Does Data Say All You Need to Improve Health Care Performance?

written by Robert McNutt, M.D. | April 20, 2016

Can data lie? In a word, yes. The answer also depends on whom you ask. That goes for all forms of data analysis, including how we evaluate health care. This presents a tough dilemma for providers, patients and other stakeholders at a time when the industry and government are heavily invested in using data to compare provider performance, attach payment to "good" providers and penalties for the others.

The fact is that the health care data "revolution" is in its infancy, and it is not so easy to identify two key facts: what causes outcomes, and what fixes them. Situations where cause and effect are clear are rare, and even rarer are solutions. With that as a given, it's essential to understand how to assess whether data-based conclusions are reliable.

How a Simple Question Can Produce Conflicting Results

Let's take an example of how easily data can tell various stories. In the graphic below, the same set of soccer data was given to 29 different research teams. Each was asked to "determine if referees were more likely to give red cards to dark-skinned players." A red card means the player is ejected from the game. Egregious mistakes lead to red cards, such as spitting on a player or violent actions. It seems a reasonable question, as any bias would devalue the game.

Same Data, Different Conclusions

Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.
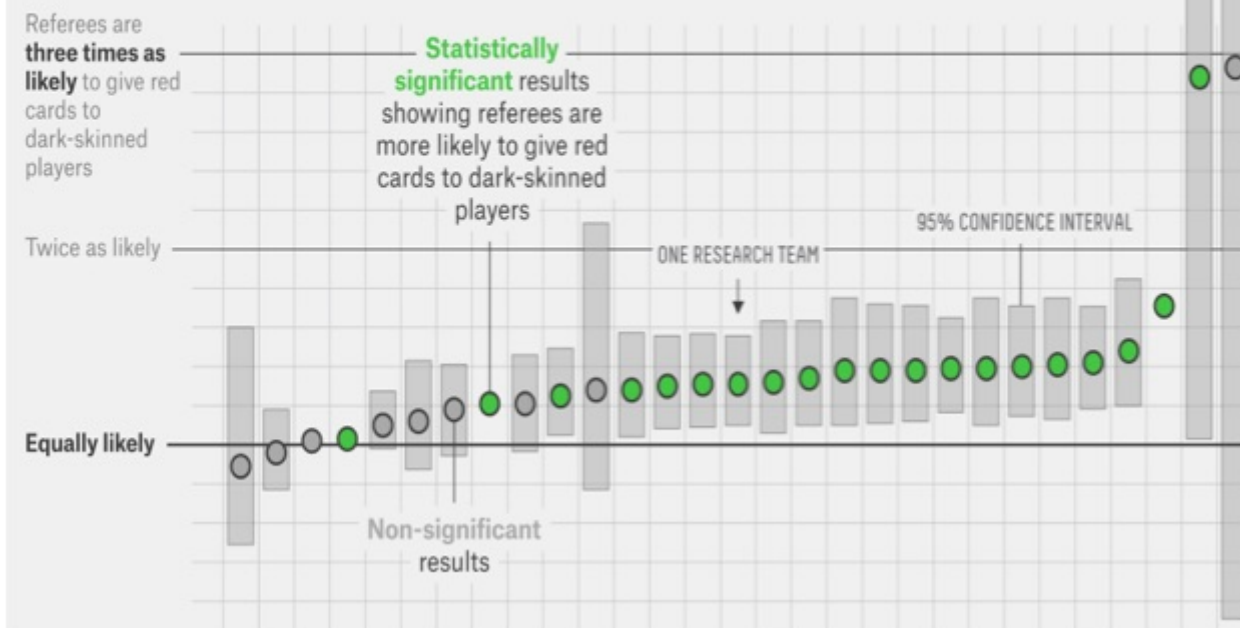
As you can see, using the exact same data, the 29 research teams gave 29 different estimates for the probability that referees give more red cards to dark-skinned players. Responses ranged from an equal likelihood (12 estimates) to more likely (17 estimates). Note, also, how different are the estimates of variations around the mean.

The above figure is stunning and should give us pause. Consider how this applies to "big data." If we are not able to analyze the same dataset uniformly, but, instead, get different estimates based on who is doing the estimating and how they are doing it, we have a science problem of profound implications.

