



Research Insights

■ To Whom Can We Apply the Results of Comparative Effectiveness Research?

Summary

Through the American Recovery and Reinvestment Act of 2009 and the Patient Protection and Affordable Care Act (ACA) of 2010, federal investment in comparative effectiveness research (CER) increased significantly, including through the creation of the Patient Centered Outcomes Research Institute (PCORI). In grappling with decisions about how to conduct this research and apply its findings to improve health care, the challenge of treatment heterogeneity has become a significant issue for researchers and policymakers. Treatment heterogeneity in CER occurs when a particular treatment leads to different incremental results in different patients when compared to an alternative treatment. While such outcomes can reflect the quality of the research, most treatment heterogeneity reflects differences in the underlying patient population.

For some patients, providers, and other stakeholders, the potential for treatment heterogeneity creates concern about the validity of applying CER results averaged over an entire study population to the care of individual patients. Other experts question the extent to which observed treatment variation is actually the result of

random variation, noting the lack of statistical significance of subgroup analyses in many studies.

Some researchers have suggested addressing the potential risks associated with treatment heterogeneity through “adaptive” study designs that allow investigators to alter their study design to focus on particular subgroups if initial findings suggest that variation in treatment results may exist. Some researchers have advocated “pragmatic” clinical trials that incorporate “real world” conditions in their study designs. Other experts suggest supplementing prospective clinical trials with non-experimental, observational studies.

The potential for treatment heterogeneity poses particular risks for policymakers who seek to use CER results to inform clinical guidelines and insurance coverage decisions. One suggested solution has been to focus disproportionate CER resources on research where there is *a priori* evidence of important clinical variation among patients. Other experts suggest that any policy changes that are based on CER findings should include evaluations of the effects of the change.

Genesis of this Brief:

AcademyHealth convened a panel of experts to share their experiences and perspectives on treatment heterogeneity in comparative effectiveness research during its Annual Research Meeting (ARM) in Seattle, Washington, in June 2011. Anirban Basu, Ph.D., University of Washington, Sheldon Greenfield, M.D., University of California, Irvine, Mark Helfand, M.D., M.P.H., M.S., Department of Veterans Affairs, Portland and Oregon Health and Science University, and Bruce Psaty, M.D., Ph.D., University of Washington participated in the session. David Atkins, M.D., M.P.H., Department of Veterans Affairs, moderated the discussion.

Introduction

The dramatic increase in funding for comparative effectiveness research (CER) and the establishment of the Patient-Centered Outcomes Research Institute (PCORI) has focused a spotlight on the challenge of translating research results into actual health policy and clinical practice. Much of the discussion revolves around large research studies, such as randomized controlled trials, and the applicability of their results to clinical practice. While randomized controlled trials are considered to be the “gold standard” from a methodological standpoint, their use in CER creates tension between the goal of determining what works best for the population at large and the goal of informing clinical decisions for individual patients. In other words, how do individual patients differ from the average patient contemplated in CER? While clinical trials may show that a particular treatment has benefit on average, how should one apply this result to a diverse population?

These issues frame some of the central challenges of translating the results of CER into clinical practice and into policy. Experts in the field debate the importance of variation among patients and the implications for a broader population. Do all variations need to be taken into account to ensure safe treatment and patient-centered care, or is this variation overstated and perhaps exploited to defend against policies that would restrict infinite patient choice? The question of how to weigh this variation, particularly when translating research evidence into actionable policy, has important implications for the future use of CER.

Background on CER

In its 2009 report *Initial National Priorities for Comparative Effectiveness Research*, the Institute of Medicine (IOM) defined comparative effectiveness research (CER) as the “generation and synthesis of evidence” that compares the benefits and harms of alternative ways to improve care delivery or to “prevent, diagnose, treat, and monitor a clinical condition.” The purpose of CER is to provide decision makers such as patients, clinicians, and policymakers with evidence to inform decisions and help to improve health care for both individual patients and populations.¹ In simplified terms, CER is an effort to compare treatments and medical services (e.g. drugs, diagnostic tests, and surgical services) and establish their relative value.

Due to successes in biomedical science, patients and clinicians often have a wealth of options to choose from when making decisions about diagnosis and treatment. Despite these options, patients and clinicians may be unsure of the best options because there is limited or no evidence comparing different therapeutic choices.² This may be particularly true for patients suffering from multiple chronic health conditions. In an era of escalating health care costs coupled with concerns about health care quality, CER seeks to fill this information

gap and represents one avenue to promote high-value, evidence-based care.

