Cost-effectiveness analysis (CEA) can be a powerful analytic tool for assessing the value of health care interventions when used in conjunction with efficacy, safety, and other supporting data in an evidence-based decision-making environment. CEA is commonly defined in terms of the comparison of costs, expressed in monetary units, with outcomes that may be expressed in a variety of ways. One of the most common forms of CEA compares costs in monetary units with outcomes quantified in nonmonetary units (e.g., cancer avoided, death avoided, or successfully treated patient). Cost-utility analysis (CUA) is a form of CEA that compares costs in monetary units with outcomes quantified as a multidimensional measure of effectiveness (e.g., utilities that are used to estimate quality-adjusted life-years [QALYs]). Cost-benefit analysis (CBA), another form of CEA that is used less frequently, compares costs and benefits (i.e., outcomes) both of which are quantified in monetary units. Overall, there has been an increasing trend in the use of CEA (CEA, CUA, and CBA) to inform decision making. This trend can be implicitly measured by the frequency of published CEAs over time. Using the Tufts Medical Center Cost-Effectiveness Analysis Registry, Neumann et al. (2009) assessed CUAs published between 1976 and 2006.\(^1\) The authors observed a substantial increase in the number of published CUAs, from less than 50 per year for the period 1976-1996 to approximately 225 per year in 2005 and 2006.

In the United States, there is conflicting evidence as to the acceptance of CEA. Continuing with the Tufts registry analysis, Neumann et al. observed that published CUAs developed from the U.S. perspective decreased from 65.6% for the period 1976-90 to 38.5% in 2006.\(^1\) Determining the factors leading to this decline is not a straightforward task. Furthermore, there is an underlying assumption that publication in a peer-reviewed journal is a proxy for acceptance of CEA. One explanation for the decline in published CUAs may be the use of QALYs in CUA, which have yet to gain significant traction in the United States. The decline in published CUAs may also be attributable to overarching concerns about the quality, reliability, and relevance of CUA (and by extension CEAs), as well as the long-held perceptions that commercially sponsored CEAs are inherently biased.\(^2,3,4\) Published CEA guidelines aim to alleviate some of these concerns by implementing standards by which to conduct the research.\(^5,6,7\) In light of the declining number of published CUAs in the United States, there is a trend towards the targeted use of CEA in formulary decision making. For example, economic evaluations (e.g., CEA/CUA and budget impact modeling) have been incorporated into the Academy of Managed Care Pharmacy’s (AMCP) Format for Formulary Submissions Version 3.0 (AMCP Format) to support reimbursement and/or formulary placement of an intervention.\(^8\) Through dossiers prepared in the AMCP Format, manufacturers have the opportunity to summarize the evidence and rationale supporting all choice(s) in a manner that is clear, transparent, and evaluable by decision makers. There has also been an increasing trend in the use of CEA/CUA by managed care organizations (e.g., WellPoint and Premera) to inform the formulary listing process and determine tier status of new interventions.\(^9,10\)

Ultimately, the acceptance of CEA in the United States and its ability to inform decision making is dependent on the methods employed, relationships between parameter inputs and model outputs, appropriate interpretation of model outputs (including a firm understanding of the uncertainty surrounding the analysis), and transparent dissemination of the CEA. Becker and Dembek address the relationship between parameter inputs and model outputs and the subsequent interpretation of CEA results in the June 2011 issue of the *Journal of Managed Care Pharmacy* (JMCP).\(^11\) They raise an interesting and vexing point of discussion with respect to how parameter inputs are selected in CEAs. This issue is subject to a further level of complexity when considering multiple sclerosis (MS) and disease-modifying drugs (DMDs). While the randomized clinical trials (RCTs) of the DMDs for the treatment of MS have demonstrated that treatment has a favorable impact on at least 1 (often several) of the short-term outcomes typically used to assess efficacy in MS trials, the RCTs had numerous methodological shortcomings\(^12,13,14\) subjecting even the best CEA practice guidelines to the ultimate stress test.

