Comparative Effectiveness Research Methods

Symposium Overview and Summary

Kathleen N. Lohr, PhD

Background: The Agency for Healthcare Research and Quality, in expanding its efforts to advance methods for comparative effectiveness research (CER), convened a second invitational symposium in June 2009 through its Developing Evidence to Inform Decisions About Effectiveness (DEcIDE) network. A distinguished Planning Committee set the agenda, reviewed more than 70 abstracts, and invited almost 30 speakers; the program emphasized issues of clinical heterogeneity and longitudinal outcomes, including patient-reported measures.

Results: Conference papers covered a varied set of points about 3 major areas of CER: study design and data collection issues; statistical and analytic techniques; and applications to policy and clinical decision-making. Both primary data collection and analyses of databases (including electronic health records and distributed networks) are crucial for the CER agenda. Methods advances enable investigators to pay greater attention to important population subgroups, including persons of low literacy, non-English speaking patients, or the frail elderly. Both established and newer statistical techniques—eg, propensity scoring and various types of modeling, including Bayesian approaches—offer analysts improved ways to address issues such as confounding and measurement bias. Finally, some articles provided guidance for and examples of extending CER into newer realms, such as evaluations of medical devices or surgical procedures, and providing better information for decision-makers, clinicians, and patients.

Key Words: comparative effectiveness, health care, methods, outcomes, research, statistics

(*Med Care* 2010;48: S3–S6)

The Agency for Healthcare Research and Quality (AHRQ), through its Effective Health Care program and the comparative effectiveness research (CER) it supports, particularly through the DEcIDE network, convened a second symposium on CER methods in June 2009. It built on a 2006 AHRQ conference on Emerging Methods in Comparative Effectiveness and Safety,[1] and directly addressed goals set out in the American Recovery and Reinvestment Act of 2009 (ARRA, Public Law 111–5). Harold Sox, in the opening keynote, cited the authoritative definition of CER from the Institute of Medicine[2] and singled out 3 key contributions from a robust, national CER agenda: generating and synthesizing evidence, documenting the effectiveness of alternative interventions, and fostering informed health decisions.[3]

The pressing issues for the 2009 AHRQ symposium concerned new and emerging methods in 2 main domains: ways to enhance the inclusion of clinically heterogeneous populations in CER studies and ways to implement longitudinal investigations that capture longer term health outcomes, including patient-reported outcomes. Cutting across these topics were 3 thematic areas: (1) study design and data collection, (2) statistics and analytic methods, and (3) policy issues and applications. The empirical results and discussions from these symposium papers move the CER field forward in important ways, but many methodological challenges, together with applications to health policy-making, call for continuing work.

OPTIMIZING CLINICAL HETEROGENEITY AND LONGITUDINAL OUTCOMES: THE ROLE OF STUDY DESIGN AND DATA COLLECTION

CER attempts to make meaningful comparisons of medical interventions taking the clinical heterogeneity of patient populations, intervention combinations, and outcomes into account. Desirable study designs and data collection methods should maximize the opportunity of collecting health and patient-reported outcomes (PRO) data on clinically heterogeneous populations while dealing with tradeoffs between internal validity (risk of bias) and external validity (applicability). In one approach toward this goal, researchers used various measures of patient-reported comorbidity and disease burden (for diabetes) to create a composite measure that could represent patients' potential for treatment response; this could then be applied in developing trial designs that can identify important patient subgroups (ie, those with differential potential for response to treatment.)[4] In another paper,

investigators demonstrated how the practice-based evidence study (PBE) methodology measures and controls for heterogeneity of patients, treatments, and outcomes seen in real-world clinical settings and can create a comprehensive set of patient, treatment, and outcome variables by which treatments associated with better outcomes for specific types of patients can be identified.[5] For studying issues such as polypharmacy or other drug-related problems, especially those relating to psychotropic medications, in nursing home residents, the Minimum Data Set offers a wide array of outcomes information and utilization data for observational studies, including those intended to inform regulatory activities such as safety warnings.[6] "Rapid learning health care" is yet another strategy that leverages diverse datasets, health information technology (including electronic PRO [ePRO] capabilities), and sophisticated iterative analyses to create a real-time framework in which clinical studies can evaluate the relative impact of therapeutic approaches on a diverse array of measures; proof of concept, feasibility, and validity studies in patients from academic cancer clinics demonstrated the applicability of these techniques for CER.[7]