In recent years, CER has received a boost from the federal government. The American Recovery and Reinvestment Act of 2009 (ARRA) created the Federal Coordinating Council for Comparative Effectiveness Research to coordinate CER across responsible federal agencies. ARRA contained \$1.1 billion for CER, with the money divided between the Agency for Healthcare Research and Quality (AHRQ), the National Institutes of Health (NIH), and the Office of the Secretary of Health and Human Services (HHS).³ As part of its mission the council laid the foundation for CER by establishing criteria for research priorities and mapping out a strategic framework for CER activity and investments.⁴

The Patient Protection and Affordable Care Act of 2010 (ACA) contained additional investments in CER. The ACA established the Patient-Centered Outcomes Research Institute (PCORI), an independent, non-profit organization that conducts CER to give patients and health care providers the best available evidence to make more informed decisions.⁵ Using a largely stakeholder-driven process, PCORI will lead efforts to prioritize and fund CER, develop methodological standards, and disseminate results to patients and providers. By 2015, total funding for PCORI will reach nearly \$500 million.⁶ In early 2012, PCORI released draft national priorities for research and the research agenda. Proposed national priorities for research include addressing disparities and accelerating patient-centered research.⁷

Issues in Randomized Controlled Trials

Randomized controlled trials are regarded as the most rigorous method for generating evidence on comparative effectiveness, and thus play a major role in an expanded national agenda for CER.⁸ However, experts have raised concerns about the design, conduct, and analysis of large trials. Two issues are especially important. Operationally, randomized controlled trials are large, expensive, and often extremely long in duration. Start-up and subject enrollment is especially time-consuming and the drawn out process means that trials are often inefficient and incompatible with the need for more rapid evidence development. From an analytic standpoint, the traditional medical statistics framework that is used to interpret the results of large trials is often inadequate for CER. The traditional method of dividing results into “significant” and “non-significant” is suited for one-time decisions, such as those for regulatory purposes. This model is not ideal for comparing the effectiveness of different interventions over time.⁹

Another problem inherent in the design of large trials that is particularly relevant for this discussion is that the recruited patient population may differ significantly from the broader general population, and even from the population with the particular disease or condi-

tion in question. Conventional randomized controlled trials strictly limit enrollment to ensure a homogenous study group and also analyze treatment effects individually, rather than in combination with other therapies, which would better represent clinical practice.¹⁰ A treatment shown to be effective in a large trial may have very different effects when applied to the population at large, particularly if there is significant selection bias among trial participants. These issues raise implications for the generalizability of these studies and their findings and their application in CER.

The appropriate role of large trials for guiding clinical and policy decisions in CER is debated among experts in the field. The design of these trials is not optimal to meet the needs of decision makers, and the trials typically do not address the questions that clinicians confront in everyday practice.¹¹ The mixture of harms and benefits that result from any treatment will vary in the general population, and clinicians, when dealing with patients one-on-one, are concerned with individualizing therapy to be most effective. In order to utilize the scientific evidence that results from large trials, it is critical to understand the forces that contribute to variation in results among the study population, and how this affects the application of a treatment to the general population.

Defining Treatment Heterogeneity

The need to individualize therapy for patients is related to the statistical problem of heterogeneity of treatment effects. Also known as *treatment heterogeneity*, this effect is present when the same treatment leads to different incremental results in different patients as compared to an alternative treatment.¹² Treatment heterogeneity has several potential causes. Some of these causes relate to the research studies themselves: e.g., quality of research design and methodology and variations in the setting in which the research is conducted. The major contributor to treatment heterogeneity, however, is individual patient characteristics. These can include variation in the initial severity of disease, patient preferences, vulnerability to side effects, responsiveness to treatment, and the presence of comorbid conditions.¹³ There is no debate regarding the existence of treatment heterogeneity, but rather the debate centers around its relative importance when interpreting the results of CER.

The design of large CER studies frequently presents several problems when trying to interpret results and apply them to a particular population. The overall results of a trial reflect population means, or average effects across patients. Using average effects across different patients can generate misleading results for physicians in clinical practice who care for individual patients—each patient comes with certain characteristics that affect treatment response.¹⁴ There is potential that results from large trials may be ineffective, or, more critically, harmful to particular subgroups of patients. This is one of several concerns among patients, providers, and other stakeholders

regarding the use of CER to guide care and policy—the fear that CER will promote “one-size-fits-all” medicine that does not take recognize subgroups or patients with special needs.¹⁵

On the other hand, measuring heterogeneity of treatment effects is a statistical phenomenon. As such, many experts question whether or not the variation is significant. There are several relevant methodological concerns—first, in any research study there is the possibility that any observed variation is random and therefore does not have significant implications for treatment decisions. Secondly, while there may be non-random variation between different groups of patients, the variation may not be statistically significant and warrant the consideration of individualizing treatment.

