Emerging Methods in Comparative Effectiveness and Safety

Symposium Overview and Summary

Kathleen N. Lohr, PhD

Background: Interest in new methods for comparative effectiveness, drug and patient safety, and related studies is burgeoning. The advent of Medicare Part D for outpatient prescription drugs has drawn significant attention to the need for efficient ways to monitor the potential benefits and harms of pharmaceuticals. These trends prompted the Effective Health Care program at the Agency for Healthcare Research and Quality and its DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) network to examine innovative approaches for such investigations through an invitational symposium in June 2006.

Results: Conference papers covered numerous points about ways to structure both interventional and database-oriented studies, particularly those concerned with adverse drug events, to avoid bias in those studies, and to apply advanced statistical tools to exploit the information from these studies to their fullest. Of particular importance are: (1) using new types of experimental designs, including cluster randomization, delayed designs, pragmatic trials, and practice-based investigations that incorporate the natural variation of data from routine clinical practice; (2) finding efficient ways to use different types of databases—eg, Department of Veterans Affairs files, Centers for Disease Control and Prevention surveillance files, Medicaid claims data, and state hospital data—for examining initiation, persistence, and adherence, and the benefits and adverse events of pharmaceutical use; and (3) inventing or refining ways to decrease the threats to validity of analyses relying on administrative or other observational data, particularly through propensity scoring, inverse probability weighting, risk adjustment, and direct or indirect methods for synthesizing comparative effectiveness information.

SETTING THE STAGE

At the Agency for Healthcare Research and Quality (AHRQ), trends in evidence-based practice and effective health care, together with Medicare Prescription Drug Improvement and Modernization Act of 2003 (MMA; Pub. L. 108–173), have converged to spotlight issues of drug effectiveness and safety. This focus is equivalent to finding the balance between benefits and harms for health care services, and the Agency's newly established Effective Health Care program is responding by examining new methods for rational, timely, and rigorous assessment of therapeutics. The aims are to maximize the likelihood that beneficial treatments are used and that harmful treatments are not used, to reduce costs and/or improve cost effectiveness, and to provide an explicit, fair, and rational method of resource allocation. Of particular concern, because of the new Medicare Part D program, are health conditions affecting the Medicare population and pharmaceutical interventions. Refining methods that will permit comparative effectiveness and other health services research projects using linked Parts A, B, and D records when and if they become available is especially important.[1]

In recognition of these issues, AHRQ held an invitational conference in Rockville, Maryland, June 19–20, 2006 at which 70 experts gathered to share their current thinking on these urgent matters. The conference papers in this special issue of *Medical Care* reflect these themes, with particular emphasis on the methodological challenges to studying patient safety and comparative effectiveness. Common problems include systematic error (eg, selection bias, exposure misclassification, or outcome misclassification), random error, confounding by clinical conditions, indications for drugs, or use of other therapies, and logistical issues in conducting effectiveness research.[2] For effectiveness and comparative effectiveness work in particular, questions of how best to integrate clinical trials with other study designs are especially knotty.[3] Reporting benefits and harms information from such studies, in a user-friendly and consistent way, is another challenge.[4]

GENERATING EVIDENCE: NEW METHODS TO EVALUATE DRUG SAFETY AND EFFECTIVENESS

New Types of Experimental Studies and New Analytic Techniques

Arguably the broadest strategy for expanding methods for comparative effectiveness and safety analyses is to expand the choice of study designs beyond either traditional randomized controlled trials (RCTs, using experimental designs that may limit generalizability) or purely observational studies. The latter comprise a wide range of nonexperimental designs but also, in this context, various types of database analyses. The emerging techniques of cluster randomized trials, designed-delay and pragmatic trials, and practice-based approaches all offer advantages, but also some drawbacks, for this research.