**Background – A Case Study in DMDs for the Treatment of MS**

Becker and Dembek wrote in response to a previously published economic model (Goldberg model) assessing the cost-effectiveness of DMDs as first-line treatment of relapsing-remitting MS (RRMS).\(^15\) The model compared 5 treatment strategies: no DMD treatment, intramuscular (IM) interferon...
beta-1a (IFNβ-1a; Avonex), subcutaneous (SC) IFNβ-1a (Rebif), IFNβ-1b (Betaseron), and glatiramer acetate (GA; Copaxone). The model was developed from the U.S. payer perspective by adopting a 2-year time horizon reflecting the duration of the DMD RCTs, as well as the U.S. payer preference for shorter time horizons when conducting economic evaluations (however, the appropriateness of a 2-year economic analysis for a chronic condition like MS is debatable). Parameter inputs were sourced from the pivotal, randomized, placebo-controlled clinical trials for the DMDs; published literature; standard sources (e.g., Red Book for drug acquisition costs and Centers for Medicare and Medicaid Services fee schedule for physician services costing); and expert opinion. Results were reported in terms of clinical outcomes (relapses and disease progression); costs and incremental cost-effectiveness ratios (ICERs) were reported as a cost per relapse avoided for each individual DMD treatment versus no DMD treatment.

In order to assess the relationship between parameter inputs and model outputs, Becker and Dembek (Becker-Dembek model) replicated the Goldberg model and should be applauded for developing an accurate replicate (with a reported absolute difference between analyses of less than 1%). The Becker-Dembek model was subsequently used to examine the parameter inputs for relapse rates associated with IM IFNβ-1a, where early termination of the RCT resulted in 2 populations that could serve as the basis for estimation of efficacy inputs (relapses and disease progression, the latter of which is not addressed) in an economic model (Table 1): the intent-to-treat (ITT) cohort (relative reduction in annualized relapse rates vs. placebo [‘relative reduction in relapse rates’]: 18.2%) and a 2-year completer cohort (relative reduction in relapse rates: 32.2%). Relapse rates based on the ITT cohort were used for the base-case analysis in the Goldberg model. Becker and Dembek argue that this cohort, where only 57.1% of subjects completed at least 2 years of therapy, does not provide a like-with-like comparison when evaluated with the ITT cohorts of the other DMD therapies, where 86%-92% of subjects completed at least 2 years of therapy. Rather, the authors propose that the relapse rates estimated from the 2-year completer cohort would provide a better comparison and resulted in significantly different ICERs when used (Goldberg model: $141,721 per relapse avoided; Becker-Dembek model: $77,980 per relapse avoided). This is not an unexpected finding given the magnitude of the difference in the relative reduction in relapse rates based on the 2 cohorts (ITT: 18.2% vs. 2-year completer: 32.2%). However, the replacement of a single parameter input without careful consideration of the methodology, assumptions, and robustness of the economic model raises questions regarding the transparency of the analysis and may perpetuate the perception of distrust of CEA, especially among U.S. decision makers.

### Challenges to Incorporating Transparency in MS Modeling

Perhaps the foremost challenge to CEA/CUA in MS is how best to address the considerable differences in the DMD RCTs. In particular, it has been noted that there were differences in the proportion of subjects completing 2 years of therapy, baseline disease severity as measured by the Expanded Disability Status Scale (EDSS), relapse and disease progression criteria (e.g., definitions, clinical activity prior to enrollment), and primary and secondary endpoints (e.g., relapses, disease progression), as well as “methodological inadequacies” such as inadequate allocation concealment, incomplete reporting of proportion of treatment dropouts, and failure to calculate treatment effects in ITT analyses. Thus, the heterogeneity across the RCTs may also violate the like-with-like comparison among the DMDs and further supports the notion that all parameter inputs and assumptions should be carefully explored and tested in sensitivity analyses to assess the impact on the interpretation of model results.

