

P E R S P E C T I V E

Evidence-Based Coverage Decisions? Primum Non Nocere

The use of evidence-based information should follow the instruction for physicians to “first, do no harm.”

by **Newell E. McElwee, S. Yin Ho, Kimberly A. McGuigan, and Mark L. Horn**

ABSTRACT: Drug class reviews are blunt tools for medical decision making. The practice of evidence-based medicine is far more than simply systematic reviews: The patient and doctor are integral. Here we highlight areas of evidence-based coverage decision making where greater balance and transparency could serve to improve the current process, and we recommend elements of a more positive approach that could optimize patient outcomes under resource constraints. [*Health Affairs* 25 (2006): w279–w282 (published online 6 June 2006; 10.1377/hlthaff.25.w279)]

THE PAPER BY Peter Neumann provides a thoughtful and stimulating review of the Drug Effectiveness Review Project (DERP).¹ The DERP has pioneered the application of systematic reviews to drug coverage decision making in the United States, and there are many lessons to be learned from their efforts. Our comments below address specific points in Neumann’s paper on DERP criticism, an important issue related to the DERP that he does not cover, and, finally, lessons and next steps in the evolution of using clinical and economic evidence from systematic reviews as the basis for technology assessment and coverage decision making.

■ **Criticism of the DERP.** Neumann’s section “Criticism of the DERP” is done using a “he said, she said” approach without debating the merits of the criticism or suggesting an ap-

proach addressing them. Concerns about the DERP have been raised by patients, physicians, and the pharmaceutical industry, and we believe that it is critical to address these concerns openly and honestly, to move the field forward. The criticism levied by these groups falls into several categories: program intent (using evidence to improve patient outcomes, decrease costs, or both); methodological issues (relative importance of internal versus external validity, study selection criteria); process issues (mainly transparency and stakeholder involvement); and how DERP reports are used by pharmacy and therapeutics (P&T) committees to make coverage decisions. Although Neumann independently considers program intent and how the reports are used in practice, we believe that these are in fact the same issue.

.....
The authors are all with Pfizer Inc. in New York City. Newell McElwee (Newell.McElwee@pfizer.com) is vice president of U.S. outcomes research in the U.S. Medical Group. S. Yin Ho is senior director of health strategy in the U.S. Planning and Business Development Group. Kimberly McGuigan is director/team leader of the Evidence-based Medicine Center of Excellence, U.S. Outcomes Research, in the U.S. Medical Group. Mark Horn is medical director of the Government Relations Group.

State Medicaid programs, the primary users of DERP reports, have had serious budgetary constraints in the past few years, and many have struggled to maintain essential medical services without raising eligibility requirements for enrollees. Restrictive drug formularies have been one mechanism for Medicaid cost containment, although the data available on the risks to patients of this approach have been scant. One health services researcher has said about restrictive Medicaid formularies:

The relative lack of data on their risks and benefits is cause for concern. It is sobering to realize that if such policies were considered for a clinical study, the possible risks of reduced access to essential medications would likely result in a failure to obtain human subject approval from the Institutional Review Board (IRB). These policies can be viewed as massive experiments on vulnerable populations.²

“The extrapolation of EBM to population-based coverage decisions for new technologies is far removed from its original intent.”

Therefore, we believe that the concern raised about the use of DERP reports as a cost containment tool has face validity and represents a potential conflict of interest.

Here is one way of thinking about potential conflicts of interest in coverage decision making for new technologies. Ideally, one would like to provide coverage for technologies that are effective (or more effective than treatment alternatives, as the case may be) and deny coverage for technologies that are not effective. One of the greatest concerns for payers is covering a technology that is not effective—the statistical equivalent of a Type I error, and a waste of money. To protect against this, payers often view with skepticism any information provided by anyone who would benefit financially from covering a technology. This potential conflict of interest is well recognized in the case of drugs and plays out in the news media almost daily. What is not as well recognized, however, is the converse situation where a technology that is effective is not covered—the statistical equivalent of a Type II error. The stakes are also high in this scenario. At

best, patients don't have access to the technologies they need; at worst, patients are hurt, resulting in excess morbidity and mortality. Our point is not to accuse payers of being insensitive to the needs of patients but rather to suggest that any solution must consider potential biases and conflicts of interest for both covering and not covering new technologies (including drugs). Guidelines are generally lacking for coverage decision making and would be helpful to address these issues.

Neumann also mentions methodological and process criticisms of DERP reports. We will leave it to others to comment on the methodological issues, mainly for the reason stated in a later section of Neumann's paper: The science of systematic reviews is evolving. Many of these methodological issues have been debated

for decades and will likely continue to be debated in the future. However, the notion that the science is evolving is worth raising because it relates to how the DERP reports are being used. We don't know, for example, whether another group of investigators looking at an identical research question would in fact reach the same conclusions or even use the same methods in the systematic review process. Further, the methods themselves are not mature enough to be standardized. Scientists generally understand these limitations, but policy decisionmakers, especially those with a potential conflict of interest, might not.

Neumann does not identify several issues we've experienced in the DERP process. For example, we don't know what happens to the extensive comments we submit on draft reports other than that they don't appear in subsequent revisions. Adjudication of comments is not transparent, and we have received no correspondence on why our comments were not considered. Another issue is that there are both time limitations (two weeks) and page-length limitations for comments that we believe are too restrictive if the overall objective is to “get it right,” especially given that DERP

projects are typically nine months in duration and the reports, hundreds of pages long. These process concerns, on the one hand, are very specific criticisms but, on the other hand, should accordingly be easy to fix.

