Outcome measures need context

Dr. Vinay Prasad, in his commentary on page 146 in this issue of *CCJM*, argues that, to best inform clinical decision-making, interventional and observational studies should measure multiple outcomes whenever possible, including all-cause mortality. He cites examples, such as calcium supplementation for bone health and aspirin for primary cardiovascular prevention, where favorable effects on focused clinical outcomes were not paralleled by favorable effects on overall morbidity. The study was a success, but the patient died.

Reading his commentary got me thinking about the many ways that the results of interventional studies and population data increasingly affect how we practice and teach medicine. Measuring an outcome in the population of interest (study volunteers, patient panels, trainees) is all the rage and is almost always more useful than only tracking interim metrics. True outcome measures are clearly useful when comparing groups and, hopefully, help assess the core reason the study was done.

Yet at the same time that group outcome measures are emphasized for many useful reasons, personalized medicine has a growing appeal: don't let the individual get lost in the group, and pay attention to the outliers as well as the mean.

Positive results from a well-designed, prospective, controlled trial provide confidence that a drug or procedure has efficacy compared with placebo or a known effective comparator. But before recommending a therapy to a specific patient, we need to carefully evaluate whether the likely benefit in an individual patient is worth the clinical and financial cost. The information to make that evaluation doesn't come easily from simply looking at a *P* value in a clinical study. Not only do we need to look at the size of the effect of an efficacious treatment and ask whether our specific patient is comparable to the study participants, but, as Dr. Prasad emphasizes, we must also look closely at the actual outcome measures of the study to see if they match our patient's short- and long-term goals.

How significant is a statistically significant finding if the measured outcome is not the one the patient cares the most about? For example, a recent extremely well-done study that led to US Food and Drug Administration (FDA) approval of branded colchicine for acute gout used the efficacy measure of 50% reduction in pain at 24 hours.[1] But what our patients really want is attack resolution (which usually requires medication in addition to what was used in the trial, increasing the risk of side effects). Proof of concept (a rational dose of colchicine has benefit) was very well demonstrated; that this dosing regimen should be standard of care, I think, remains unsupported.

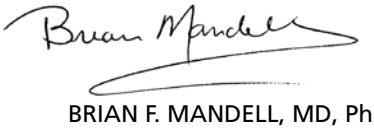
We must also try to assess the long-term relevance (clinical outcome) of results based initially on surrogate markers. For example, not all drugs that increase bone density reduce the long-term fracture rate, and not all drugs that lower the blood glucose level reduce cardiovascular complications of diabetes. This has seemingly become a linchpin concept in the FDA's approach to drug approval, with attendant increases in the cost and time to get drug approval.

We teach that the tools of evidence-based medicine should be routinely and appropriately employed in clinical practice. The premises of evidence-based medicine

doi:10.3949/ccjm.82b.03015

are deeply rooted in clinical studies. But our patients' genetic background, individual preferences, and specific concerns regarding management of their disease and the side effects of medications should also be seriously discussed. We can then jointly define individualized outcome goals in the examination room. These may not exactly match the outcomes chosen by clinical investigators in designing their studies, and the plan may not match the policy of an insurance plan or a "pay-for-performance" metric. I hope that the opportunity for reconciliation of these differences will always be available.

The increasing demand for physicians and health systems to meet specific outcome and performance measures brings up the same concerns that arise when applying the results of a clinical study to a specific patient: will striving to match a group-based outcome be beneficial to the patient in front of us? My major goal as a physician is to care for the individual patient. My patient may not exactly match the population studied to prove that an intervention worked (or didn't), so the data from that study may not fully apply. In the same way, care for all of our patients with the same diagnosis may not fit into the same performance rubric. The same attention that goes into determining appropriately relevant outcome measures for clinical studies needs to go into dictating performance outcome metrics by which physicians and health care systems are measured. They should be patient-centered and, to maintain face validity, somewhat flexible. On any given night, what keeps me awake is not population-based outcomes, but concern over the outcome of the individual patients I saw in clinic that day.



BRIAN F. MANDELL, MD, PhD
Editor in Chief

1. **Terkeltaub RA, Furst DE, Bennett K, Kook KA, Crockett RS, Davis MW.** High versus low dosing of oral colchicine for early acute gout flare: twenty-four-hour outcome of the first multicenter, randomized, double-blind, placebo-controlled, parallel-group, dose-comparison colchicine study. Arthritis Rheum 2010; 62:1060–1068.