Predictive analytics: Poised to drive population health

White Paper

As health care moves toward value-based payments and accountable care, providers need better tools for population health and risk management. The ability to prevent unnecessary hospitalizations is a major piece of the puzzle. Doing this well means proactively identifying the highest-risk patients and prioritizing them for care coordination and targeted interventions.

Historically, providers have struggled to accurately identify these high-risk patients. Some providers have created patient lists by manually combing through EMR data. This proves to be a resource-intensive and inaccurate process. Predictive analytics has long held the promise of solving this problem. Only recently, however, has data become available in sufficient quality and quantity to bring predictive analytics closer to fulfilling that promise. Powered by vast quantities of high fidelity clinical and claims data, predictive analytics can identify high-risk patients with greater speed and accuracy than ever before.

Why today is the tipping point for predictive analytics in health care

Predictive analytics uses regression models on underlying data to predict outcomes. This is not new to health care. The challenge, in past years, has been the underlying data. Up until a few years ago, the main source of digital health care data was claims data. But claims data does not get at a patient's overall health or disease-specific functioning. This clinical data was often handwritten, dictated, or incomplete. Hence, predictive modeling relied on relatively small data sets, often of poor quality, and with limited variables. The result was marginally predictive models.

Because of the recent and rapid adoption of EMRs, large and diverse digital health care data sets are now available. Technology also exists that can aggregate this clinical data with other data sources, from across care settings and organizations. This data can then be better structured to enable analytics. Natural language processing (NLP) can be used to access unstructured data. The result is "bigger and better" data with higher predictive ability.

Many other industries have already benefited from today's inexpensive data storage and massive parallel computing power. Health care has lagged because data availability lagged. But the quantity and quality of digital data that is required now exists.

	OLD MODELS	**MODERN** MODELS
SIZE OF DATA SET	• **Limited** digital health care data	• **Large** sets of digital health care data
SOURCES OF DATA	• Usually **claims** data • Usually inpatient data only	• **Clinical** (EMR) + **claims** + **socioeconomic** + **care management** data • Data from all care settings (inpatient, outpatient, ED, etc.)
QUALITY OF DATA	• **Poor** data fidelity • No way to access unstructured data	• **Excellent** – cleaned, normalized, validated • **Natural language processing (NLP)** to access unstructured data
	Poor predictive abilities (c-stats of <0.7)	*Good predictive abilities (c-stats of 0.7–0.8+)*

Figure 1: Modern predictive analytics in health care

How providers can maximize the power of predictive analytics

Now that the data is available, the question is how can providers best use it? First, providers need a solution that leverages a data set of robust size, scope and quality. This enables the best predictive models, which providers can then apply to their own patients. This is where usability comes in. An analytics solution must incorporate provider data in a timely way, and it must answer important questions in an actionable way.

Predictive power: Size, Sources and Quality of Data Set

The size of a sample data set matters when it comes to building predictive clinical models.[1] As a sample size grows, the level of a model's uncertainty and degree of bias decreases. Increasing the sample size increases the chances of seeing all likely events and patient variation. Using a large and diverse sample of patients from many health systems, geographic regions and demographics, you can reduce the likelihood of skewing a study with a homogeneous sample. For example, a small sample of patients from the same health system with similar demographics would not effectively represent the larger heterogeneous congestive heart failure (CHF) population because it does not take into account the variability of all possible independent variables that could lead to a dependent event. Hence, the relative predictive power of a statistical model increases exponentially when using millions of patients instead of hundreds of patients.

Relevant and varied data sources are critical for uncovering the most predictive variables. For this reason, clinical, claims, socioeconomic and care management data should all be integrated into one data set. Models that use little or no clinical data have an incomplete picture of a patient's health. Claims data indicates whether a procedure occurred, but not what actually happened to a patient's health. If a physician were asked to predict a patient's outcome, clinical status would be paramount — whether lab tests, vital signs or comorbidities. A strong model must therefore include clinical data in addition to claims data.

