Sever al years ago at a national meeting, when Fairman et al. first presented their attempt to validate the assumptions and findings of 2 frequently cited pharmacoeconomic (PE) models, reaction was mixed. Although some audience members welcomed the concept of model validation, others expressed concern and even irritation with the work. In the lively discussion that followed, one man, who identified himself as the employee of a pharmaceutical manufacturer, expressed his view: “Your work seems to assume that models are supposed to represent real life,” he said. “They aren’t. It doesn’t make sense to validate a model. It’s not real life; it’s only a model.”

In early 2004, Fairman and Motheral were judged to have contributed the most important article to managed care pharmacy published in JMCP in 2003. The 2003 JMCP Award for Excellence was bestowed on these authors for their work in applying actual health plan data to the assumptions used in PE decision-analytic models; specifically, 2 widely cited models of the medical and drug costs associated with combination regimens of antimicrobial and antisecretory drugs to eradicate Helicobacter pylori (H. pylori) infection. Study evidence suggested that the H. pylori treatment models had overstated the cost-effectiveness of brand drug combination regimens such as a proton pump inhibitor with clarithromycin (PPI-C) and had understated the cost-effectiveness of generic drug combinations such as bismuth with metronidazole and tetracycline (BMT). For example, the findings of one model had indicated that costs per successfully treated patient were $1,001 for BMT and $980 for PPI-C. After Fairman and Motheral empirically adjusted that model for actual health plan utilization, costs per successfully treated patient were $852 for BMT and $1,118 for PPI-C.

Two months after the publication of the work by Fairman and Motheral, Cox et al. performed a similar analysis to assess the validity of a PE model of the use of cyclooxygenase-2 (COX-2) inhibitors. Like Fairman and Motheral, Cox et al. found that the proposed PE model was inaccurate. When the model’s assumed use rates for gastroprotective agents (e.g., PPIs, histamine-2 receptor antagonists, misoprostol) were replaced with actual utilization data, the cost per year of life saved for a patient taking a COX-2 inhibitor increased from $18,614 in the original model to $106,192 in the empirically adjusted model.

It is often forgotten that the follow-up analysis of actual pharmacy and medical claims data to test the assumptions in predictive modeling is a fundamental principle in the use of PE modeling results in drug formulary decision making by pharmacy and therapeutics (P&T) committees. Early in 2003, the Task Force on Good Research Practices of the International Society for Pharmacoeconomics and Outcomes Research advised that PE models should only aid decision making and not be represented as “statements of scientific fact,” and should be “continually assessed against data, and models should be revised accordingly.” At the time that the Fairman and Motheral article was published in JMCP in September-October 2003, this research was groundbreaking, with no other similar work (application of actual health plan data to the assumptions contained in a PE model) published in the peer-reviewed literature. Today, these 2 studies by Fairman and Motheral and Cox et al. are the only published research in validation of PE models using actual medical and pharmacy claims; 1 German study published in 2006 used data from cancer registries to test a decision-analytic model for cervical cancer screening in Germany.

In an editorial that accompanied the Fairman and Motheral critical analysis of model assumptions in H. pylori eradication, Hakim opined that all PE models produce results at a given point in time that are expected to be wrong at a future point in time, and that it is more important to focus on the interaction between assumptions and outcomes in a PE model rather than on specific results. To understand that interaction clearly, Hakim argued, model transparency is critical. For the same reason, Hakim advocated use of simple rather than complex models when possible, because simple models facilitate replication and validation. A very different view of modeling was put forth 3 years later by Eddy. He argued that since the point of a model is to get the correct answer, a complex model that is not well understood but that closely simulates real-life processes is preferable to a simpler transparent model that produces a less accurate result. To assess model accuracy, Eddy advocated use of complex mathematical simulations of clinical trials, performed using technology that encompasses object-oriented (“virtual world”) programming and systems of differential equations. Such an approach seems to ignore the problem of markedly flawed input data, which even the most sophisticated mathematical procedure cannot overcome and which was the chief problem uncovered in the 2 analyses performed by Fairman and Motheral and Cox et al. Also overlooked is the difficulty that health plans would encounter in attempting to validate or modify the assumptions and findings of complicated (and proprietary) mathematical models in their own populations.

