Physician Comparisons Based on Performance Don’t Tell the Right Story

written by Robert McNutt, M.D. | June 8, 2017



Medical decision-making requires a comparison. There is, most often, more than a single option for your care. New tests and treatments are constantly being added to the medical portfolio by scientific inquiry. The only way to advance care, in fact, is by comparing options.

Comparing incites a difficult task, however: the compared option that is best for your disease-related outcome may be worse for your test- or treatment-related outcomes. For example, for men with early stage prostate cancer, surgery may reduce the chance of dying of prostate cancer from 8 to 6 percent over 10 years, but surgery increases, simultaneously, the chance of impotence by 20 to 80 percent. Such a trade-off requires individuals to balance the chance of added benefit and harm to know if the potential value to gain is worth the potential value to lose.

Medical Care Should Inform Individuals About Trade-Offs

Given this vision of decision making, the goal of medical care is to inform individuals of trade-offs and allow them to choose the one test or treatment that is best for them. “Best for them” means that it is likely that two people will choose differently. The goal of the best medical care, then, is to maximize individuals’ variations, rather than the population of individuals’ similarities.

As a result, the best medical care leads to variable decisions made by patients. This means that I may have patients with some diseases who make different choices than yours. And my patients’ choices may lead to differences in their outcomes compared to yours. This is a good thing if patients are making their choices after being informed.

Comparing Physicians Based on Outcomes, Alone, Fails to Account for Patients’ Decisions

But it may not good for comparing physicians’ care if these individual decisions are not taken into account. The present, similarity-based approach to physician comparisons focuses on outcomes, not decisions. For example, someone comes up with an idea about a valuable outcome measure of care (for example, A1C), measures that item in the population of patients you care for, and then compares your measure of that item to a benchmark derived from many other practices measuring the same thing. If your group of patients has a lower average A1C than another group of patients cared for by other physicians, you are rewarded in kind. But when the focus of care is on individuals, patients may choose to forgo side-effects of more aggressive

treatments aimed to lower the A1C, and, hence, have higher A1C levels. Individualized care and population-based care may well be at odds.

I can understand the desire to compare physicians and health systems. However, it is a complex task from a statistical standpoint. There are more than 240 quality-of-care benchmark measures, each with different numbers of patients involved in the measures, with different prevalence rates of base-line performance among compared physicians. In addition, these measures are being taken out of context. Patients' probability estimates for outcomes will vary based on clinical and personal characteristics (the context). But none of these personal variations will be measured, and presently there is no reasonable risk-adjustment for unmeasured and immeasurable patient variations. Now, throw into the mix the complexities of how people will value outcomes in terms of how those outcomes will affect their lives, and you have a pretty sticky statistical problem on your hands.

Aggregate Data Comparing Physicians Can Be Misleading

It's not a new idea that comparing individual physicians will be a difficult task. In a study of 11 million patients, only about 2 to 8 percent of physicians were comparable on even the most reliable of quality-of-care measures. Grouping measures improved the number to about 15 percent. These percentages, while paltry, likely overestimate the subset of physicians who can be compared. Why?

Because important measures of patients and their decision-making practices are presently missing from data sets. For

instance, no patients in this study were asked about their choices and trade-offs, and there were no measures of their unique, clinical and personal nuances (I am ignoring the fact that those physicians being compared may not even be involved in similar types of measures or cases). This means that the estimates from this study are not “risk-adjusted” for those attributes of individuals’ informed choices.

With these personal, co-dependent, confounding measures of patients’ choices added to data sets in the future, the comparable number of physicians might be as low as 0 percent because of the large number of confounding patient factors that are unequally distributed among physician groups. Increasing variation in a data set makes comparison more difficult, as greater numbers of patients will be needed for reliable comparative estimates. If individual physicians are to be compared on these sorts of aggregate data sets, we need far more information about their patients and the process their patients use to make choices.

Accurate Physician Comparisons Need Measures of the Doctor-Patient Decision-Making Process

But, maybe I am wrong; maybe 11 million people are not enough people to adequately test the veracity of comparing individual physicians. Maybe 11 million is too small of a “big” data set, and someday we will have 100s of millions (really big data). This may improve estimates of the numbers of physicians who can be compared, but these population measures, I claim, will still aim at the wrong target. Medical care is practiced behind closed doors between patient and physician. The duo should be spending time discussing the consequences of the patient’s

choice; they should be determining whether the patient wants a better A1C, or what he or she may be willing to give up to lower cholesterol.

Averaged out, large data is at odds with the small data relationship that mirrors a physician-patient bond. As we learn more about comparing groups of individual physicians, I hope we also cultivate, in parallel, the development of a useful measure of the cottage industry of physicians and their patients working to maximize what matters to the patient. For instance, we may need to measure a patient's numeric understanding of the consequences of the choices being made and, concomitantly, how well the physician informed the patient of those consequences.

Science should inform the progress of medical care, and measurement is a key component of science. Measuring physicians and their patients will require more than just data, however. It will involve knowing if patients understand what they are getting into when they encounter the medical care system, and how they, ultimately, direct the system to perform best for them, and not others.

Founded as ICLOPS in 2002, Roji Health Intelligence guides health care systems, providers and patients on the path to better health through Solutions that help providers improve their value and succeed in Risk. Roji Health Intelligence is a CMS Qualified Clinical Data Registry.

Image Credit: Dietmar Becker