Research also points to the importance of socioeconomic and care management data for predictive analytics in health care. In 2011, the Veterans Health Administration (VHA) reviewed risk prediction models specifically for hospital readmission.[2] It found that social and environmental factors such as access to care, social support, and substance abuse contributed to readmission risk in some models. In addition, the authors discuss how care management factors such as discharge follow-up and coordination of care with primary care physicians likely impact readmission risk.

Finally, data quality is paramount. It is one thing to aggregate data from diverse sources, settings and organizations, but unless the resulting set is cleaned, normalized and validated, it will not be useful in predictive models. Each data set must be prepared the same way, using the same clinical classification (a single consistent ontology) or else statistical power will suffer. In addition, natural language processing capabilities will add breadth and depth to the data.

Usability

For a provider to deploy predictive modeling in their organization, their own data must run through the models in a timely manner. Otherwise, they risk predicting outcomes too late to do anything about them. Data should be extracted automatically and continuously. Some clinical data elements can and should be incorporated in as little as two days from occurrence. Claims data, on the other hand, is 2–6 months old by the time it is typically reported.

While timeliness is one side of usability, another is the ability to generate actionable insights. Predictive models should predict events that have significant impact on the quality and cost of care. Importantly, providers must also be able to positively impact these events. A model may be able to predict patients that will develop CHF, but it is not clear that providers can play a role in preventing this outcome. However, for patients that already have CHF, providers can take concrete actions to help prevent hospitalization. Similarly, other chronic conditions such as diabetes (DM), chronic obstructive pulmonary disorder (COPD) and asthma also have well-proven approaches for admissions prevention — and are great candidates for predictive modeling.

In addition, once outcomes are predicted, providers need to be able to easily see what actions to take, for which patients. They should be able to easily stratify patients based on their level of risk, and then prioritize resources accordingly. This risk stratification should also link to actionable lists that highlight each patient's clinical status and specific care gaps.

