

4 Essential Lessons for Adopting Predictive Analytics in Healthcare

By David Crockett, Ph.D.

“The willingness to intervene is the golden key to harnessing the power of historical and real-time data.”

Information plus context equals knowledge. But predictions made solely for the sake of making a prediction are a waste of time and money.

Predictors are most useful when their knowledge can be transferred into action. The willingness to intervene is the golden key to harnessing the power of historical and real-time data. More importantly, to best judge the efficacy and value of forecasting a trend and ultimately changing behavior, both the predictor and the intervention must be integrated back into the same system and workflow where the trend originally occurred. Case in point: mediating hospital readmissions.

THE ECONOMY OF PREDICTION

Academically speaking, predicting hospital readmissions is a very active topic. Thus far in 2013, 36 peer-reviewed journal articles have been published on the subject along with three additional review articles. Highlighting this rapidly growing interest are recent papers focused on simplified readmission scoring for elderly patients,¹ the relationship between readmission and mortality rates,² and a systematic review of tools for predicting severe adverse events.³ Prediction discussions associated with specific areas such as heart failure^{4,5} or within pediatric populations are also very active.^{6,7}

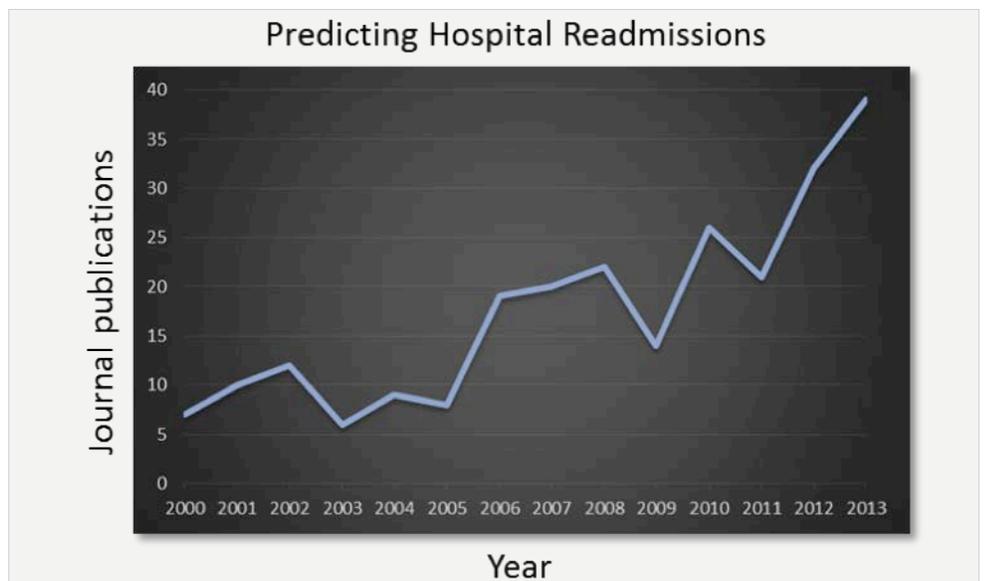


Figure 1. Opportunities for improved patient care and motivation to avoid financial penalties have resulted in a rapidly growing interest in predicting hospital readmission.

...to best leverage the data, predictors should not be used in isolation.

This year alone, researchers are on track to publish approximately the same number of papers on using prediction in healthcare as were published in the entire 1990s. The motivation? Improving patient care while avoiding financial and reimbursement penalties for hospitals.

Research among academics aside, the common challenge remains: how can the industry successfully move promising ideas from academic research circles to fully developed and working implementations in a live hospital IT environment? And how can predictive analytics be used to help control costs and improve patient care?