Cluster randomized trials are especially valuable for evaluating outcomes under conditions of actual use of health care services (rather than through RCTs); randomization occurs by "clusters" that can be defined by physicians, practices, parts or all of different types of facilities, health systems, and geographic regions. Using this approach sets up some statistical issues (eg, power, imbalances in covariates) and epidemiologic questions (eg, low levels of adherence or expected benefits), but it offers considerable advantages for doing comparative effectiveness studies of pharmaceuticals. Especially attractive are applications within health plans with extensive, automated data systems.[5,6] Like many cluster randomization trials, pragmatic trials are done in real-world settings with everyday patients and practices. They may have particular appeal for health program administrators who need to make often far-reaching policy decisions about coverage and benefit issues, including those that turn on data about the effectiveness and safety of pharmaceuticals. "Delayed design" approaches invoke randomized communities and clusters of physician practices (such that some groups of subjects receive an intervention only after a specific amount of time had passed); such "delay" in the randomized context, can provide a useful instrumental variable.[7] Another emerging study design, applied in health systems as a form of participatory action research, can exploit the natural variation in data produced by routine clinical practice to determine what works, for whom, when, and at what cost; this framework may yield results that health systems can act on more rapidly than RCTs can produce.[8]

Databases and Data Analysis Approaches Applicable to Medicare Part D Data

Increasingly, investigators are using databases, such as those from state Medicaid programs, state-wide hospital records, or the Department of Veterans Affairs, to examine initiation, persistence, adherence, and continuity of pharmaceutical use and to study benefits and adverse events of medications. Clarifying the strengths and limitations of such studies is especially crucial in anticipation of the availability of Medicare Parts A, B, and D data.

For decades, researchers have used Medicaid data for research on the initiation, continuation, and patient outcomes of prescription drug therapies (including issues of disparities and use by population groups defined by diagnosis). "Good practices" have accumulated for tasks such as verifying data validity, modeling use of pharmaceuticals over time, and increasing the utility of Medicaid data files, and these will be relevant for studies involving Medicare Part D data.[9] Similarly, innovative techniques can identify new episodes of drug use, apply survival analysis to evaluate persistence and the medication possession ratio to explore adherence within episodes, and include risk scores in analyses to account for measured confounders.[10] Simulation techniques work well for investigating numerous effectiveness, economic, or safety questions in health databases. Examples include examining health insurance benefits, out-of-pocket outlays, and the continuity of pharmaceutical prescription fills and use.[11] Finally, for studying adverse drug events (ADEs), advanced nonlinear methods (eg, hierarchically optimal classification tree analysis) permit creation and validation of ADE surveillance rules that improve on simple, expert-generated rules.[12]

Active Surveillance of Adverse Effects

Much of the work of AHRQ's pharmaceutical portfolio concerns itself with adverse drug (or vaccine) events and, in particular, methods for detecting rare but potentially severe side effects. Such harms may not appear in RCTs done in preparation for licensure applications; they may not even arise to any substantial degree in later studies (eg, head-to-head trials or those done for additional labeling purposes). For that reason, regulators, researchers, and the public are increasingly calling for better ways to monitor pharmaceuticals once they are introduced or in widespread use so that uncommon but serious safety problems can be detected as early as possible.

One option is a specialized surveillance system – maximized sequential probability ratio testing – ie, an enhanced signal detection technique applicable to either continuous or time-period data as they are collected. It permits flexibility in choosing outcomes, selecting controls, and accommodating variations across time or settings; it offers promise for early detection of such adverse events.[13] The National Electronic Injury Surveillance System-Cooperative Adverse Drug Event Surveillance project, a nationally representative surveillance system based on emergency department clinical records maintained at the Centers for Disease Control and Prevention, is another approach for detecting ADEs treated in such settings (ie, outpatient rather than inpatient).[14]

Methods to Control Confounding and Reduce Bias

Among the significant challenges to effectiveness and safety analyses is decreasing the threats to validity that bedevil analyses of administrative or other observational data. Several approaches offer means for achieving this, including propensity scoring, inverse probability weighting (IPW), instrumental variables (IV), and various direct or indirect methods for amassing and synthesizing comparative effectiveness information.

