

## IS TREATMENT HETEROGENEITY AN ACHILLES' HEEL FOR COMPARATIVE EFFECTIVENESS RESEARCH?

G. Caleb Alexander, MD, MS<sup>1-4</sup>  
Bruce L. Lambert, PhD<sup>4</sup>

1. Department of Medicine, University of Chicago Hospitals, Chicago, Illinois
2. MacLean Center for Clinical Medical Ethics, University of Chicago, Chicago, Illinois
3. Center for Health and the Social Sciences, University of Chicago, Chicago, Illinois
4. Department of Pharmacy Administration, University of Illinois at Chicago College of Pharmacy, Chicago, Illinois

### Funding

Dr. Alexander was supported by the Agency for Healthcare Research and Quality (RO1 HS0189960) and National Heart, Lung and Blood Institute (R01 HL 107345-0). Dr. Lambert was supported by grant number U18HS016973 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. These funding sources had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; and preparation, review, or approval of the manuscript for publication.

### Acknowledgments

Dr. Alexander is a consultant for IMS Health. The authors gratefully acknowledge William Galanter and Gordon Schiff for helpful comments received on previous manuscript drafts.

### Correspondence

G. Caleb Alexander, MD, MS  
The University of Chicago  
5841 S. Maryland, MC 2007  
Chicago, IL 60637  
Phone: 773-834-9177  
Fax: 773-834-2238  
Email: galexand@uchicago.edu

### Length and word count

Text 1160; references 10

The American Recovery and Reinvestment Act (ARRA) included \$1.1 billion to support comparative effectiveness research (CER). Additional investments to expand CER have since been made, including support for the Patient Centered Outcomes Research Institute, a non-profit institute tasked with prioritizing and funding CER and disseminating the results of these efforts to various stakeholders.

Efforts to increase the value of health care through CER have met with both enthusiasm and resistance. Proponents argue that knowing more about the comparative effectiveness of different treatments will empower decision makers to choose the safest and most effective options, ultimately increasing the value of the healthcare dollar. Opponents claim that CER forces a “one-size fits all” approach. Criticism highlights individual differences in treatment response and warns against the perils of overreliance on “average effects.” For example, a review of lessons learned from ALLHAT concluded that the landmark trial demonstrated the value of comparative effectiveness research, but then emphasized that studies based on population averages risk overlooking the needs of individual patients. The same review cautions readers that misapplication of “average effect” data might prevent patients from receiving optimal care (1). Others have argued that CER is “rigged against new drugs” and “designed to eliminate the individual differences that are at the heart of the next generation of personalized medicines” (2). These arguments echo those used in the 1990s to object to standardized treatment guidelines as “cook-book” medicine (3). The looming threat is that the unique needs of individual patients will be ignored as policy-makers intrude upon the clinical encounter.

The concept of heterogeneity of treatment effects is complex and may be underappreciated by patients and providers alike. Different patients may respond differently to the same treatment. This is particularly relevant to the interpretation of evidence from clinical trials that generally report results as differences between the average outcome(s) of interest across treatment

groups. Treatment heterogeneity may arise from observable and unobservable characteristics of patients, their diseases, and the contexts of their care (4). So if the idea of “average effects” is a myth, and no one is average, why isn’t treatment heterogeneity an Achilles’ heel, or an insurmountable flaw, of comparative effectiveness research?

Some argue that CER threatens truly personalized medicine, defined as the delivery of care that is based on genomic tests or other biomarkers to guide treatment choice. Despite the appeal and promise of such therapies, they remain at the margin of clinical practice. Consider pharmacotherapy, where a considerable amount of CER has been done (5). For the foreseeable future, most medications will not be amenable to personalization because they lack at least one of three characteristics required for stratified therapies: (a) differing biologic mechanisms, (b) treatment options with heterogeneous responses, and (c) a clinical biomarker (6). Treatment heterogeneity is not uniform across medicines and therapeutic areas; there are clinically significant individual differences in treatment effects in some areas, small differences in others, and in most areas, we simply don’t know. Similarly, despite a few clinical settings in which biomarkers may be extremely valuable in guiding clinical treatments (e.g., HER2 in breast cancer, EGFR receptor in lung cancer), in most clinical settings no such biomarkers exist.