WHAT YOU SEE IS WHAT YOU GET

Evidence-based medicine is a powerful tool to help minimize treatment variation and unexpected costs. Best-practice guidelines contribute further to the goal of standardized patient outcomes and controlling costs. But a well-known Chinese proverb states, “jùng dōu shòu dōu, jùng gwà, shòu gwà.” (If you plant beans, you get beans. If you plant squash, you get squash.) Algorithm and computer types know this better as “garbage in, garbage out.” Similar to notation in the Six Sigma approach to quality improvements, for predictive analytics to be effective, Lean practitioners must truly live the process to best understand the type of data, the actual workflow, the target audience and what action will be prompted by knowing the prediction. In military intelligence terms, this is “boots on the ground;” in a hospital setting, “nurses on the floor.”

In short, decision makers cannot be isolated or far removed from the actual point of decision. At the same token, to best leverage the data, predictors should also not be used in isolation, although in the healthcare industry today, readmission risk profilers are often used as standalone applications.

Returning to the Chinese garden, Confucius is credited with saying, “If you think in terms of a year, plant a seed; if in terms of 10 years, plant trees; if in terms of 100 years, teach the people.” For the long-term success of predictive analytics in healthcare, it’s necessary to do all of the above. However, the process can be jump-started by learning from other industries and expertise.

EXISTING EXPERTISE

Fortunately for healthcare, there are numerous existing models from other industries that are very efficient at risk stratification in the realm of population management. Take, for example, the casino industry, which carefully models and manages population risk in terms of how many people walk through the door and how much the pay outs will cost the casino.^{8,9} Similarly, the statistical work of actuaries in the task of managing population life insurance risk and payout is also well understood.^{10, 11}

In healthcare, the irony of a technology-driven, more generalized prediction model that inputs big data and global features is that the targeted utility is usually lost. Prediction focused on a specific clinical setting or patient need will always trump a generic predictor in terms of accuracy and utility.

Beyond industry expertise, studying history will likely ease some of the potential pains and pitfalls that could accompany healthcare's adoption of predictive analytics. Key lessons are listed below; illustrations of each follow.¹²

1. **More data does not equate to more insight:** It can be difficult to extract robust and clinically relevant conclusions, even from reams of data.
2. **Insight and value are not the same:** While many solid scientific findings may be interesting, they do little to significantly improve current clinical outcomes.
3. **Ability to interpret data varies based on the data itself:** Sometimes even the best data may afford only limited insight into clinical health outcomes.
4. **Implementation itself may prove a challenge:** Leveraging large data sets successfully requires a hospital system to be prepared to embrace new methodologies; this, however, may require a significant investment of time and capital and alignment of economic interests.

For the healthcare industry, like other industries, predictors will always be more useful in the framework of a more complete set of data, where the knowledge can be fully leveraged to action. Furthermore, full clinical utility of prediction or risk stratification is only possible in a data-rich enterprise warehouse environment. But perhaps most importantly, these predictor-intervention sets can best be monitored and measured within that same data warehouse environment. If the predictor is used standalone or housed elsewhere (siloes), this important evaluation step may not be possible.

SPECIFIC TRUMPS GLOBAL

Lesson #1: Don't confuse more data with more insight. This first lesson from history is similar to the love affair we humans have with new technology. The sometimes not-so-obvious irony is that without having the proper technology framework in place, with context and metadata for meaningful use, new technology is really not very useful. In fact, it is often a waste of time and money. Thus in healthcare, the irony of a technology-driven, more generalized prediction model that inputs big data and global features is that the targeted utility is usually lost. Prediction focused on a specific clinical setting or patient need will always trump a generic predictor in terms of accuracy and utility. The following example from genetics illustrates this concept.

Thyroid cancer can be caused by mutations in the RET gene. This proto-oncogene is a tyrosine kinase that modulates phosphorylation signals on the cell surface into the cell. In RET mutations, a very specific protein location (transmembrane) and certain amino acid residue (cysteine) is changed more frequently (hotspots) as compared to surrounding locations. These hotspot changes are very characteristic to RET kinase and can be modeled with high accuracy when looking only at the RET gene. But if this gene-specific situation is diluted into a mutations data set derived from the entire genome, that specificity and accuracy are completely lost.¹³

The very features that characterize a condition well are the attributes that can train an accurate predictor. But if those features (variables) do not stand out above the background noise, the predictor only predicts the noise well. The full power of prediction is best realized when specific variables are gathered, a targeted clinical need is met and participants are willing to act.

