Big Data: Making a Difference in Clinical Practice

By Mehdi H. Shishehbor, DO, MPH, PhD

When interventionalists are confronted with the question, "What would change your clinical practice?", one oft-repeated response is, "More clinical trials!" But an understandable concern is that patients and lesions from randomized controlled trials for peripheral artery disease (PAD) don’t always represent the full range of what physicians encounter in daily practice. Strict enrollment criteria often mean that lesions are simpler and patients’ anatomies are less complex than what we see in the clinic. Registries and single-arm studies sometimes incorporate patients with severe PAD and often have larger cohorts, but they rely on site-reported data and can lack the rigor of a randomized clinical trial. Furthermore, design, endpoints, and definitions vary from study to study; few studies report long-term outcomes; and there is a distinct paucity of head-to-head comparisons.

As a result, extracting meaning from any of these studies can be challenging: How do you use these data to make clinical decisions for real-world patients? It’s our responsibility as physicians to understand and apply the data when making treatment decisions, but the heterogeneous nature of the evidence can make this much more challenging.

Hence, we as interventionalists must continue to advocate for studies that will enhance our understanding of each treatment and, ultimately, improve how we care for patients. There are several creative ways that data can be used and reused to help us further understand the role of each treatment in the landscape; but before these uses are possible, the data must be available.

Clinical trial results need to be transparently reported and promptly published in peer-reviewed journals. Many clinical trial results remain unreported and unpublished, although it’s difficult to know exactly what percentage linger in this way considering the lack of a single universal clinical trial registration site. For those studies that are reported, however, there can be considerable lag times between completion of the trial and reporting of the results. Indeed, one group reported in *PLOS* that the median time between primary completion and posting of results was 19 months! And that was only for trials that reported results. This doesn’t even begin to address the deficiencies of the published studies in reporting outcomes by sex, geography, or ethnic background, and in incomplete reporting of adverse events. Certainly, the FDA and other regulatory bodies continue to urge the inclusion and reporting of these components, but to date, the metaphorical carrots and sticks haven’t amounted to much.

Once clinical trial results are published in a peer-reviewed journal, the data are available for additional purposes, like in meta-analyses or the development of evidence-based treatment guidelines. Meta-analyses take advantage of the increased number of patients to identify larger overall trends in treatment. One disadvantage of this big-data approach is that different meta-analyses can produce different results, even when using the same data set. This can happen when different meta-analyses use different methods, for instance, when normalizing patient populations across each of the included studies.

Another big-data approach is to pool data from clinical studies that have already been conducted, which increases sample size and opens the door to performing meaningful subgroup analyses. These studies are different from meta-analyses because they use raw data, and data may be pooled only from trials with uniform endpoints. By analyzing big data this way, you can tackle a wide variety of questions about procedural characteristics, outcomes, and other clinical questions of interest.
Both meta-analyses and pooled analyses can provide substantial amounts of new data, without the enormous investment of time and money required to conduct a new clinical trial with planning, enrollment, and follow-up. These big-data approaches also address one of the ethical concerns of using only clinical trials to answer clinical questions. Clinical trials can be risky and dangerous to a patient. As such, subjecting new patients to therapies or treatments for a question that could be answered with data that already exist is unethical and, frankly, wasteful.\(^4\)

Data sharing has also become a hot topic in the clinical trial space. Slowly but surely, data sets are being made available so conclusions can be independently verified, but also so other groups can address potential hypotheses of interest.\(^5\) However, the privacy concerns here are significant, and sharing of data must be done carefully and cautiously to protect patient confidentiality.\(^6\) Trial sponsors, whether they are hospitals or companies, have the responsibility to think critically about how to incorporate data sharing into future clinical trials. If done right, this leap forward will make an incredible difference in the ability of physicians to improve clinical care by asking questions after a clinical trial is long over.\(^5\)

Big-data analyses are an important complement to randomized controlled trials, single-arm studies, and registries, and they provide greater opportunity to learn more about outcomes in a large and diverse patient population. It behooves interventionalists at large to consider these analyses as an important answer to the question, “What changes clinical practice?” As it turns out, it can be the big data that have been there all along.\(^1\)