Critics of CER point out that randomized clinical trials tend to highlight group mean differences in treatment response, and thus blur individual differences between patients within a treatment group. These differences can be substantial and, as with the example of ALLHAT argued above, may lead to an “average effect” being misapplied, thereby preventing a patient from receiving optimal care. To some degree, innovations in the design and conduct of clinical trials may allow for greater adaptability and ultimately may enhance the value of information that the research provides (7). However, without additional information about patients’ sociodemographic or clinical characteristics that would suggest differential responsiveness, the

mean effect derived from a treatment group in a clinical trial *is* the best unbiased predictor of an effect on an individual within that group. Thus, comparative effectiveness research can show, on average, what may be likely to be the most effective treatment. When little is known about a patient's responsiveness, the choice of *initial* therapy may reasonably be guided by such evidence. In the case of pharmacotherapies, such information is particularly valuable when incorporated into the development of formularies and other methods of applying incentives to guide value (8). These decisions must be about what is best, on average, for a population. Of course, customization of treatments in response to treatment failures, side effects, adverse events and other unique features of patients' responsiveness to therapy remains important, but such customization is not, ipso facto, prevented by CER.

The promulgation of CER need not occur in a vacuum, and legitimate interest in the relative effectiveness of a test or treatment among a subpopulation can also be used to help inform comparative effectiveness research, especially as genetic data become increasingly available in clinical databases. However, since CER is a set of methods for comparing treatments, the methods themselves embody no bias for or against personalized medicine and can be deployed either to highlight treatment heterogeneity or to measure average effects.

The challenge for policy-makers is to translate CER results into meaningful policies that assist clinicians and patients in making routine clinical decisions. Indeed, in this respect, one irony regarding arguments that use the promise of personalized medicine to object to CER is how poorly our current health system is structured to achieve even the most rudimentary customization of therapies based on basic patient or clinical characteristics. For example, abundant evidence highlights the need to more reliably customize therapy with respect to basic patient characteristics, such as selecting drug or dose based on weight, kidney or liver function, antibiotic resistance or drug-drug interactions. Indeed, in some clinical settings, innovative

technologies may be so poorly implemented that their very value may be undermined, suggesting that consistent application of known best practices may be a more direct route to quality improvement than reliance on the as-yet-unfulfilled promise of genomic personalization (9,10).

Our goal is not to suggest that treatment heterogeneity doesn't exist, nor to deny its potential clinical significance. Rather, it is to highlight misuse of the concept by those seeking to diminish any leverage that CER may be able to achieve in improving health care value. Emphasis on subgroup differences, heterogeneity of treatment effects, and the importance of patient-centered care are too often cited in support of fallacious arguments that suggest any effort to generate and implement evidence to leverage greater value will lead to inflexible, low-quality care. Personalized medicine celebrates variation, but much of what we have learned about quality improvement in healthcare emphasizes reduction of inappropriate variation, by creating reliable systems to steer clinicians toward evidence-based best practices. Greater use of comparative effectiveness research means greater knowledge of individual treatments and where they stand relative to alternatives. In a marketplace where treatments have to compete based on their real-world safety and effectiveness, the stakes may be higher for those having more to lose.

## REFERENCES

1. Partnership to Improve Care. Available at: <http://www.improvepatientcare.org/> (Accessed on April 8, 2011).
2. Goldberg R. "Medical Quackery". The Washington Times. September 7, 2007.
3. Timmermans S, Mauck A. The promises and pitfalls of evidence-based medicine. *Health Aff.* 2005;24:18-28.
4. Greenfield S, Kravitz R, Duan N, Kaplan SH. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. *Am J Med.* 2007;120:S3-S9.
5. Alexander GC, Stafford RS. Does comparative effectiveness have a comparative edge? *JAMA.* 2009;301:2488-2490.
6. Trusheim MR, Berndt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nature Reviews Drug Discovery.* 2007;6:287-293.
7. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA.* 2007;298:1209-1212.
8. Alexander GC, O'Connor AB, Stafford RS. Enhancing prescription drug innovation and adoption. *Annals of Internal Medicine.* 2011;154:833-837.
9. Woolf SH, Johnson RE. The break-even point: when medical advances are less important than improving the fidelity with which they are delivered. *Ann Fam Med.* 2005;3:545-52.
10. Dai C, Stafford RS, Alexander GC. National trends in Cox-2 inhibitor use since market release: non-selective diffusion of a selectively cost-effective innovation. *Archives of Internal Medicine.* 2005;165:171-177.