■ **An important issue not covered.**

Neumann characterizes the DERP experience as an experiment in operationalizing evidence-based medicine (EBM); indeed, we've heard the DERP referred to by others as "EBM." EBM, as defined in David Sackett's most recent textbook for clinicians, requires "the integration of the best research evidence with our clinical expertise and our patient's unique values and circumstances."³ EBM is a tool born of medicine, nurtured in the medical literature, and intended to empower physicians and patients to make individualized and intimate decisions. The extrapolation of EBM to population-based coverage decisions for new technologies is far removed from its original intent. Sackett warned ten years ago that "some fear that evidence-based medicine will be hijacked by purchasers and managers to cut the costs of health care. This would not only be a misuse of evidence based medicine, but suggests a fundamental misunderstanding of its financial consequences."⁴

David Eddy has proposed a new term for the use of evidence in population-based decision making: evidence-based guidelines (EBG).⁵ We do not yet know whether EBM, as traditionally defined, can (or should) meaningfully drive population-based coverage decisions, but we are skeptical. We agree with Eddy that population-based and individual patient decisions are different and that different terms should be used to describe them and different models to inform them. In fact, much of the controversy surrounding the DERP and EBM revolves around the differences between population-based and individually based decision making.

Richard Kravitz and colleagues have elo-

quently described the limitations of making inferences from systematic reviews to individual patients.⁶ Systematic reviews report averages and are therefore applicable to individual patients only to the extent that a particular patient is exactly like the average. Heterogeneity of treatment effects between patients can be, but often isn't, evaluated statistically in systematic reviews. However, statistical tests for heterogeneity assess only interpatient variation in treatment response to a single drug, not inpatient variation in treatment responses for different drugs within the same therapeutic class. These variations in drug effectiveness (responsiveness) and drug safety (vulnerability) can be influenced by multiple mechanisms, including genetic differences among people with different ethnic backgrounds.

Thus, even with perfect DERP methods and a perfect DERP process, the use of these reports for making coverage decisions must be done with caution so that individual patients, especially those who differ from the average, are not hurt.

The DERP model does not, and likely cannot, fully ascertain the impact of the heterogeneity of treatment effect—that is, assess intra- and interpatient variability in vulnerability to adverse events and intra- and interpatient variability in treatment effectiveness. Therefore, DERP reports should not be used to limit therapeutic choices for individual patients without a comprehensive assessment of the potential impact of these decisions on patient outcomes. We believe that payers should be required to measure the impact of formulary restrictions on actual patient outcomes, much the same as the FDA requires so-called Phase IV commitments for pharmaceutical companies. In the future, breakthroughs in genomics might offer objective scientific explanations for the heterogeneity of treatment effects, allowing us to more fully realize the potential for individualized medicine. Meanwhile, until we can accurately predict which patients will re-

“The most appropriate clinical choice for an individual patient is not necessarily found in analyses of populations.”

spond to treatments and which will have adverse events, we must be wary of population-based approaches to coverage decisions.

■ **Lessons from the DERP, and moving forward.** The first step toward a more responsible and responsive evidence-based approach begins with proper governance, transparency, and broad participation. Patients, physicians, specialty societies, payers, health policy and medical experts, and research organizations should all be included. The process must embrace some humility, mindful that evidence and its derivative recommendations continuously evolve. While meta-analyses and systematic reviews are excellent tools for answering certain questions, particularly for populations of patients, there is a gap between population and individualized care; the most appropriate clinical choice for an individual patient is not necessarily found in analyses of populations.

We agree that the process of formulary decision making should be predicated on the best clinical evidence regarding the safety and efficacy of therapeutic alternatives. We also support a process that objectively uses EBM to optimize patient outcomes: one that delivers all relevant information needed by clinicians and patients in a timely manner. We advocate flexibility as these systems evolve; rapidly evolving knowledge and technology demand this flexibility if we are to “first, do no harm.”

.....
The views expressed herein are the authors' and not necessarily the views of their employer, Pfizer Inc.

NOTES

1. P.J. Neumann, “Emerging Lessons from the Drug Effectiveness Review Project,” *Health Affairs* 25 (2006): w262–w271 (published online 6 June 2006; 10.1377/hlthaff.25.w262).
2. S.B. Soumerai, “Benefits and Risks of Increasing Restrictions on Access to Costly Drugs in Medicaid,” *Health Affairs* 23, no. 1 (2004): 135–146.
3. Because David Sackett was a co-editor on the first two editions of *Evidence-based Medicine*, and because he is known by many as the grandfather of EBM, we (and we think others) refer to the third edition as “Sackett’s most recent text,” even though he is not associated with this edition. S.E. Strauss et al., eds., *Evidence-based Medicine: How to Practice and Teach EBM*, 3d ed. (New York: Elsevier, 2005), 1.
4. D.L. Sackett et al., “Evidence Based Medicine: What It Is and What It Isn’t,” *British Medical Journal* 312, no. 7023 (1996): 71–72.
5. D.M. Eddy “Evidence-based Medicine: A Unified Approach,” *Health Affairs* 24, no. 1 (2005): 9–17.
6. R.L. Kravitz, N. Duan, and J. Braslow, “Evidence-based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages,” *Milbank Quarterly* 82, no. 4 (2004): 661–687.