These competing forces are challenging for CER and the application of its results to policy. Some experts believe that rules are lacking for distinguishing true and clinically important treatment heterogeneity from chance findings. While some recognize that true heterogeneity should be evident in replication of trials, it is often a challenge to identify significant heterogeneity. The remainder of this brief outlines the varying perspectives on the issue of treatment heterogeneity. This includes:

- Evidence for and against the importance of treatment heterogeneity;
- How to identify factors that are important influences on the balance of benefits and harms for an individual; and
- Whether and how clinical guidance and policies based on these studies needs to be altered in light of possibly subgroup or individual variation.

The Importance of Individual Effects

There are two major perspectives on treatment heterogeneity and its importance when applying the results of CER. On one side of the debate are experts who believe that treatment heterogeneity is real and always present in large trials. Applying average results in the face of this heterogeneity can lead to inefficient, incorrect policy choices. For example, if a large research study shows that a particular drug is, on average, more effective than others, it could lead to concrete policy changes such as a change in coverage policy for an insurer. While the medication is more effective on average, it may not be effective, or may even be harmful, for certain subgroups of patients. If results lead to a one-size-fits-all policy, individual patients may not be able to gain access to the most beneficial treatment.

This issue surfaced in a well-known large trial: the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) trial funded by the National Institutes of Mental Health. This \$42.6 million study compared the effectiveness of one first-generation antipsychotic, perphenazine, and all second-generation antipsychotics available in the United States. The results of CATIE raised questions about how

To Whom Can We Apply the Results of Comparative Effectiveness Research?

CER results should be designed, interpreted, and used to make public policy decisions. The initial results of CATIE reported that most of the second-generation antipsychotics were not superior to perphenazine on the primary outcomes of interest: discontinuation of the medication for any cause and greater time to discontinuation.¹⁶

Cost-effectiveness analysis following the publication of these results showed that treatment with the first-generation antipsychotic, perphenazine, was less costly than treatment with second-generation antipsychotics with no significant differences in effectiveness. While investigators disagreed about the implications that CATIE results should have for access to second-generation drugs, the perspective that spending public funds on second-generation antipsychotics is wasteful was present in the media and adopted by some pharmacy benefit managers.¹⁷ Additionally, the results led to an increase in state Medicaid programs requiring prior authorization for second-generation antipsychotics, and there is a waning commitment among manufacturers for new neuroscience drugs. In fact, a recent review shows that the effect of prior authorizations for second generation antipsychotics have led to increased costs and most likely decreased quality of care.¹⁸

The reliance on average results in CATIE may have overshadowed the heterogeneity present in the study and the important differences between study subjects. This experience has implications for the broader interpretation of CER studies. If CER is to play a large role in transforming the U.S. health care system, it must allow policymakers to alter individual-level choices of treatments—not through restricting access, but through investing in research that will generate individualized information that can guide treatment decisions more precisely than is currently possible.^{19,20}

Experts who believe in the importance of individual effects point to examples such as CATIE and other large trials to show that an overreliance on average results can lead to inefficient and incorrect policy choices. Reliance on average results generates a “winner-take-all” framework in which the full mixture of harms and benefits to real patients is obscured. Experts in the field point to the need for paradigm shift in the design and evaluation of CER studies—one that acknowledges that heterogeneity in treatment effects is the most relevant information in CER. This shift would include recognizing that treatment heterogeneity is the result of both observed and unobserved characteristics. Even when there is a variation in a clear subgroup—such as gender or age—there may be unobserved characteristics which further confound the results.

In another study, researchers found that type 2 diabetes pa-

tients with several comorbidities have very different response to therapy to lower blood sugar (in terms of five-year cardiovascular events) than patients with low levels of comorbidity. This high-comorbidity subgroup was substantial, representing 30 to 50 percent of patients.²¹ This is one example of a patient characteristic with important ramifications for treatment results. In order to truly understand heterogeneity, it is important to consider all possible characteristics and risk factors. Experts believe that this could hopefully lead to more nuanced decision making at the individual patient level, which would take full advantage of the results of CER.

Is Heterogeneity Overstated?

On the other side of the debate over how to interpret and apply the results of CER are experts who believe that while variation certainly exists, strong evidence for significant subgroup variation is often lacking. They are less concerned with unrecognized subgroup effects than with the failure to implement findings from completed trials. From this perspective, it is necessary to have strong, reproducible statistical evidence that treatment responses differ between subgroups.