Healthy skepticism in viewing the results of PE models extends beyond the need for applying real-world outcomes data to the assumptions. Previously, Curtiss described some of the fruit harvested from examining closely the methods used to derive the inputs for PE analyses of self-injectable drugs used for treating multiple sclerosis. For example, he cited 4 “methodological inaccuracies” in the clinical trials of these drugs plus 2 characteristics of the results available from clinical studies that reduce PE analyses of these data to exercises in frustration.
(1) failure to report treatment side effects and adverse events after 2 years of follow-up, and (2) no information reported on the impact of treatment, including side effects, on the quality of life of patients. These omissions have particular importance in the PE analysis of these drugs because patient interviews revealed that treatment side effects were the most common reason for discontinuation of therapy, cited by 52% of patients.11

In the current issue of JMCP, Yuan et al.12 base their PE model primarily on 2 data sources: (1) the Benefits of Entecavir for Hepatitis B Liver Disease (BEHoLD) clinical trial, in which 709 patients with hepatitis B e antigen (HBeAg)-positive chronic hepatitis B (CHB) were assigned to treatment with entecavir (n = 354) or lamivudine (n = 355);13 and (2) the Risk Evaluation of Viral Load Elevation and Associated Liver Disease/Cancer-Hepatitis B Virus (REVEAL–HBV) longitudinal (mean follow-up time 11.4 years) cohort analysis of CHB-infected Taiwanese community residents, of whom 85% were HBeAg-negative.14 Model estimates of total treatment costs associated with entecavir and lamivudine were based on calculations of liver disease (hepatocellular carcinoma and cirrhosis) rates expected over a 10-year time horizon. To calculate these expected rates, Yuan et al. applied post-treatment viral load (HBV DNA) levels from the BEHoLD trial to the REVEAL study's observed liver disease rates, stratified by viral load categories. For example, in the REVEAL study, after statistical adjustment for age, gender, cigarette smoking, and alcohol consumption, adjusted hazard ratios for hepatocellular carcinoma and cirrhosis were 1.2, 2.9, 9.5, and 15.2 for serum HBV DNA levels of 300-9,999, 10,000-99,999, 100,000-999,999, and ≥1 million copies per mL, respectively, compared with the reference category, undetectable viral load (<300 copies per mL). Because 69.1% of entecavir-treated patients in the BEHoLD trial were treated to an undetectable viral load, the model assumed a hazard ratio of 1.0 (baseline risk only) for 69.1% of the hypothetical cohort of patients treated with entecavir in clinical practice. For lamivudine-treated patients, the percentage treated to an undetectable viral load (REVEAL) was 39.8%.

Feld and Ghany recently summarized eloquently the limitations of the clinical trials of the drugs used to treat CHB, including the use of surrogate endpoints that poorly predict long-term remission of CHB.15 Two commentaries in this issue of JMCP describe numerous limitations of the PE model proposed by Yuan et al.16,17 Two of these limitations deserve special attention. First, in the REVEAL study, only 1.4% (n = 8) of HBeAg-positive patients had an undetectable viral load, and no HBeAg-positive patients had a viral load of 300-9,999 copies per mL. Thus, REVEAL's mathematical relationship between viral load level and liver disease rates was based almost entirely on HBeAg-negative patients, yet REVEAL's hazard ratios were applied to a population consisting entirely of HBeAg-positive patients. Second, Yuan et al.'s model assumes that a patient who is treated to an undetectable viral load level in a clinical trial with a minimum viral load level of 3 million copies per mL for study entry (BEHoLD) has the same liver disease risk as a community resident in a cancer screening program who enters a cohort study with an undetectable viral load (REVEAL). In future years, will validations of the Yuan et al. model demonstrate that it accurately predicted liver disease rates for patients treated with entecavir and lamivudine or will it, like the models studied by Fairman and Mothaler and by Cox et al., be largely refuted by empirical evidence?