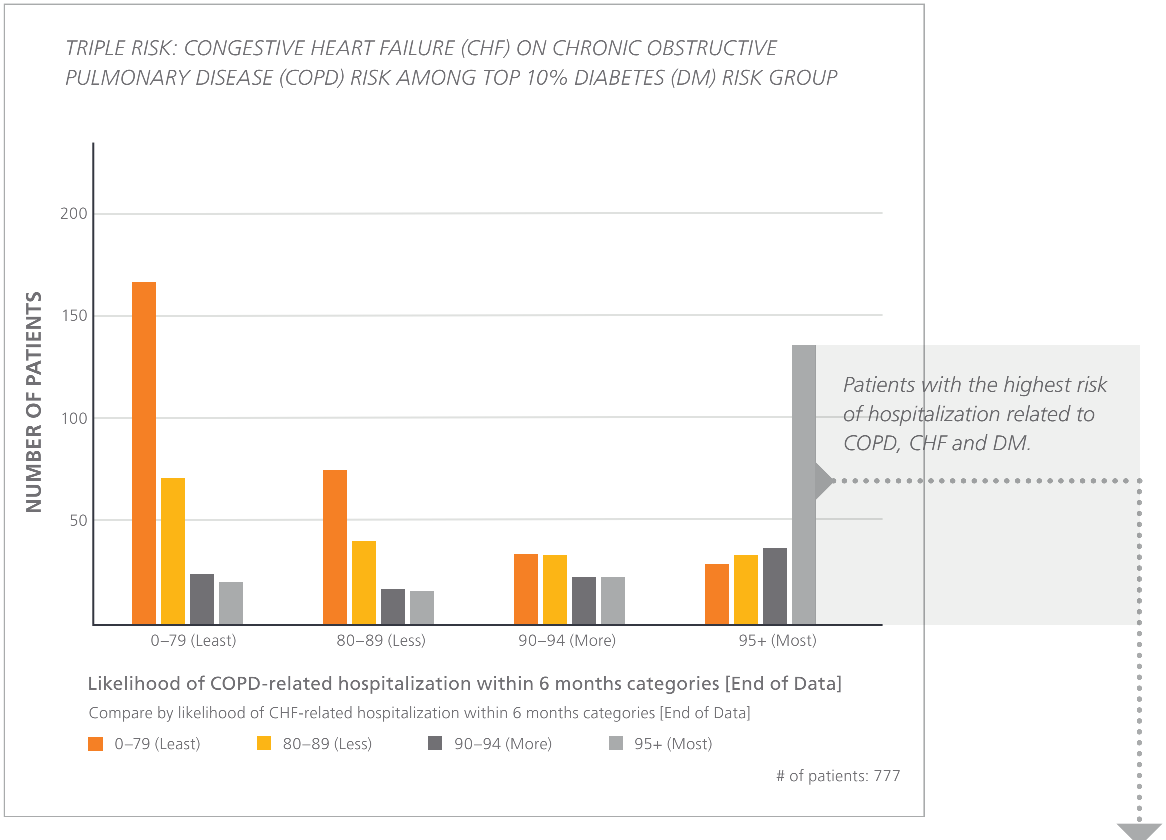
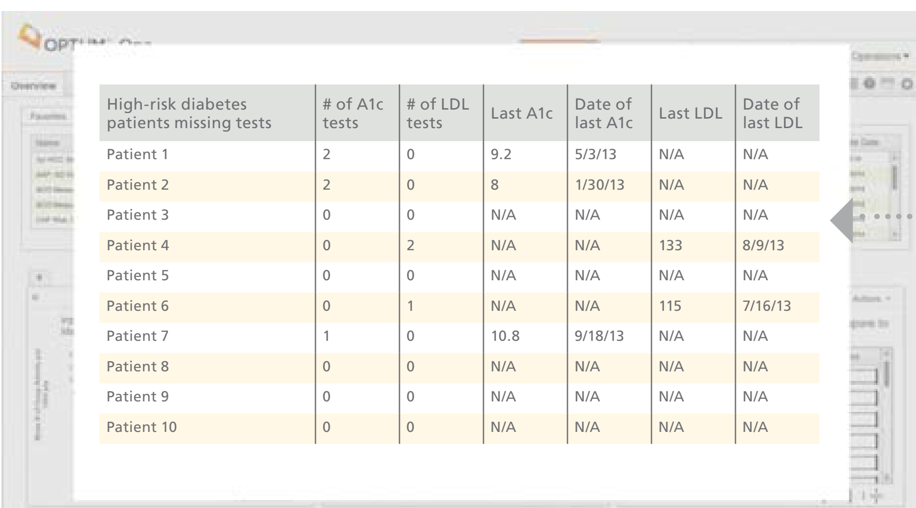


Figure 2: Predictive models help providers focus resources on highest-risk patients



High-risk diabetes patients missing tests	# of A1c tests	# of LDL tests	Last A1c	Date of last A1c	Last LDL	Date of last LDL
Patient 1	2	0	9.2	5/3/13	N/A	N/A
Patient 2	2	0	8	1/30/13	N/A	N/A
Patient 3	0	0	N/A	N/A	N/A	N/A
Patient 4	0	2	N/A	N/A	133	8/9/13
Patient 5	0	0	N/A	N/A	N/A	N/A
Patient 6	0	1	N/A	N/A	115	7/16/13
Patient 7	1	0	10.8	9/18/13	N/A	N/A
Patient 8	0	0	N/A	N/A	N/A	N/A
Patient 9	0	0	N/A	N/A	N/A	N/A
Patient 10	0	0	N/A	N/A	N/A	N/A

Figure 3: Automated patient lists highlight specific care gaps for care coordination