Additionally, some experts question the ideological underpinnings of an emphasis on treatment heterogeneity. Aligned with treatment heterogeneity are certain characteristics of the health care system in the United States—an enthusiasm for technology and personalized medicine, and a strong emphasis on individualism. These forces can sometimes run counter to implementation of evidence-based guidelines.

Experts do believe, however, that subgroup analysis is potentially helpful in identifying groups that are more likely to experience serious adverse events in response to a treatment. Where they differ from the other side of the debate on subgroup effects is on the issue of treatment efficacy, or effectiveness. They believe that there should not be a presumption of subgroups with regard to treatment efficacy. In the absence of such strong evidence to prove otherwise, then the best estimate for a subgroup may be the mean of the entire population.

Potential Solutions and Policy Implications

Experts on both sides of the debate have suggested solutions for dealing with treatment heterogeneity in large trials. There are alternative methodological approaches for improving the design and interpretation of these studies. One popular solution is to conduct more stratified trials. This is a variation on the traditional trial design where subjects are split into pre-specified groups according to risk models. This type of design allows researchers to better recognize subgroup effects in different populations.

Another solution in the form of a new trial design is the use of

adaptive and pragmatic trials. Adaptive trials use features that can “adapt” in response to information generated during the trial. This flexibility can allow researchers to focus on clinically important subgroups if they present during the study.²² Pragmatic or practical clinical trials represent a way of trying to better reflect “real-world” conditions in study design. Traditional trials typically exclude relevant patient subgroups and also are designed primarily to determine treatment efficacy, following the model that is necessary to gain regulatory approval. Pragmatic trials attempt to be more informative to decision makers and study patients with comorbid conditions in diverse settings. Studied outcomes are patient-centered and reflect what is most important to patients and providers.²³ The pragmatic trial model is more likely to predict results across a wide range of practice settings and eliminates a common obstacle to physician implementation of research findings—lack of applicability.²⁴

There are also potential solutions in non-experimental research designs. Some experts have advocated for the use of well-designed observational studies in CER. One of the crucial elements in conducting a well-designed observational study is correctly identifying subgroups. Observational studies also have the ability to include a wider range of patients and address multiple outcomes better than randomized controlled trials.²⁵

In a policy environment where a large amount of federal resources have been invested in CER, results from studies will have implications for health care. How will results be used to best guide clinical and policy decisions? While the full spectrum of expert opinion on the issue places different levels of importance on treatment heterogeneity, policymakers are tasked with the difficult prospect of translating CER results into actionable policy.

CER creates challenges for writers of clinical guidelines because there may be a tendency to overemphasize the average treatment effects which could be at odds with the best clinical care in certain situations. Experts recognize that there are some areas where average treatment effects may be best. One example is trauma care—in this situation there is no time to individualize care, and studies showing average results can be helpful. On the other hand, providing ongoing treatment to patients with a chronic condition (often in the presence of several other comorbid chronic conditions) is a different endeavor.

One way for policymakers to tackle potential heterogeneity is to use the results from CER studies to identify characteristics of clinical situations which are likely to result in important heterogeneity. Some experts suggest that this would help to focus efforts on studies in clinical areas where heterogeneity is important for applying

the results of CER, as opposed to other areas where it may not be as significant.²⁶

Finally, some experts suggest that if CER is going to be used to directly inform policy—such as changes in coverage policy—then research should include direct evaluations of policy challenges. Such evaluations could include observational studies of policy changes or large-scale social experiments.²⁷ This would help to provide evidence on how policy changes might be effected by treatment heterogeneity. In order to make full use of CER and improve care at the level of the individual patient, the causes and effects of treatment heterogeneity must be understood. Better understanding heterogeneity will help to translate the theoretical benefits of CER to actionable policy to improve the value of health care.

Conclusion

There are significant challenges ahead as the nation seeks to use its investment in CER to improve the quality and value of health care. When translating the results of CER into policy, decision makers will weigh the importance of treatment heterogeneity and its implications for applying study findings. The complexities of these issues provide a good foundation for policymakers to discuss how they want to use the nation’s investment in CER to realize greater value from our health care system and ultimately to improve the health of the nation.

About the Author

Sarah Katz was formerly an associate at AcademyHealth and is currently director of development at Christ House in Washington, DC.