Confounding is a particular threat to many types of comparative effectiveness and safety studies, especially with epidemiologic and nonexperimental studies. To deal with it,

analysts can use IPW and IV techniques. For example, known drawbacks to observational data are sometimes addressed through matching by propensity scores or stratification, but when those methods are insufficient, use of inverse propensity score estimators or IPW estimators permit analysts to deal with problems such as residual confounding or censoring of data.[15] Individual-agent simulations (in which individuals, rather than populations, are the unit of interest) can also be applied to analyze ADEs and their severity and impact; controlling confounding by factors that influence the occurrence of the adverse event, not by factors that influence the outcome, may be the better approach in using such simulations.[16]

IV methods also allow investigators to estimate the effects of treatments (eg, both the effectiveness and safety of pharmaceuticals) in situations involving unobserved confounding. Here the goal is to deal with factors related to the treatment in question but unrelated to the outcome under study. One such variable may be physicians' preferences for using one type of pharmaceutical rather than another, although such methods remain in development and testing stages.[17] Some debate continues about the strengths and limitations of risk adjustment versus IV estimates of treatment effectiveness from observational data, especially when treatment benefits are heterogeneous. Risk-adjustment estimates can provide information on treatment benefits for patients who received treatment similar to information derived from RCTs, but decisionmakers need to take directions of possible bias in account. IV estimates are useful in assessing the effect of treatment rate changes, but decisionmakers need to consider the characteristics of patients whose choices were affected by the IVs in question.[18]

Restricting study populations, as a means of making patients more homogeneous in secondary database analyses, is another effective mechanism for controlling for confounding risk factors that might influence drug use. Increasing restrictions will change rate ratio estimates and reduce bias in studies of treatment effects. Tradeoffs with respect to the applicability of information for policymakers or clinicians, however, may call into question the utility of very narrowly defined restrictions.[19]

Observational studies pose numerous additional hurdles to robust findings and conclusions. They arise from differences in populations, prescribers' behaviors, and confounding by indication or by variables missing from automated databases; yet another complication can develop if patients' characteristics lead to changes in treatments over time. Propensity score methods are increasingly popular as a means to address these, but proper application calls for attention to several points. One is the extensiveness of the set of variables used to derive propensity scores; those based on "reduced" sets may be problematic.[20] Furthermore, these methods may require strong assumptions (eg, that biases in both measured and unmeasured confounders go in the same direction). Innovative techniques, such as subclassification on a longitudinal propensity score, may reduce the multidimensionality of observational data, including treatments changing over time.[21] Nonetheless, pharmacoepidemiologic studies can benefit from using these approaches in examining comparative effectiveness or safety or developing estimates of population use of pharmaceuticals.[22]

Finally, an issue of special concern to AHRQ's Evidence-based Practice Centers (doing MMA-related comparative effectiveness reviews), Centers for Education and Research in Therapeutics (CERTs), and DEcIDE centers involves when and how to use direct evidence (eg, head-to-head trials), indirect evidence (eg, trials of 2 pharmaceuticals each compared only with a placebo or a common active comparator), and combinations of the two. Although unadjusted indirect comparisons are never acceptable, various methods for doing adjusted indirect comparisons are comparably accurate; frequentist and Bayesian methods for comparative effectiveness meta-analyses may have some advantages.[23]

FINAL THOUGHTS

Pharmaceuticals play an increasingly central role in health care, and never more so than with the initiation of the Part D benefit for prescription medications in the Medicare program. With this pivotal place in the medical armamentarium comes growing concerns about the true benefits and risks of pharmaceuticals, both alone and in comparison across drugs. Traditional studies, particularly RCTs, postmarketing surveillance and monitoring, voluntary reporting systems, and the like do not provide the information that the clinical community and the public need to ensure effective, safe health care.

The rising interest in putting to full use all possible sources of data on the impact of pharmaceuticals on patient outcomes is not surprising. The challenges to accomplishing this lie in overcoming possible biases to studies from these sources and using proven methods to extract as much *information* as possible relevant to the decisions that policymakers, practitioners, and patients need to make about their health care options. This symposium provided numerous and compelling ideas about data sources, study designs, and statistical methods to help analysts meet those challenges and, where more work is needed, to set out the critical issues for future research and methods development.

REFERENCES

1. Greenwald LM. Medicare Part D data: major changes on the horizon. *Med Care*. 2007;45(Suppl 2):S9–S12.
2. Strom BL. Methodologic challenges to studying patient safety and comparative effectiveness. *Med Care*. 2007;45(Suppl 2):S13–S15.