Various distributed or federated electronic data systems offer important platforms for CER work that can help overcome drawbacks of traditional observational studies; their critical feature is use of electronic health records. One example is the Distributed Ambulatory Research in Therapeutics Network (DARTNet), which facilitates remote, point-of-care data collection triggered by an electronic prompt for additional information at patients' visits (in this case to nonintegrated primary care clinics providing diabetes care).[8] Other investigators drew important lessons about establishing a distributed research network across multiple data sources. Important conclusions centered on incrementalism (for software development and implementation), security coupled with autonomy (eg, letting data holders query a central site rather than having queries penetrate any system firewalls), and auditing features.[9]

Cluster randomized trials continue to be an attractive approach applicable to many CER topics in both inpatient and ambulatory settings. They are appropriate for networks of hospitals, health plans, or medical practices with centralized administrative and informatics capabilities and for a range of services and settings (as exemplified by a study of methicillin-resistant *Staphylococcus aureus* [MRSA] infection in intensive care units[10]). Another trial-related challenge arises in studying the effectiveness of devices (eg, for artificial hip implantation), especially because such studies are likely to be small and few in number. Comparative assessments of such interventions benefit from using a statistical framework, in this case hierarchical generalized linear models, that can combine diverse information sources from premarket and postmarket settings.[11]

Several papers focused specifically on the challenges of PRO measurement regardless of study design. Measurement bias (dissimilar responses from people who differ on some characteristics but have identical health problems) needs to be addressed in ways that create equivalent subgroups. Multiple-group, multiple-indicator, multiple-cause models can evaluate and correct for measurement bias (as demonstrated for alcohol abuse behaviors) and produce more accurate conclusions across heterogeneous populations.[12] Being of low literacy or non-English speaking (or both) are major barriers to accurate measurement of self-reported behaviors or outcomes across diverse populations in CER projects. A multimedia touchscreen program proved feasible in comparing PRO responses across literacy levels in Spanish-speaking cancer patients and was well accepted by patients; it offers a prototype approach for reaching these especially problematic patient subgroups.[13]

OPTIMIZING CLINICAL HETEROGENEITY AND LONGITUDINAL OUTCOMES: THE ROLE OF STATISTICAL TECHNIQUES AND ANALYTIC MODELS

Using both traditional analytic strategies and innovative statistical methods, specifically focused on clinically heterogeneous populations, is imperative for robust CER work. Of interest are ways to combine multiple data sources, especially for dealing with measurement error, missing data, and heterogeneity of patients or treatment settings. Also critical are techniques, such as simulation or other modeling approaches, that can improve identification of patient subgroups and evaluation of outcomes. Electronic health records are increasingly an important source of CER data, but they pose their own challenges. In some cases, no specific statistical technique, whether traditional or innovative, is an obvious choice to overcome the challenges of heterogeneity or long-term outcomes; in such cases, multiple modeling studies may be needed to determine which account best for myriad confounding factors confronted in different types of databases.

One article illustrated, for CER on the safety of biologics, how using a propensity-score approach enabled the required pooling of data from multiple data sources, allowed for extensive confounder adjustment, and protected patient privacy. This technique proved superior to analyses that used individual covariates and entailed a reasonable tradeoff between strong statistical analyses and flexibility of conducting the needed studies.[14]

For CER on very complex topics, dealing adequately with multiple possible therapies and possible outcomes (both benefits and harms) can be very challenging; available information can be extremely heterogeneous in terms of baseline patient characteristics, settings, and the like. Drug treatment for patients with HIV (including those who have various drug-resistant strains) is a case in point. Mechanistic models of the disease in question can be used to conduct virtual therapeutic trials with the goal of predicting outcomes, some of which are long term and may not be observed within standard trial lengths (eg, deaths or quality-adjusted life expectancy).[15] Another article reported on the so-called competing risks problem, in which investigators cannot easily determine the probability of an outcome in the presence of competing outcomes. Three newer statistical approaches can be used: (1) cause-specific hazard, (2) cumulative incidence function, and (3) event-free survival (EFS). The utility of each technique depends on the specific circumstances of the

study, but all offer improvements over commonly applied estimators of risks and probabilities.[16]