Comprehensive outcomes data is often missing in our current healthcare system. By not capturing the final outcome the utility of machine-learning tools is severely limited and thus becomes one of the obstacles to widespread adoption and trust.

INTEGRATED PREDICTION

Lesson #2: Don't confuse insight with value. People who work in a database discipline understand that data plus context equals knowledge. In a similar vein, prediction in a comprehensive data warehouse environment is superior to standalone applications, as illustrated by the potential synergy of the existing Rothman Index, an early indicator of wellness.¹⁴ This proven algorithm captures trends from multiple data feeds of vital signs, lab values and nursing assessments. This data, taken as a whole, will often provide early warning as a patient begins to fail, where even a careful human observer cannot possibly connect the dots between so many unrelated data points simultaneously. One key to the success of the algorithm is first obtaining all of the necessary data. Assessing only part of a picture often yields an incorrect view.

LEVEL OF TRUST

Lesson #3: Don't overestimate the ability to interpret the data. Another interesting irony is the level of trust that people place in computational prediction or forecasting. With hurricane season winding down, isn't it comforting to know that computer models accurately predicted seven of the last three storms? Yes, you read that right, the computer overshoot by more than 200%. What it really comes down to is that comparison between the weather forecast and someone whose joints ache whenever a storm is coming—which one is truly more accurate? At the end of the day, it either rains or it doesn't, regardless of what the forecast said. It's all about the outcome.

The same holds true for predictions in healthcare. The complication is that comprehensive outcomes data is often missing in our current healthcare system. By not capturing the final outcome the utility of machine-learning tools is severely limited in this particular setting and thus becomes one of the obstacles to widespread adoption and trust. Without a class outcome (label) to train the algorithm, supervised (structured, forward chain) models cannot be easily built.

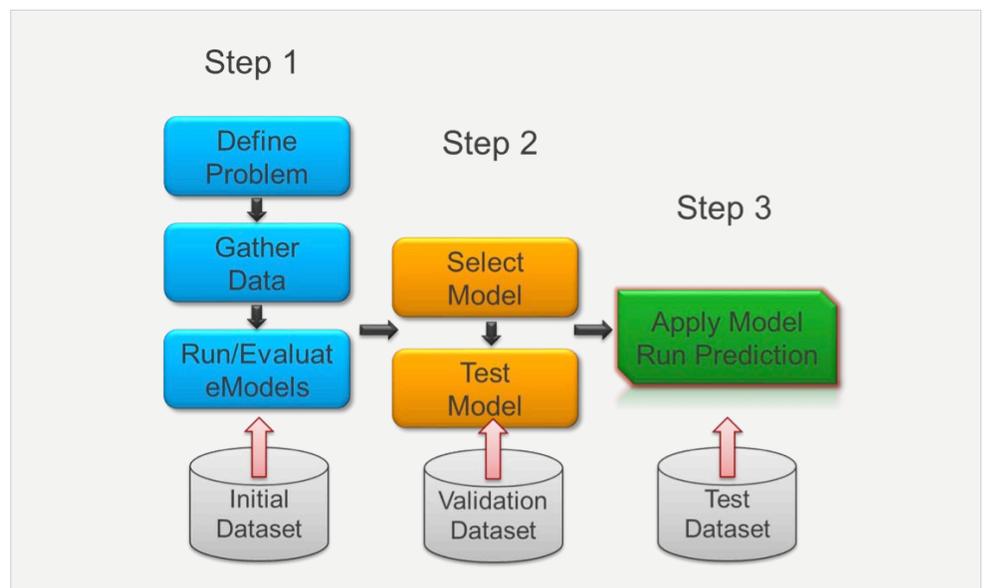


Figure 2. Overview of the machine learning modeling process.