About AcademyHealth

AcademyHealth is a leading national organization serving the fields of health services and policy research and the professionals who produce and use this important work. Together with our members, we offer programs and services that support the development and use of rigorous, relevant and timely evidence to increase the quality, accessibility, and value of health care, to reduce disparities, and to improve health. A trusted broker of information, AcademyHealth brings stakeholders together to address the current and future needs of an evolving health system, inform health policy, and translate evidence into action. For additional publications and resources, visit www.academyhealth.org.

Endnotes

1. IOM, *Initial National Priorities for Comparative Effectiveness Research*, 2009.
2. Federal Coordinating Council for CER, Report to the President and the Congress on Comparative Effectiveness Research. See also: <http://www.hhs.gov/recovery/programs/cer/cerannualrpt.pdf>.
3. U.S. Department of Health and Human Services, Comparative Effectiveness Research Funding. See also: <http://www.hhs.gov/recovery/programs/cer/index.html>.

To Whom Can We Apply the Results of Comparative Effectiveness Research?

4. Federal Coordinating Council Report, 2009.
5. PCORI, About Us, <http://www.pcori.org/about/>.
6. Center for Medical Technology Policy, "Overview of the Patient Centered Outcomes Research Institute," <http://www.cmtpnnet.org/comparative-effectiveness/overview-of-the-patient-centered-outcomes-research-institute>.
7. PCORI. *Draft National Priorities for Research and Research Agenda, Version 1*. 2012. See also: <http://www.pcori.org/assets/PCORI-Draft-National-Priorities-and-Research-Agenda.pdf>.
8. Luce, B.R. et al. "Rethinking Randomized Clinical Trials for Comparative Effectiveness Research: The Need for Transformational Change," *Annals of Internal Medicine*, Vol. 151, No. 3, 2009, pp. 206-210.
9. *Ibid.*
10. Horn, S.D. and Gassaway, J. "Practice Based Evidence: Incorporating Clinical Heterogeneity and Patient-Reported Outcomes for Comparative Effectiveness Research," *Medical Care*, Vol. 48, No. 6, Supplement 1, 2010, pp. S17-S22.
11. Tunis, S.R. et al. "Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy," *JAMA*, Vol. 290, No. 12, 2003, pp. 1624-1632.
12. Kravitz et al., "Evidence-Based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages," *The Milbank Quarterly*, Vol. 82, No. 4, 2004, pp. 661-87.
13. Kaplan, et al. "Who Can Respond to Treatment? Identifying Patient Characteristics Related to Heterogeneity of Treatment Effects," *Medical Care*, Vol. 48, No. 6, Supplement 1, 2010, pp. S9-S16.
14. Kent, D.M. and Hayward, R.A. "Limitations of Applying Summary Results of Clinical Trials to Individual Patients: The Need for Risk Stratification," *JAMA*, Vol. 298, No. 10, 2007, pp. 1209-1212.
15. Docteur, E. and Berenson, R. "How Will Comparative Effectiveness Research Affect the Quality of Health Care?" Urban Institute and Robert Wood Johnson Foundation, February 2010. See also: http://www.urban.org/uploaded-pdf/412040_comparative_effectiveness.pdf.
16. Meltzer, D.O. et al. "Comparative Effectiveness Research for Antipsychotic Medications: How Much is Enough?," *Health Affairs*, Web first, 2009, w794-w808.
17. Meltzer, 2009.
18. Basu A, Meltzer HY. Tying Comparative Effectiveness Information to Decision Making and the Future of CER Designs: The case for Antipsychotic Drugs. *Journal of Comparative Effectiveness Research* 2012; 1(2): 171-180.
19. Basu, A. "Individualization at the Heart of Comparative Effectiveness Research: The Time for i-CER Has Come," *Medical Decision Making*, November/December 2009, pp. N9-N11.
20. Basu A. Economics of individualization in comparative effectiveness research and a basis for patient-centered health care. *Journal of Health Economics* 2011; 30(3): 549-559.
21. Greenfield, S. et al. "Comorbidity Affects the Relationship Between Glycemic Control and Cardiovascular Outcomes in Diabetes," *Annals of Internal Medicine*, Vol. 151, No. 12, 2009, pp. 854-860.
22. Luce, B.R. et al, 2009.
23. *Ibid.*
24. Tunis, S.R. et al, 2003.
25. Greenfield, S. and Platt, R. "Can Observational Studies Approximate RCTs?" *Value in Health*, Vol. 15, 2012, pp. 215-16.
26. Helfand, M. Presentation at AcademyHealth Annual Research Meeting, Seattle, June 14, 2011.
27. Meltzer, D.O. et al., 2009.