3. Atkins D. Creating and synthesizing evidence with decision makers in mind: integrating evidence from clinical trials and other study designs. *Med Care*. 2007;45(Suppl 2):S16–S22.

4. Sedrakyan A, Shih C. Improving depiction of benefits and harms: analyses of studies of well-known therapeutics and review of high-impact medical journals. *Med Care*. 2007;45(Suppl 2):S23–S28.

5. Mazor KM, Sabin JE, Boudreau D, et al. Cluster randomized trials: opportunities and barriers identified by leaders of eight health plans. *Med Care*. 2007;45(Suppl 2):S29–S37.

6. Glynn RJ, Brookhart MA, Stedman M, et al. Design of cluster-randomized trials of quality improvement interventions aimed at medical care providers. *Med Care*. 2007;45(Suppl 2):S38–S43.

7. Maclure M, Carleton B, Schneeweiss S. Designed delays versus rigorous pragmatic trials: lower carat gold standards can produce relevant drug evaluations. *Med Care*. 2007;45(Suppl 2):S44–S49.

8. Horn SD, Gassaway J. Practice-based evidence study design for comparative effectiveness research. *Med Care*. 2007;45(Suppl 2):S50–S57.

9. Crystal S, Akincigil A, Bilder S, et al. Studying prescription drug use and outcomes with Medicaid claims data: strengths, limitations, and strategies. *Med Care*. 2007;45(Suppl 2):S58–S65.

10. Grijalva CG, Chung CP, Arbogast PG, et al. Assessment of adherence to and persistence on disease-modifying antirheumatic drugs (DMARDs) in patients with rheumatoid arthritis. *Med Care*. 2007; 45(Suppl 2):S66–S76.

11. Rupper RW, Bair BD, Sauer BC, et al. Out-of-pocket pharmacy expenditures for veterans under Medicare Part D. *Med Care*. 2007;45(Suppl 2):S77–S80.

12. Nebeker JR, Yarnold PR, Soltysik RC, et al. Developing indicators of inpatient adverse drug events through nonlinear analysis using administrative data. *Med Care*. 2007;45(Suppl 2):S81–S88.

13. Lieu TA, Kulldorff M, Davis RL, et al. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care*. 2007; 45(Suppl 2):S89–S95.

14. Jhung MA, Budnitz DS, Mendelsohn AB. Evaluation and overview of the National Electronic Injury Surveillance System-Cooperative Adverse Drug Event Surveillance project (NEISS-CADES). *Med Care*. 2007;45(Suppl 2):S96–S102.

15. Curtis LH, Hammill BG, Eisenstein EL, et al. Using inverse probability-weighted estimators in comparative effectiveness analyses with observational databases. *Med Care*. 2007;45(Suppl 2):S103–S107.

16. Samore MH, Shen S, Greene T, et al. A simulation-based evaluation of methods to estimate the impact of an adverse event on hospital length of stay? *Med Care*. 2007;45(Suppl 2):S108–S115.

17. Brookhart MA, Rassen JA, Wang PS, et al. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Med Care*. 2007;45(Suppl 2):S116–S122.

18. Brooks JM, Chrischilles EA. Heterogeneity and the interpretation of treatment effect estimates from risk adjustment and instrumental variable methods. *Med Care*. 2007;45(Suppl 2):S123–S130.

19. Schneeweiss S, Patrick AR, Stürmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care*. 2007;45(Suppl 2):S131–S142.

20. Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. *Med Care*. 2007;45(Suppl 2):S143–S148.

21. Segal JB, Griswold M, Achy-Brou A, et al. Using propensity scores subclassification to estimate effects of longitudinal treatments: an example using a new diabetes medication. *Med Care*. 2007;45(Suppl 2):S149–S157.

22. Stürmer T, Glynn RJ, Rothman KJ, et al. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Med Care*. 2007;45(Suppl 2):S158–S165.

23. Vandermeer BW, Buscemi N, Liang Y, et al. Comparison of meta-analytic results of indirect, direct, and combined comparisons of drugs for chronic insomnia in adults: a case study. *Med Care*. 2007;45(Suppl 2):S166–S172.