“Healthcare providers need to partner with groups that have a keen understanding of the leading academic and commercial tools, and the expertise to develop appropriate prediction models.”

In reality, however, clinicians make judgments and medical decisions using incomplete information every day. Granted, these are typically sound judgments based on training, past experience and collective knowledge of trusted colleagues. But at the end of the day, treatment decisions made on incomplete information and educated guesses are quite common in the current health system.

In the end, the goal is the same: to leverage historical patient data to improve current patient outcomes. Predictive analytics is a powerful tool in this regard.

STATE OF THE INDUSTRY

Lesson #4: Don't underestimate the challenge of implementation. So many options exist when it comes to developing predictive algorithms or stratifying patient risk. This presents a daunting challenge to health care personnel tasked with sorting through all the buzzword and marketing noise. Healthcare providers need to partner with groups that have a keen understanding of the leading academic and commercial tools, and the expertise to develop appropriate prediction models. Representative examples of open source tools include popular software such as R and Weka. The statistical package R can be found on The Comprehensive R Archive Network (CRAN) at <http://cran.us.r-project.org> hosted on servers at the Fred Hutchinson Cancer Research Center. This is a widely used open source tool, with thousands of specific libraries or “packages” for a variety of applications. As of September 2013, the CRAN package repository features 4,849 available packages and that number is growing exponentially. Packages are user submitted (shared) to assist with statistical computing for topics such as biology, genetics, finance, neural networks, time series modeling and many others.

The Waikato Environment for Knowledge Analysis (Weka) incorporates several standard machine learning techniques into a software workbench that's issued under the GNU General Public License. This is a Java implementation of tools for data pre-processing, feature selection, classification, regression, clustering, association rules and visualization, hosted by the Computer Science Department at the University of Waikato in New Zealand, <http://www.cs.waikato.ac.nz/ml/weka>. Using Weka, a specialist in a particular field can leverage machine-learning methods to discover useful knowledge from datasets much too large to analyze by hand.

Representative examples of commercial offerings include Spotfire and Predixion Software. Spotfire (<http://spotfire.tibco.com>) is a long-standing data visualization tool kit that originated out of the University of Maryland in 1996 and acquired into the TIBCO product suite in 2007. Similar to SAS, SPSS and other analytics vendors, Spotfire supports in-database analysis in conjunction with Oracle, Microsoft SQL Server and Teradata data warehouse platforms. Visualization applications cover a wide range of industries and expertise. The software is built to deploy reports across cloud, in-house hardware and mobile devices. Current versions include predictive capabilities such as tools for regression, clustering and custom R scripting.

Predixion Software (<http://www.predixionsoftware.com>) has recently introduced a new release of its collaborative predictive analytics platform, Predixion Enterprise Insight.™ The software is also available in the cloud, on-premise or as a managed

Prediction should link carefully to clinical priorities and measurable events such as cost effectiveness, clinical protocols or patient outcomes. These predictor-intervention sets can be evaluated most effectively when they're housed within that same data warehouse environment.

appliance. Features include data-preparation tools, model building, performance evaluation and model management tools. While the developer interface is the familiar Microsoft Excel environment, completed models can be deployed in various and flexible ways. The software can also leverage the PMML-based predictive models from SAS, SPSS, R and other platforms.

LESSONS LEARNED

Machine learning is a well-studied discipline with a long history of success in many industries. Healthcare can learn valuable lessons from this previous success to jumpstart the utility of predictive analytics for improving patient care, chronic disease management, hospital administration and supply chain efficiencies. The opportunity that currently exists for health care systems is to define what “predictive analytics” means to them and how can it be used most effectively to make improvements within their system.