When evaluating an economic analysis, decision makers should acknowledge that the model developed for the CEA/ CUA is by no means perfect; it is oftentimes a simplistic view of a very complex real world. In the case of MS, the heterogeneous aspects of the RCTs may be difficult to fully account for in an economic model, resulting in a number of simplifying assumptions, each of which has implications on the interpretation of model results. The focus on a single parameter input, which results in a more favorable outcome, is somewhat disingenuous considering that the article by Becker and Dembek was sponsored by the manufacturer of IM IFNβ-1a, the arm of the model being changed, and may contribute to the perception that...
TABLE 2  Summary of IM IFNβ-1a Relapse Rate Data Used in Previous Multiple Sclerosis Economic Models

<table>
<thead>
<tr>
<th>Reference</th>
<th>Population (Country)</th>
<th>Time Horizon (Cost Year)</th>
<th>Active Treatment Comparators</th>
<th>Effectiveness Measure</th>
<th>Relative Reduction in Relapse Rates (IM IFNβ-1a vs. Placebo)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Becker (2011)11</td>
<td>RRMS (US)</td>
<td>2 years (2008)</td>
<td>1M/SC IFNβ-1a IFNβ-1b GA</td>
<td>Relapses avoided</td>
<td>0.322</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.242–0.403</td>
<td>Based on RCT (Jacobs 1996)16</td>
</tr>
<tr>
<td>Nuijten (2010)29</td>
<td>RRMS (Germany)</td>
<td>4 years (2008)</td>
<td>1M/SC IFNβ-1a IFNβ-1b GA</td>
<td>Relapses avoided</td>
<td>0.183</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.137–0.229</td>
<td>Based on RCT (Jacobs 1996)18</td>
</tr>
<tr>
<td>Chao (2009)25</td>
<td>RRMS/SPMS (US)</td>
<td>2 years (2008)</td>
<td>Natalizumab 1M/SC IFNβ-1a IFNβ-1b GA</td>
<td>Relapses avoided</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.24–0.40</td>
<td>Based on RCT (Jacobs 1996)18</td>
</tr>
<tr>
<td>Goldberg (2009)15</td>
<td>RRMS (US)</td>
<td>2 years (2008)</td>
<td>1M/SC IFNβ-1a IFNβ-1b GA</td>
<td>Relapses avoided</td>
<td>0.183</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.137 – 0.229</td>
<td>Based on RCT (Jacobs 1996)18</td>
</tr>
<tr>
<td>Tappenden (2006)12,</td>
<td>RRMS/SPMS, 51 year olds (US)</td>
<td>Lifetime (2005)</td>
<td>1M/SC IFNβ-1a IFNβ-1b GA</td>
<td>QALYs</td>
<td>0.18</td>
<td>Based on RCT (Jacobs 1996)18</td>
</tr>
<tr>
<td>Tappenden (2009)26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Gani (2008)27</td>
<td>HARRMS, 36 year olds (UK)</td>
<td>30 years (2006)</td>
<td>Natalizumab IFNβ-1b GA</td>
<td>QALYs</td>
<td>0.19</td>
<td>Based on Cochrane review of IFNβs (Rice 2003)16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(IFNβs combined) 0.26–0.11</td>
<td></td>
</tr>
<tr>
<td>Bell (2007)17</td>
<td>RRMS (US)</td>
<td>Lifetime (2005)</td>
<td>1M/SC IFNβ-1a IFNβ-1b GA</td>
<td>Time spent in EDSS 0-0.5 Relapse free QALYs</td>
<td>0.27</td>
<td>Based on published reviews (Filippini 2003, 14 Simpson 2002,28 Khan 200221)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.088–0.452</td>
<td></td>
</tr>
<tr>
<td>Prosser (2004)19</td>
<td>Non-PPMS, 30 year olds (US)</td>
<td>10 years (1999)</td>
<td>IM IFNβ-1a IFNβ-1b GA</td>
<td>QALYs</td>
<td>0.1914</td>
<td>Based on RCTs ( Jacobs 1996,18 Rudick 1997 30) and adjusted for patients who discontinued during the trial</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1340–0.2489</td>
<td></td>
</tr>
<tr>
<td>Chikoti (2003)31</td>
<td>RRMS/SPMS, 30 year olds (UK)</td>
<td>20 years (2001)</td>
<td>1M/SC IFNβ-1a IFNβ-1b GA</td>
<td>QALYs</td>
<td>N/A</td>
<td>Based on RCT (Jacobs 1996)38 – data not provided in publication (source cited as published trial and commercial in-confidence trial data)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Nuijten (2002)32</td>
<td>RRMS, 30-year-old females (UK)</td>
<td>Lifetime (1998)</td>
<td>IFNβ-1b</td>
<td>QALYs</td>
<td>N/A</td>
<td>IM IFNβ-1a not assessed in model</td>
</tr>
</tbody>
</table>