Today's hospital admissions models: Strong evidence of predictive ability

Statistical validation to assess predictive ability

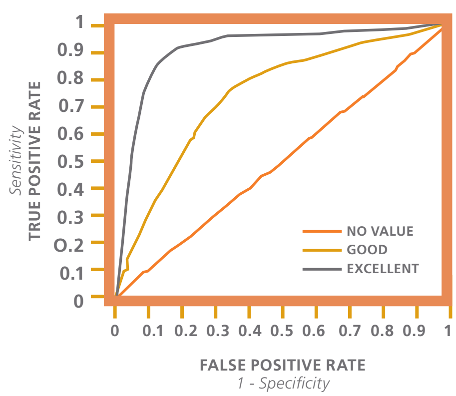
Predictive models are commonly assessed using c-statistics (equivalent to the "area under the receiver operating curve"). This metric describes a model's ability to predict a positive or negative outcome such as a hospital admission. A c-statistic of 0.5 indicates a model that performs no better than chance. In general, if a model's c-statistic exceeds 0.7 it is considered to have acceptable predictive ability.

In the Veterans Health Administration 2011 review of risk prediction models for hospital readmission, most models had poor predictive ability.[3] Of the studies they reviewed, 21 reported a c-statistic and only six of them had c-statistics above 0.7. Many models relied upon retrospective administrative data, and these tended to have poor discriminative abilities. The review highlighted the need for future models to consider variables related to severity of illness, overall health, and social determinants of health.

QUICK TUTORIAL

COMPARING ROC CURVES

Receiver Operating Characteristic (ROC) curves measure a model's predictive ability by plotting both sensitivity and specificity



COMPARING C-STATISTICS

A c-statistic = the area under the ROC curve

EXCELLENT	GOOD	ACCEPTABLE	POOR	FAIL
0.9–1.0	0.8–0.9	0.7–0.8	0.6–0.7	0.5–0.6

SOURCE: *http://darwin.unmc.edu/dxtests originally from Metz CE. Basic principles of ROC analysis. Sem Nuc Med. 1978;8:283-298*

Figure 4: Predictive models are commonly assessed using c-statistics

A lot has changed in the past few years. Today's models can predict hospital admissions in a range of conditions such as CHF, DM, COPD and asthma. Because it is easier to integrate data from outpatient and inpatient settings, models are able to predict both initial admissions and readmissions. And, in contrast to the VHA's findings, today's models have been statistically validated with strong results. For example, Optum has developed predictive hospital admissions models with c-statistics of 0.757 for CHF, 0.833 for COPD, 0.765 for DM, and 0.784 for pediatric asthma. Unlike past models, these models rely heavily on clinical data, which includes actual lab results to better determine risk levels.

As data sets grow in size and scope, today's models can be retrained and become even more robust. For example, Optum's CHF model had an original c-statistic of 0.733 for its IDN model, which was based on data from a total patient pool of 20M patients (inclusive of all conditions). Retraining at a later date with an additional 10M patients resulted in a c-statistic of 0.757. In addition, efforts are now being made to broaden the scope of variables that can be included. Including data from care management assessments, for example, will allow for new variables related to patients' psychosocial backgrounds, barriers to care, and quality of care. Evolving data sets in this way will push predictive abilities even further.

Area Under the Receiver Operating Curve (C-STATS)

HOSPITAL ADMISSIONS MODELS	IDN MODEL	NON-IDN MODEL
CONGESTIVE HEART FAILURE MODEL		
Training sample	0.757	0.742
Avg of testing samples	0.739	0.708
CHRONIC OBSTRUCTIVE PULMONARY DISEASE MODEL		
Training sample	0.833	0.802
Avg of testing samples	0.830	0.799
DIABETES MELLITUS MODEL		
Training sample	0.765	0.754
Avg of testing samples	0.781	0.765
PEDIATRIC ASTHMA MODEL		
Training sample	0.784	0.739
Avg of testing samples	0.761	0.716

NOTE: *Models developed using data from over 30M patients (inclusive of all conditions). All models predict both initial admission and readmission, for both inpatient and emergency department. Pediatric asthma model also predicts observation visits.*

Figure 5: Examples of statistically validated admissions models from Optum

Post-hoc analysis to assess predictive ability

To further assess the utility of today's predictive hospital admissions models, a team at Optum conducted a post-hoc analysis on CHF patients who did not receive any intervention. Of the highest-risk patients (the riskiest 5% as predicted by the model), an average of 55% were admitted within six months. This illustrates how powerful predictive models can be for identifying candidates for targeted care coordination. The 55% is based on over 30 months of data, across 17 medical groups with a total of 1.2M CHF patients. While the result varied by month, the lowest monthly average was 52% and the highest was 62%.