While we await the answer to that question, information about actual histologic improvement for BEHoLD subjects may represent the best "rubber-meets-the-road" assessment of the model that we currently have, especially since the BEHoLD study authors describe histologic improvement as the study's "primary efficacy end point." The BEHoLD trial report indicates a higher rate of "histologic improvement" with entecavir (72%; 226/314) than with lamivudine (62%; 195/314). However, that comparison included in the denominator all study patients with baseline biopsy specimens and did not exclude patients with missing follow-up biopsies. It is only in a footnote to their primary data table for histologic improvement that the authors identify the counts for "adequate pairs of biopsy specimens" as 292 for entecavir and 269 for lamivudine. Using those denominators, the percentages of cases with histologic improvement are 77% for entecavir (226/292) and 72% for lamivudine (195/269), yielding a nonsignificant P value of 0.204 for the comparison (Fisher's exact test); that is, there was no significant difference in the primary efficacy endpoint of histologic improvement for entecavir versus lamivudine when only the cases with evaluable liver biopsies are compared.

Given all of these sources of uncertainty created by the application of epidemiological data (REVEAL) to clinical trial data (BEHoLD) and the apparent lack of difference between entecavir and lamivudine in the primary endpoint of histologic improvement in BEHoLD, the model proposed by Yuan et al. may reveal less about the cost-effective treatment of CHB than it does about the current state of the art in PE modeling. If we are concerned about the accuracy of answers produced by PE models, we are not alone. Skepticism about model results is common, with some observers arguing that models are so obtuse and difficult to scrutinize that authors should be required to submit electronic versions as part of peer review to enable verification by journal editors and reviewers.18-22 Inadequacies in economic models submitted as part of product dossiers using AMCP Format for Formulary Submissions have been reported.23,24 In a review of PE analyses contained in dossiers submitted to one health plan between 2002 and 2005, only 43% contained sensitivity analysis, 38% described the study perspective, 20% described assumptions clearly, and 18% described caveats to study conclusions.

For some observers, lack of confidence in PE model results is amplified by pharmaceutical manufacturer sponsorship. For example, in 1994 the New England Journal of Medicine established
a conflict of interest policy for PE models similar to its existing policy for review articles. PE models would be “excluded from consideration” if any author had “a personal financial conflict of interest” in model results. Such policies, although well intentioned, may be insufficient to address the problem of potentially biased or incorrect model assumptions when results of a sponsored clinical trial are applied to a model developed and published at a later date. Only a careful reader of the BEHoLD clinical trial report, published in the New England Journal of Medicine in 2006, will find the following statement contained well within the methods section and NOT in the disclosures at the end of the article: “The sponsor collected the data, monitored the conduct of the study, performed the statistical analyses, and coordinated the writing of the manuscript with all authors.”

The credibility gap in PE modeling is longstanding among pharmacy and medical directors involved in P&T decision making and is unlikely to be resolved by increasingly complex mathematical solutions, no matter how theoretically accurate or sophisticated. Authors of PE models might do better by focusing on basic face validity and on careful documentation and support of model assumptions. Otherwise, readers of PE models might just decide that the man who raised the point years ago was right—a model is not real life; it’s only a model.

**Authors**

KATHLEEN A. FAIRMAN, MA, is associate editor and senior methodology reviewer of the Journal of Managed Care Pharmacy; FREDERIC R. CURTISS, PhD, RPh, CEBS, is editor-in-chief of the Journal of Managed Care Pharmacy.

AUTHOR CORRESPONDENCE: Kathleen A. Fairman, MA, Kathleen Fairman LTD, PO Box 31278, Phoenix, AZ 85046, Tel.: 602 867.1343, E-mail: kathleenfairman@qvnet.net

**REFERENCES**


2. Academy of Managed Care Pharmacy. JMCP Award for Excellence. Available at: www.amcp.org/amcp.ark?p=F0382B76.