Producing information on expected benefits and harms of alternative therapeutic options is the hallmark of CER, but providing such information tailored to individual patients is far more problematic. Identifying subjects for CER projects from sources such as electronic health databases can be relatively straightforward, but identifying them before clinical diagnosis, so as to understand better their risk factors and other characteristics is challenging. One paper documented how logistic regression or "machine learning techniques" can be used to detect diagnoses (in this case heart failure) well before a clinical diagnosis is recorded, potentially improving selection of members of a broader set of important clinical subgroups for such research.[17]

Finally, controlling for confounding in studies using various kinds of health databases can be especially difficult because of the multiplicity of variables with which to contend. No single statistical technique may serve all purposes, especially in situations of considerable uncertainty about the meaning or importance of different variables or the underlying disease mechanics. Comparative effectiveness investigators may sensibly run statistical models using a variety of specifications and then report all those results, so that users and decision makers can more easily understand how sensitive findings are to model specifications.[18]

COMPARATIVE EFFECTIVENESS RESEARCH METHODS: POLICY AND PRACTICE APPLICATIONS AND IMPLICATIONS

Much comparative effectiveness research to date has focused on pharmaceuticals and on traditional health care delivery systems and insurance schemes. Increasingly, however, policymakers and clinicians are directing questions to other types of health care interventions, patient populations, and insurance programs. Studies involving medical devices and especially implantable devices, for example, can be especially complex, because they call for study designs and implementation strategies quite different from those for drugs. To improve the conduct of comparative effectiveness investigations in this area, one paper outlined a conceptual framework for investigators to use in taking account of problems such as randomization, blinding, and allocation concealment or analytic challenges such as adjusting for confounding and technical features of devices or clinician expertise.[19]

Procedures can pose yet another set of challenges for this type of research. A "coverage with evidence development" project done for the Washington State workers' compensation agency evaluated spinal cord stimulation for chronic back and leg pain after spine surgery (failed back surgery syndrome) and determined that SCS offered no long-term benefits relative to either evaluations in a pain clinic or usual care and was associated with potential harms, leading the program to continue a policy of not covering SCS.[20] Various interested parties in the industry criticized both the study, which went beyond features associated with traditional randomized trials, and the ensuing state policy, leading to a decision that a state technology assessment office should review all relevant evidence about this therapy.

Health care decisionmakers, in seeking to be more evidence-based, are showing increasing interest in the application of Bayesian techniques for synthesizing data in the published literature taken from trials on drugs, devices, and other interventions. The question is whether such information, when subjected to Bayesian hierarchical modeling and other advanced analytic techniques, can yield improved estimates of treatment effects such as mortality or help predict the utility of future trials of the same interventions.[21] Such approaches may accomplish these goals overall, although they may provide only limited help in understanding outcomes for patient subgroups.

An important barrier to widespread use of comparative effectiveness research stems from the need to communicate and disseminate results in ways that work for a wide array of audiences and stakeholders. The Institute for Clinical and Economic Review is working to make such work, including systematic reviews, "fit for purpose"[22] through an evidence rating scheme that combines separate ratings of comparative clinical effectiveness and comparative value. The aim is to present a rating format that supports coverage and payment decisionmaking in both the public and private sectors.

Finally, decisionmaking about and choosing among drugs, devices, or other interventions can be especially challenging when the benefits of the alternative services are relatively similar but the side effects or potential harms are not; the problems can be particularly knotty when clinicians also try to take patient values and preferences into account. Using antidiabetic medications as the example, one paper illustrated how shared decisionmaking, including use of well-designed decision aids, can turn comparative effectiveness findings into information that patients can understand and apply and thus help disseminate evidence into routine, patient-centered practice.[23]

CONCLUSION

The recent increase in national interest in producing and using information on the comparative effectiveness of health care interventions in decisionmaking is striking. Clinical and policymaking communities, and the public at large, seek reliable, timely data that can help ensure safe and effective health services for all. Meeting that need is associated with numerous challenges to the design, conduct, analysis, and reporting of a wide variety of studies—including studies that adequately reflect the diversity of the US population and that provide findings about outcomes important to patients themselves. AHRQ supported this symposium to generate up-to-date ideas about methods that will help investigators, and their target audiences, meet these challenges. In some cases, of course, the issues remain incompletely resolved; future research and methods development will be needed to advance the field of comparative effectiveness research.