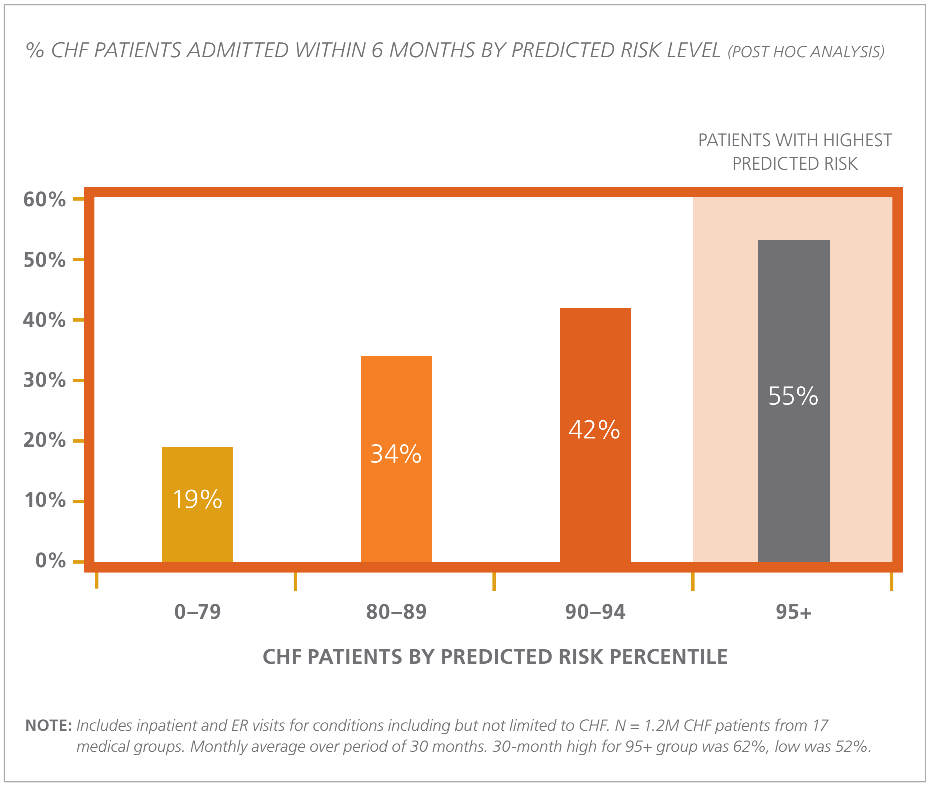


Figure 6: 55 percent of highest-risk CHF patients were admitted within six months

In addition to monthly variation, predictive ability also varied significantly amongst medical groups. At the high end, one medical group saw 64% of their highest-risk patients admitted within six months. The group with the lowest percentage was 49%. The main driver of the variation was access to both clinical and claims data in both inpatient and outpatient settings. Most of the groups that saw relatively low predictive results either lacked clinical inpatient data, or lacked inpatient data altogether. In contrast, the groups with more diverse and complete data tended to see the strongest results (p-value <.0001).