This sort of query is not dissimilar to our medical queries. Your practice and system are always asking questions of data. A question may be, "Will doctors more likely send their patients to subspecialists if their patients have worse A1C levels?" This, like the soccer question, is a prediction question. A health care system that must plan for the number of subspecialists that may be needed would like to know how those referring might act. Our measures of quality of care are asking these sorts of questions on a routine basis. Is my practice better than another, and can my practice's past performance predict this year's performance? These are big-time questions, as revenues ebb and flow with these predictions.

Results Are Heavily Influenced by the Method of Measurement

This report from the 29 research groups points to important lessons. First, we must pay attention to who is doing our data analysis. We need to know how they work and how they use the methods of prediction. Regression analysis is the statistical method used to find associations between a dependent (outcome) variable and independent (predictor) variables. Each of the research teams for the soccer study used regression methods, but they each still derived different answers from the same tool. Hence, who is doing the work is more important than the tool being used. All of the teams were statistical teams, so the variation in results must be due to how they used the tool. Hence, we must consider how regression can "go wrong":

The way data is generated influences the choice of regression methods, and, most often, the generating process is not known. This forces those doing regression to make assumptions. These assumptions can foster misleading results, especially if the effects under consideration are small. The differences in this dataset regarding red cards in soccer were small, as red cards are uncommon in the first place, and we do not know how the data were generated; it could have been a non-representative sample, for example. This may be one reason why the results of the different teams varied; they did not know how the data were generated; hence, their assumptions about how to use the data in the regression models varied.

If researchers allow too much complexity on the independent side of a regression, prediction worsens. This seems counterintuitive, but is noted often in empirical studies. The reasons for this are many, but, most often, the independent or predicting variables being added to the regression are co-dependent; so, they individually add little, while, simultaneously, their addition worsens prediction. The best prediction models use smaller numbers of more strongly associated independent variables. We do not know how the researchers in the soccer study chose and numbered their independent variables, but it could be a source of the difference in their results.

There may be little prior research on which independent variables might matter in the first place. This may be another problem with the red card data set. For example, financial prediction models include well-established prior independent variables like price elasticity when predicting stock market fluctuations. We really don't know why referees may or may not issue red cards. Most of the reasons for issuing a red card are subjective in nature or may be missed during play. In medical care, these high quality independent variables are few and far between, as in soccer. Hence, assumptions or hypotheses about the strength of associations may influence how researchers build their models.

Measurement of the independent and dependent variables may be poor. For example, how was "dark-skinned individual" determined in the red card data set? How reliable is that measure? What are the beliefs of those determining the darkness of the skin when developing the dataset? How do dark skin and the country of origin of the soccer team co-

vary? How variable were the referees in their prior probabilities of giving red-cards in general, irrespective of the "darkness" of the players skin? At the time of the red card, what was the distribution of players' skin colors?

Statisticians attempt to overcome these issues in numerous ways. For example, they may develop "no-change" models, a sort of "placebo like" reference model, to test against their better predicting models. Some statisticians take their two best models and then average these to make a single prediction. However, misleading and overconfident estimates are common, and, as this example of the red card dataset shows, prediction results vary and, hence, make inference impossible and causality nonsensical.

The main points to learn from this soccer study are, again, that just analyzing large datasets will not lead to uniform results. Humbly interpret any observational study and regression analysis; you will never know if you have a useful prediction or not from a single regression study. Second, pick your research teams well and allow them to help in the data collection and planning. Third, stick with a single research team and a single methods approach. It is always best to reduce variation, and this study shows us another way to do that; don't let different ideas address the same sets of data.

Implications for Health Care Performance Improvement

Turning to the more complex scenario of improving outcomes and performance in health care, it's easy to see how we get tangled up in measuring and improving performance. Organizations want to simplify the process of change by using data that are easy to get and employing processes that require a minimum of time and resources. But this approach hides biases and poor methodologies, as well as flawed data.

With more comprehensive data and the maturing of Registry science, the best way to determine how to save money—and the patient—is a well-structured research experiment using Registry data. Previous posts have outlined options for how this should occur. A Clinical Data Registry with stable estimates of outcome measures, data that are pristinely gathered and collated, paired with well-planned experiments, is the way to ensure that what we do actually helps those who are ill to get better.

In science, methods matter at least as much as data, if not more.

Founded in 2002, ICLOPS has pioneered data registry solutions for improving patient health. Our industry experts provide comprehensive Solutions that help you both report and improve your performance. ICLOPS is a CMS Qualified Clinical Data Registry.

Contact ICLOPS for a Discovery Session.

Photo Credit: dawilken