of health costs. The positive predictive value of a predictive modeling application might be as high as 80%, but only for high-risk patients. Applied to people with moderate health risks, the same predictor might have a lower positive predictive value. Predictive analytics can forecast which patients will go to the ER with good accuracy in some cases. But because of the unpredictable nature of some ER visits, which may be related to car accidents or various types of trauma, the software doesn’t predict ER visits as well as hospitalizations.\textsuperscript{17}

However, it is possible to gauge the likelihood that a particular patient will generate high costs in the following year. To calculate that probability, an organization must have data on a variety of risk factors, including information on the individual’s prior costs and utilization of services, current health status, diagnoses, lab results, and medications; it would also help to know something about the non-clinical factors that are discussed below. Applying an algorithm to those variables yields a risk score for each patient, based on their individual characteristics, and an average score can be computed from that.\textsuperscript{18}
To do predictive modeling, organizations must have access to multiple sources of data that describe the health status of individuals and populations as completely and as currently as possible.

The organization benchmarks its average risk score against national standards or its historical costs. If its average cost to care for a patient is $1,000 per year, for example, it multiplies that amount times the number of patients and their average risk score to predict what it will spend in the next year. The organization can then decide whether the capitation payment it’s being offered is sufficient to cover its expected costs.

Each risk contract that an organization negotiates covers a separate population, comprised of patients that are covered by a particular health plan. So providers may have to use the predictive modeling approach described above multiple times for each payer from which they are planning to take capitation. (Payment bundling, while it also involves risk, requires different calculations based on episodes of care.)

The health risks of individuals are always changing, of course, and a few “outliers” could have catastrophic costs in the next year. Large organizations have a better ability to withstand the financial consequences of these catastrophes than small ones do. But no provider organization should take on financial risk without stop-loss insurance to cushion it against these unexpected losses.

Predictive modeling can help organizations factor in these outliers and prevent at least some of them from racking up huge expenses. By tracking catastrophic cases over time, an organization can get a sense of which patients are likely to hit the stop-loss limit, which might be $100,000. It can then provide extra resources to ensure those patients receive appropriate care. While it’s impossible to forecast all catastrophic events, focusing intensively on those that are most likely to occur can have strongly positive results for both the patients and the bottom line.

**Data Sources**

To do predictive modeling, organizations must have access to multiple sources of data that describe the health status of individuals and populations as completely and as currently as possible. The information must be very timely to be actionable for care management.

“The more data you have, the better able you are to predict outcomes,” Patrick Gordon, executive director of the Colorado Beacon Consortium, said recently. “Access to more actionable data within a process driven by clinical judgment and shared patient decision-making improves the ability of a practice team to proactively align resources with patient needs.”

Nevertheless, the data available for predictive modeling today has some serious deficiencies. Claims data is neither timely nor precise; clinical data is usually limited to a single organization; and patient-reported data, except for patient satisfaction surveys, is largely missing. Until the information that healthcare organizations can apply to predictive modeling improves, it will be more useful for some purposes, such as risk stratification of populations, than for others, such as predicting the health risks of individuals with a high degree of accuracy.

Nevertheless, some healthcare systems are beginning to combine claims and clinical data in ways that enable them to use predictive analytics more effectively. And as risk-bearing organizations seek to engage patients in their own care, they are beginning to recognize the importance of patient-reported data.

**Claims Data**

Claims data usually lags the provision of services by one to three months, but it offers the broadest view of the healthcare services that patients have received and the prescriptions they’ve filled. In the view of Jonathan Weiner, a professor of health policy at Johns Hopkins University, accountable care organizations (ACOs) and other entities that manage population health will be heavily dependent on claims data for the next decade or longer.

For purposes of calculating the financial costs and risks of a particular patient population, there is no substitute for claims information. The clinical data available to a healthcare organization is generally limited to the care provided within that enterprise, but everybody who provides services to insured patients submits claims.

Some health plans make claims data available to providers and/or ACOs. Other healthcare organizations that are self-insured employers have begun the journey toward population health management by using the claims data for their own employees. But unless an organization includes a health plan—such as Kaiser Permanente, HealthPartners, or Geisinger—it is unlikely to have access to complete claims or encounter data for most or all of its patient population.

The analytic tools now available to healthcare providers are mainly those that insurers have historically applied to claims. Today, when clinical data is combined with claims, it must be integrated into that framework.
But eventually, the approach to predictive modeling will become more clinically oriented, and claims data will be used to round out the picture.

**Clinical Data**

The spread of electronic health records in recent years has led to a massive growth in the amount of digitized clinical data. But much of this data is unstructured, making it unavailable to predictive modeling and other analytic tools. Moreover, because patients receive healthcare from multiple providers, clinical data generated by one organization may not be sufficient to describe what has happened to a patient or that person’s current health status. Health information exchange is improving, but still has a long way to go.