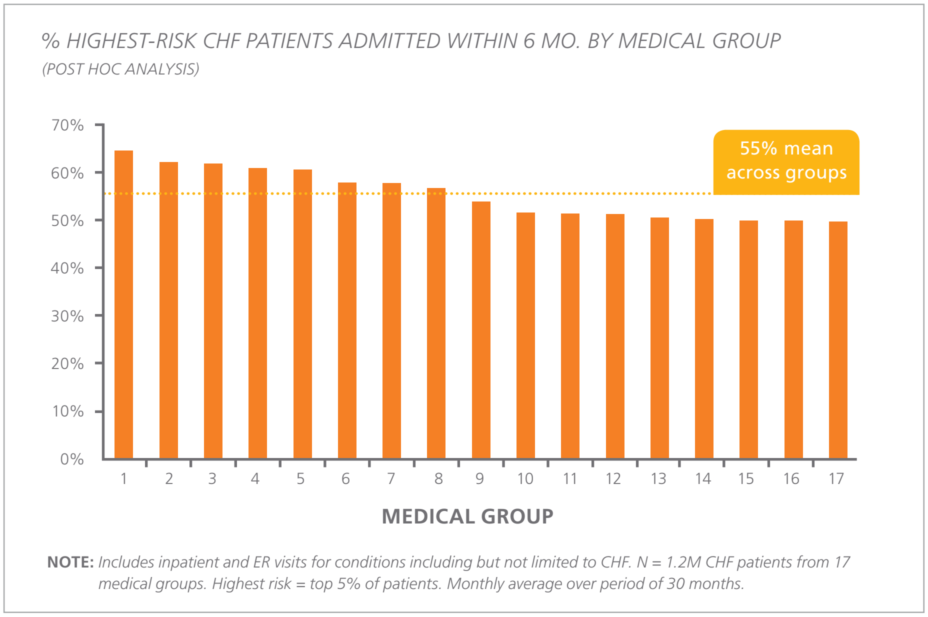


Figure 7: Predictive ability varies by medical group

Providers actively piloting predictive analytics

Many providers have begun to use predictive models to identify high-risk patients, particularly those at risk of hospital admissions. Sentara Medical Group is one such provider. Sentara has 380 primary and specialty care physicians in Virginia and North Carolina. They recently received the NCQA's highest level of recognition for their Patient Centered Medical Home (PCMH) program because of their commitment to chronic disease management and service excellence. They are using Optum One, (Optum's combined analytics and care management platform) to take this model of care even further.

Sentara uses Optum One's predictive analytics to stratify patients with CHF, COPD and DM by risk of future hospitalization. They began by selecting 11 pilot sites called "Transformation of Care Sites." The Quality team used predictive analytics to identify a small number of high-risk patients for each site, and equipped them with detailed individual patient information via a "Patient Profile." This profile provides a 36-month view of an individual patient's clinical parameters (e.g., BMI, blood pressure, ejection fraction), utilization parameters (e.g., ER visits), and treatments (e.g., medication changes). This visual picture of what is going on with a patient helps physicians recognize and act on gaps in patient care or changes in health status.