In order to be successful, clinical event prediction and subsequent intervention should be both content driven and clinician driven. Importantly, the underlying data warehouse platform is key to gathering rich data sets necessary for training and implementing predictors. Notably, prediction should be used in the context of when and where needed—with clinical leaders who have the willingness to act on appropriate intervention measures. The more specific term is prescriptive analytics, which includes evidence, recommendations and actions for each predicted category or outcome. Specifically, prediction should link carefully to clinical priorities and measurable events such as cost effectiveness, clinical protocols or patient outcomes. Finally, these predictor-intervention sets can be evaluated most effectively when they're housed within that same data warehouse environment.

References

1. Ben-Chetrit, E., et al., A simplified scoring tool for prediction of readmission in elderly patients hospitalized in internal medicine departments. *Isr Med Assoc J*, 2012. 14(12): p. 752-6.
2. Krumholz, H.M., et al., Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. *JAMA*, 2013. 309(6): p. 587-93.
3. Overland, Dina, “Blues Plans Bundle Payments for Knee, Hip Replacements,” *FierceHealthPayer*, May 23, 2012.
4. Hosein, F.S., et al., A systematic review of tools for predicting severe adverse events following patient discharge from intensive care units. *Crit Care*, 2013. 17(3): p. R102.
5. Psofka, M.A. and J.R. Teerlink, Strategies to prevent postdischarge adverse events among hospitalized patients with heart failure. *Heart Fail Clin*, 2013. 9(3): p. 303-20, vi.
6. Bradley, E.H., et al., Hospital strategies associated with 30-day readmission rates for patients with heart failure. *Circ Cardiovasc Qual Outcomes*, 2013. 6(4): p. 444-50.
7. Berry, J.G., et al., Pediatric readmission prevalence and variability across hospitals. *JAMA*, 2013. 309(4): p. 372-80.

8. Bardach, N.S., et al., Measuring hospital quality using pediatric readmission and revisit rates. *Pediatrics*, 2013. 132(3): p. 429-36.
 9. Eadington, W.R., The economics of casino gambling. *The Journal of Economic Perspectives*, 1999. 13(3): p. 173-192.
 10. Croson, R. and J. Sundali, The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty*, 2005. 30(3): p. 195-209.
 11. Frees, E.W. and E.A. Valdez, Understanding relationships using copulas. *North American actuarial journal*, 1998. 2(1): p. 1-25.
 12. Cox, S.H. and Y. Lin, Natural hedging of life and annuity mortality risks. *North American Actuarial Journal*, 2007. 11(3): p. 1-15.
 13. Shaywitz, D. Turning Information Into Impact: Digital Health's Long Road Ahead. 2012 9/17/2013]; Available from: <http://www.forbes.com/sites/davidshaywitz/2012/12/30/turning-information-into-impact-digital-healths-long-road-ahead/>.
 14. Crockett, D.K., et al., Utility of gene-specific algorithms for predicting pathogenicity of uncertain gene variants. *J Am Med Inform Assoc*, 2012. 19(2): p. 207-11.
 15. Rothman, M.J., S.I. Rothman, and J.t. Beals, Development and validation of a continuous measure of patient condition using the Electronic Medical Record. *J Biomed Inform*, 2013. 46(5): p. 837-48.
-



About the Author

David K. Crockett, Ph.D. is Health Catalyst's Senior Director of Research and Predictive Analytics. He brings nearly 20 years of translational research experience in pathology, laboratory and clinical diagnostics. His recent work includes patents in computer prediction models for phenotype effect of uncertain gene variants. Dr. Crockett has published more than 50 peer-reviewed journal articles in areas such as bioinformatics, biomarker discovery, immunology, molecular oncology, genomics and proteomics. He holds a BA in molecular biology from Brigham Young University, and a Ph.D. in biomedical informatics from the University of Utah, recognized as one of the top training programs for informatics in the world. Dr. Crockett builds on Health Catalyst's ability to predict patient health outcomes and enable the next level of prescriptive analytics – the science of determining the most effective interventions to maintain health.