REFERENCES

1. Lohr KN. Comparative effectiveness research methods: symposium overview. *Med Care*. 2007;45(suppl 2):S3–S6.
2. Institute of Medicine (IOM). Initial national priorities for comparative effectiveness research. Washington, DC: The National Academies Press; 2009.
3. Sox HC. Defining comparative effectiveness research: the importance of getting it right. *Med Care*. 2010;48(suppl 1):S7–S8.
4. Kaplan SH, Billimek JT, Sorkin DH, et al. Who can respond to treatment? Identifying patient characteristics related to heterogeneity of treatment effects. *Med Care*. 2010;48(suppl 1):S9–S16.
5. Horn SD, Gassaway J. Practice-based evidence: incorporating clinical heterogeneity and patient-reported outcomes for comparative effectiveness research. *Med Care*. 2010;48(suppl 1):S17–S22.
6. Crystal S, Gaboda D, Lucas J, et al. Assessing medication exposures and outcomes in the frail elderly: research challenges in nursing home pharmacotherapy. *Med Care*. 2010;48(suppl 1):S23–S31.
7. Abernethy AP, Ahmad A, Zafar SY, et al. Electronic patient-reported data capture as a foundation of rapid learning cancer care. *Med Care*. 2010;48(suppl 1):S32–S38.
8. Libby AM, Pace W, Anderson HO, et al. Comparative effectiveness research in DARTNet primary care practices: point of care data collec-tion on hypoglycemia and over-the-counter and herbal use among patients diagnosed with diabetes. *Med Care*. 2010;48(suppl 1):S39–S44.
9. Brown J, Brown JS, Holmes JH, et al. Distributed health data network: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care*. 2010;48(suppl 1):S45–S51.
10. Platt R, Takvorian SU, Septimus E, et al. Cluster randomized trials in comparative effectiveness research: randomizing hospitals to test methods for prevention of healthcare-associated infections. *Med Care*. 2010;48(suppl 1):S52–S57.
11. Normand S, Marinac-Dabic D, Sedrakyan A. Rethinking analytical strategies for surveillance of medical devices: the case of hip arthroplasty. *Med Care*. 2010;48(suppl 1):S58–S67.
12. Carle AC. Mitigating systematic measurement error in comparative effectiveness research in heterogeneous populations. *Med Care*. 2010;48(suppl 1):S68–S74.
13. Hahn EA, Du H, Garcia SF, et al. Literacy-fair measurement of health-related quality of life will facilitate comparative effectiveness research in Spanish-speaking cancer outpatients. *Med Care*. 2010;48(suppl 1):S75–S82.
14. Rassen JA, Solomon DH, Curtis JR, et al. Privacy-maintaining propensity score-based pooling of multiple databases applied to a study of biologics. *Med Care*. 2010;48(suppl 1):S83–S89.
15. Roberts MS, Nucifora K, Braithwaite RS. Using mechanistic models to simulate comparative effectiveness trials of therapy and to estimate long-term outcomes in HIV care. *Med Care*. 2010;48(suppl 1):S90–S95.
16. Varadhan R, Weiss CO, Segal JB, et al. Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications. *Med Care*. 2010;48(suppl 1):S96–S105.
17. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, Strategies, and a Comparison of Machine Learning Approaches. *Med Care*. 2010;48(suppl 1):S106–S113.
18. Brookhart MA, Stürmer T, Schneeweiss S. Confounding control in healthcare databases: challenges and potential approaches. *Med Care*. 2010;48(suppl 1):S114–S120.
19. Sedrakyan A, Marinac-Dabic D, Normand ST, et al. A framework for evidence evaluation and methodological issues in implantable device studies. *Med Care*. 2010;48(suppl 1):S121–S128.
20. Turner JA, Hollingworth W, Comstock B, et al. Comparative effective research policy: experiences conducting a coverage with evidence development study of a therapeutic device. *Med Care*. 2010;48(suppl 1):S129–S136.
21. Berry SM, Ishak KJ, Luce BR, et al. Bayesian meta-analyses for comparative effectiveness and informing coverage decisions. *Med Care*. 2010;48(suppl 1):S137–S144.
22. Ollendorf DA, Pearson SD. An integrated evidence rating to frame comparative effectiveness assessments for decision makers. *Med Care*. 2010;48(suppl 1):S145–S152.
23. Shah ND, Mullan RJ, Breslin MA, et al. Translating comparative effectiveness into practice: the case of diabetes medications. *Med Care*. 2010;48(suppl 1):S153–S158.