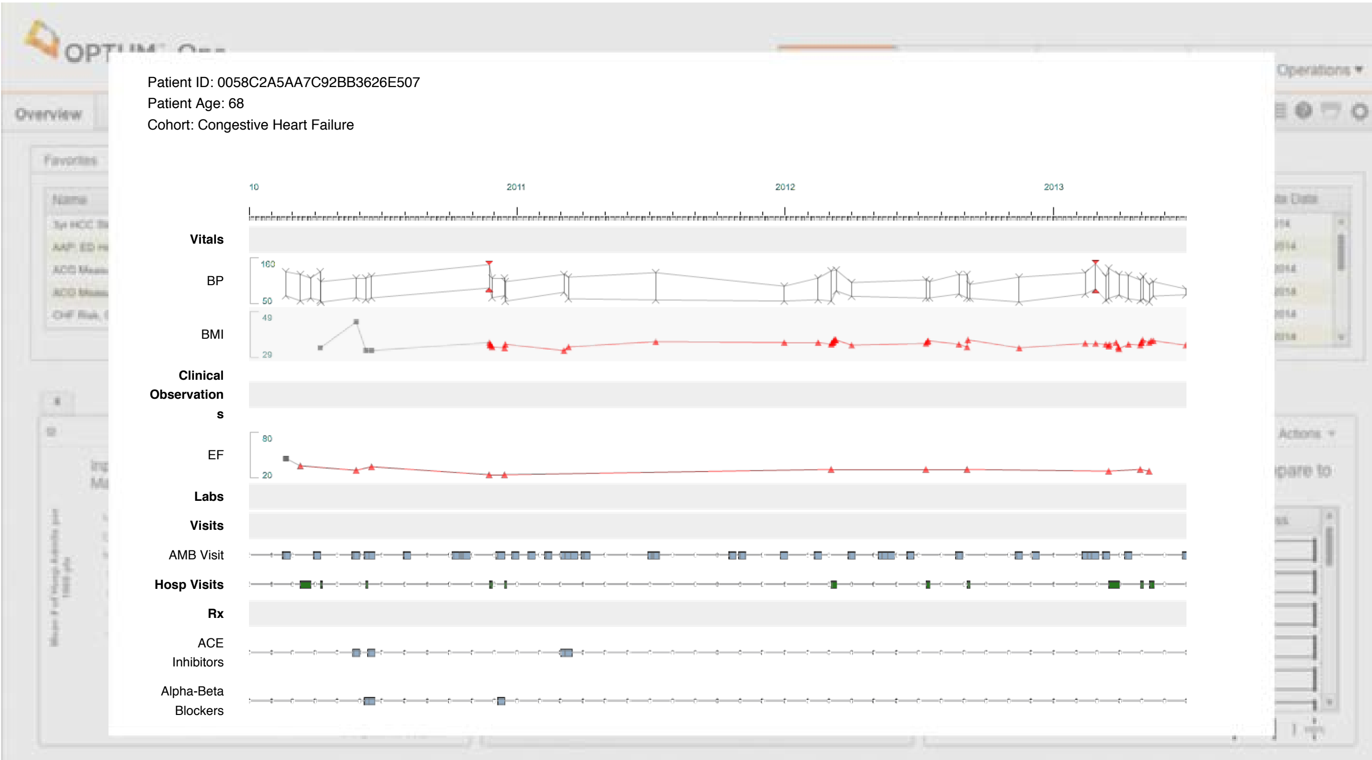


Figure 8: Patient Profiles equip providers with holistic view of high-risk patients

Sentara's early results from its use of predictive analytics are promising. They have been well received by physicians and have had a significant impact on high-risk patient lists. In one practice, of the 44 high-risk patients identified, only one had been part of previous high-risk lists. In addition, rates of engagement in care coordination programs have improved, attaining 50%+ of eligible patients in some cases. Sentara is now expanding its use of predictive analytics to the remaining PCMHs and is also introducing the pediatric asthma model as an additional tool.

A tool for moving from reactive to proactive care

Sentara is one of many providers beginning to integrate predictive analytics into their organizations. They are using it to help rebalance their care model in favor of more proactive care and less reactive care. By homing in on high-risk patients sooner and with more accuracy, providers can focus their resources where they will have the highest impact, and succeed in an environment rapidly moving toward value-driven health care.

Sources

[1] P.M. Lukacs et al., Model Selection Bias and Freedman's paradox, Ann Inst Stat Math (2010) 62:124.

[2] Devan Kansagara et al. Risk Prediction Models for Hospital Readmission (Department of Veterans Affairs, Health Services Research & Development Service: October 2011). http://www.hsrd.research.va.gov/publications/esp/readmission-REPORT.pdf

[3] Ibid.

About Optum

Optum is an information and technology enabled health services company serving the broad health care marketplace, including care providers, health plans, life sciences companies and consumers and employs more than 65,000 people worldwide. For more information about Optum and its products and services, please visit **www.optum.com**.

About Humedica

Humedica, an Optum company, is the foremost clinical intelligence company that provides private cloud-based business solutions to the health care industry. Humedica's sophisticated analytics platform transforms disparate clinical data into actionable, real-world insights. Powered by the largest and most comprehensive clinical database, Humedica solutions move beyond claims data to offer a more complete, longitudinal view of the patient population. Through its award-winning solutions, Humedica, which is headquartered in Boston, empowers its partners and customers to make confident, value-based decisions about patient care in a rapidly changing health care market.

OPTUM™ T 800.765.6619 | **www.optum.com**

13625 Technology Drive, Eden Prairie, MN 55344