EDSS = Expanded Disability Status Scale; GA = glatiramer acetate; HARRMS = highly active relapsing remitting multiple sclerosis; IFNβ = interferon beta; IFNβ-1b = interferon beta-1b; IMSC IFNβ-1a = intramuscular/subcutaneous interferon beta-1a; N/A = not available; PPMS = primary-progressive multiple sclerosis; QALYs = quality-adjusted life-years; RCT = randomized controlled trial; RRMS/SPMS = relapsing-remitting multiple sclerosis, SPMS = secondary-progressive multiple sclerosis, UK = United Kingdom, US = United States.

commercially sponsored CEAs/CUAs are inherently biased. As noted by Becker and Dembek, there are at least 9 published models evaluating the cost-effectiveness of the DMDs for MS (Table 2). At a minimum, these analyses provide additional perspective on how IM IFNβ-1a relapse rates were addressed and, when used more broadly, offer an opportunity for decision makers to compare across the studies (CEA and CUA) to assess if the results are consistent or not and what the key drivers of each individual analysis may be. In the context of Becker and Dembek’s article, a summary and discussion of these models would have provided a foundation for readers to build upon when evaluating how best to use the IM IFNβ-1a efficacy data in the context of an economic analysis.

Becker and Dembek state that the objective of their research was to highlight the impact that varying input parameters can have on the output of a CEA. The results of the Becker-Dembek model demonstrated that the Goldberg model and the subsequent interpretation of results were sensitive to changes in a single parameter input—the relative reduction in relapse rates associated with IM IFNβ-1a compared with placebo. Again, this finding was not unexpected, as Goldberg et al. (2009) observed that their model was sensitive to changes in the relapse rate inputs, with the model being more sensitive to using an absolute rate reduction rather than relative rate reduction, but that the overall rank order of results did not change.15 In as much as Goldberg et al. discussed the
differences among the DMD RCTs while outlining the methodology employed in developing the model (referencing the IM IFNβ-1a RCT completion rate of 57.1%), there was no discussion of the 2 IM IFNβ-1a cohorts (ITT vs. 2-year completer cohort) and the implications on parameter input selection and the subsequent interpretation of model results using one set of inputs versus the other. A question arises as to what extent Goldberg et al. should have explored the differing estimates of relapse rates, tested the rates in the model, and discussed the implications on the interpretation of results. While sensitivity analyses were conducted in the Goldberg model (parameter inputs were varied by ±25% including the clinical data on relapse rates), the high end of the range tested in sensitivity analyses for IM IFNβ-1a relapse rates was still below the 2-year completer rate from the IM IFNβ-1a clinical trial (relative reduction in relapse rates: 22.9% vs. 32.2%).

Determining the Appropriate Parameter Inputs—What Does the Evidence Tell Us?