According to a HIMSS Analytics white paper on analytics, the data challenges to healthcare organizations include:

- Getting data into the system in a structured way, whether it’s collected on paper or comes from another source, such as prescription fill data from pharmacies
- Issues with extracting data from source formats and combining them into a usable aggregated database
- Missing data elements required for analysis. In some cases, this occurs because providers fail to enter data in the correct fields. Data may also be unobtainable if providers cannot exchange information electronically or if the data is housed in multiple databases within an enterprise

Even if an organization has an enterprise data warehouse, it might find that it takes too long to aggregate and normalize the data for the predictive analytics that are used in care management. A healthcare system might solve this problem by building a registry within the warehouse and making sure that the registry receives updates on clinical data, such as lab results, within 24 hours of it becoming available. In an organization that includes multiple inpatient and ambulatory EHRs, one solution is to create a registry that receives data directly from the organization’s or ACO’s internal health information exchange.

**Patient-Reported Data**

To increase the accuracy of predictive analytics and risk stratification, it is essential to obtain information on how patients regard their own health status, their non-clinical risk factors, including health behavior, and their obstacles to managing their own health. Some of this data can be collected during visits to their providers, but much of it changes continually and must be gathered between visits or after discharge from the hospital. Consequently, organizations must provide ways for patients to report this data themselves on a regular basis.

The importance of patient-reported data cannot be overestimated. For example, a particular patient might be considered at moderate risk based on clinical data such as slightly elevated blood pressure and obesity. But that patient’s propensity to become seriously ill is much greater if one considers their lifestyle, socioeconomic status, and ability to obtain healthy food. The chance of a recently discharged patient being readmitted, similarly, will be higher if that person has no one to take care of him or her at home, is depressed, and can’t afford the copayments for prescription drugs.

Among the types of patient surveys that have been developed for collecting information pertinent to health risks are health risk assessments (HRAs), patient activation surveys, and functional status surveys. HRAs, which are used mostly by large, self-insured employers, ask people about a wide range of health and lifestyle factors. Activation instruments measure a patient’s knowledge, skills and confidence in managing their own healthcare. Functional status surveys, which some providers use to measure outcomes after hospitalizations or post-acute care, ask patients how they’re feeling and how well they’re functioning. Both generic and condition-specific instruments are available for this purpose.

The use of patient-reported data in predictive modeling is rare today. But the Cincinnati Beacon Community—one of 17 around the country that are funded by the government to explore the frontiers of health IT—has used HRAs to help hospitals reduce readmissions. Some hospitals and rehabilitation facilities use functional status surveys, but the data from them is not being entered in EHRs.
Conclusion

Predictive analytics are emerging as must-have tools for any organization that wants to do population health management. These analytics cover a wide range of applications, including those that forecast patients’ future health, classify them by their current health status, predict hospitalizations and readmissions, and adjust providers’ performance evaluations by their case mix. In addition, predictive modeling is being used extensively to help organizations calculate the likely cost of caring for a particular population. This is an increasingly important function as more and more organizations take financial risk for care.

Predictive modeling has some serious limitations. The biggest challenges have to do with the available data. Claims and clinical data each have their own issues, and patient-reported data—which could form a much fuller picture of a patient’s situation—is largely missing. But predictive analytics are already invaluable tools in the new healthcare delivery models. As the data improves and new algorithms are devised, their value will increase further, but only if they’re connected to workflow automation solutions that make their insights actionable.

Five Ways to Leverage Predictive Modeling

1. **Risk stratification.** Classify patients as low, medium or high risk. Use that information to allocate resources at a population-wide level, identify high-risk patients, alert providers and care managers about those patients, and design interventions to prevent other people from becoming high risk.

2. **Workflow automation.** Couple predictive modeling with automation tools that enable providers to reach out to patients with care gaps and allow care managers to touch more patients in various ways, ranging from high-touch case management to web-based education and coaching.

3. **Readmission prevention.** Use preventive modeling to identify which patients are most likely to be readmitted. Intervene with these patients so they receive the support they need to avoid readmission.

4. **Provider attribution and risk adjustment.** Apply risk adjustment to evaluate the performance of individual providers, sites, and your whole organization in comparison to others. Use risk adjustment to measure variations in care, improve quality, and show payers how your organization ranks in utilization and quality.

5. **Financial risk calculations.** Employ predictive modeling to calculate how much care delivery will likely cost for your population in the coming year. Use these figures to determine whether the organization will lose or make money under proposed risk contracts.
Notes


11. Terry, “ACOs Need Claims Data for Analytics.”


17. Personal communication with Ron Russell, Verisk Health.
20. “Clinical Analytics in The World of Meaningful Use.”
23. Office of the National Coordinator for Health IT, Factsheet, Greater Cincinnati Beacon Collaboration (Cincinnati, OH).