Volume to Value: the Role of Big Data, Machine Learning, and Estimation of Treatment Effects

William H. Crown, PhD, 2013-2014 ISPOR President and Chief Scientific Officer, Optum Labs, Cambridge, MA, USA

Health care systems around the world are struggling with the “volume to value” transition. There are many dimensions to this transition but one key dimension has to do with informing coverage decisions and reimbursing providers and manufacturers based upon the value of clinical interventions, rather than just reimbursing on the basis of providing services. The big data evolution is important because it dramatically improves our ability to measure dimensions of value that were not previously possible. For example, as health care claims data are merged with electronic medical records (EMR) data, it becomes possible to measure relative changes in clinical outcomes per dollar/euro/pound (or other currency) spent.

This has driven a tremendous amount of interest in “big data.” It seems that almost every day there is an article focusing on the power of big data to uncover new insights—particularly insights into consumer behavior. In the life sciences arena, numerous big data partnerships have been formed over the past several years among health plans, research organizations, regulatory authorities and biopharmaceutical firms, in an attempt to pool resources and data to harness the power of big data [1]. The volume, complexity, and currency of health care data are expanding rapidly; there is little doubt that the big data revolution has arrived in health care. The role of this revolution in the move from volume to value, however, is evolving.

Most of the assessments of value currently taking place continue to rely upon traditional statistical methods from epidemiology, health services research, and health economics. There are some early attempts, however, to use new methodologies from the machine learning field to draw insights from health care data. One prominent example is the partnership between Wellpoint and IBM Watson to examine whether it is possible to provide physicians with point of care treatment recommendations based upon clinical data and evidence regarding treatment effectiveness from the literature, coupled with the artificial intelligence and high speed computing power of Watson.

Machine learning approaches use extremely sophisticated algorithms to cluster observations together [2]. The algorithms have exotic names (at least they seem exotic to me) like random forest, support vector machines, neural networks, and lasso regression. Machine learning and traditional statistical models for treatment evaluation are related but different. Multivariate methods for estimating treatment effects attempt to generate unbiased and efficient estimates of an unobserved population treatment effect using sample data. Because there is nothing about multivariate regression models that ensures that two treatment groups are even comparable to begin with, it is common practice to first match the treatment groups using propensity score methods. By predicting the probability of treatment and then matching treated and untreated patients on the basis of having a similar probability of treatment, the groups are balanced on the basis of their observed covariates. Treatment effects are then assessed within these matched samples.

It is interesting that machine learning approaches bear more similarity to the matching process than to the treatment estimation problem. In fact, one could argue that machine learning approaches are sophisticated methods for predicting membership in groups. This also suggests that if one is interested in treatment effects it is not enough to stop with machine learning. Rather, it is the first step in the evaluation process.

One of the challenges faced by health care systems around the world is the effective treatment of patients with multiple co-morbidities. Treatment guidelines are developed for specific diseases such as diabetes, asthma, cardiovascular disorders, etc. For a physician attempting to treat a patient with several such comorbidities, the sum of the individual guidelines might suggest that there are fifteen things the physician should do. But, perhaps, the most effective and efficient protocol would be to do five things in a particular order. Machine learning methods should excel at putting patients into groups. Traditional statistical methods can then be used to evaluate outcomes associated with alternative treatment pathways using the health care data for the patients in a particular group.

Another important characteristic of machine learning is to build the models on one set of data and test them on a separate set of data. This approach is sometimes used in epidemiology and health economics statistical analysis as well but it is fair to say that it is not standard procedure to do so. The concept of “learning” and “test” samples is very valuable because it is all too easy to estimate a model that performs great on one data set but doesn’t perform well at all on another data set. The intuition that I use to think about this comes from the standard ordinary least squares regression...
model. It is fairly easy to show that as the number of variables in a regression model approaches the sample size, the r-square will approach one. Thus, even with very large datasets, models with huge numbers of interactions, etc. can be made to predict group membership with a high degree of accuracy for a given sample. Such models, however, may not predict well at all for a new sample.

As we move down the pathway toward more rapid insights from machine learning approaches, how do we guard against the risk of drawing incorrect conclusions? I am reminded of a paper by John Seeger and colleagues about ten years ago which found that a standard multivariate analysis of statins was associated with elevated risk of subsequent acute myocardial infarction (AMI) [3]. After using propensity score analysis to match patients who received statins with those that did not, however, Seeger et al. found that statins were associated with a significant reduction in AMI risk.

In a situation like this, presenting the physician with the descriptive data on AMI risk for statin treatment, and even traditional multivariate results, would be highly misleading. Only after the propensity score matching was the correct answer obtained. I recently used this as an example at a conference and asked one of the panelists how we could avoid making this kind of mistake. The panelist’s reply was that big data would solve the bias problem. This is not the case.

As I discussed in the last issue of ISPOR CONNECTIONS, big data can help to reduce bias if it facilitates incorporation of missing variables (e.g., through data linkage). But more of the same type of data will just reduce standard errors on biased estimates—that is, make the biased estimates more efficient! In the statin example, the bias was introduced by a purely methodological issue—that lack of common support for the covariates in the treatment and comparison groups. In that particular study, the data were actually quite rich—augmenting claims data with a number of data elements abstracted from medical charts. Having more of the same data would not have changed the result.

This leads me to conclude that data mining approaches, particularly in combination with statistical methodologies, can be extremely valuable for identifying groups of patients to be studied, but the estimation of treatment effects probably needs to be more “hands on.” This approach would enable the researcher to investigate alternative estimation approaches and results from specification tests to arrive at the “best” estimate of treatment effects. These approaches should help identify real opportunities to identify, evaluate and support opportunities to move from volume to value in health care.

REFERENCES

Looking to publish an article on economic evaluations, policy analysis, or outcomes assessment?
Submit to ISPOR CONNECTIONS today at: http://www.ispor.org/news/ICOnnections.aspx