A key question remains: What is the most appropriate parameter input for IM IFNβ-1a relapse rates that should be considered in an economic evaluation of the DMDs in MS? Analysis of the IM IFNβ-1a RCT resulted in 2 estimates of the relative reduction in relapse rates, both of which were associated with significant improvements in favor of IM IFNβ-1a (ITT: 18.3%, \(P = 0.04\); 2-year completer: 32.2%, \(P = 0.002\)).18 When compared in the same model, the choice of parameter estimate results in substantially different ICERs for the cost-effectiveness of IM IFNβ-1a versus no DMD treatment (Goldberg model: $141,721 per relapse avoided; Becker-Dembek model: $77,980 per relapse avoided). Commercial in-confidence data for IM IFNβ-1a suggests that the equivalent ITT estimate for efficacy (relapses and disease progression) would result in considerably less favorable ICERs than the central estimate of cost-effectiveness based upon the public domain data.12

In systematic reviews, the IFNβs were found, on the whole, to reduce the frequency of relapses compared with placebo (relative relapse rate [95% confidence interval]: first year of treatment = 0.73, [0.54-0.99]).15,16 However, sensitivity analysis assessing the effect of discontinuation (e.g., withdrawal or loss to follow-up) suggested that the IFNβ treatment effect on relapses compared with placebo was inconclusive when assuming discontinuations progressed in their disease (relative relapse rate [95% confidence interval] = 1.11, [0.73-1.68]), an outcome largely driven by the IM IFNβ-1a RCT population (relative relapse rate [95% confidence interval] = 1.78, [1.46-2.17]).14

In one of the first prospective, open-label, comparative effectiveness studies of the DMDs, 156 RRMS patients who had an EDSS score between 0 and 4 elected to either receive no treatment or treatment with IM IFNβ-1a, IFNβ-1b, or GA for 18 months.10,20 After 12 months of treatment, patients in the untreated group had a relapse rate of 0.97, whereas patients in the IM IFNβ-1a, IFNβ-1b, and GA groups had relapse rates of 0.85 (relative reduction in relapse rates vs. untreated group = 12%), 0.61 (relative reduction in relapse rates vs. untreated group = 37%), and 0.62 (relative reduction in relapse rates vs. untreated group = 36%), respectively.

In as much as the open-label design and limitations of such require cautious interpretation of results (Kahn 2002), head-to-head RCTs of the IFNβs support the findings observed in the open-label comparative effectiveness study. The Independent Comparison of Interferon (INCOMIN) study randomized patients to IM IFNβ-1a or IFNβ-1b, and a statistically significant reduction in annualized relapse rates in favor of IFNβ-1b was observed (relative relapse rate: 0.71, \(P = 0.03\)).22 The Evidence for Interferon Doseresponse: European-North American Comparative Efficacy (EVIDENCE) study randomized patients to IM IFNβ-1a or SC IFNβ-1a, and a statistically significant reduction in relapse rates in favor of SC IFNβ-1a was observed at 24 weeks (relative relapse rate: 0.73; \(P = 0.022\)) and at 48 weeks (relative relapse rate: 0.84; \(P < 0.001\)).23

Accounting for the INCOMIN and EVIDENCE data, the Agency for Healthcare Research and Quality (AHRQ) economic evaluation estimated relative relapse rates for each intervention that were very similar to those estimates derived solely from the RCTs (IM IFNβ-1a: 0.82 vs. 0.83; SC IFNβ-1a: 0.68 vs. 0.68; IFNβ-1b: 0.70 vs. 0.66); however, relative hazard ratios for disease progression were substantially different (IM IFNβ-1a: 0.58 vs. 0.79; SC IFNβ-1a: 0.60 vs. 0.70; IFNβ-1b: 0.71 vs. 0.52).12 When using the mixed treatment estimates in the CUA, the AHRQ authors noted the following: "The incorporation of evidence from the head-to-head trials resulted in some marked differences...the effectiveness of [IM IFNβ-1a] was reduced considerably. While [IM IFNβ-1a] appears to have a comparatively favorable cost-effectiveness profile when based upon the placebo-controlled trial data alone, this appears considerably less favorable when cost-effectiveness estimates were based also upon evidence from head-to-head trials."12 From an economic evaluation perspective, the IM IFNβ-1a relapse rates estimated from the ITT population (relative reduction in relapse rate versus placebo: 18.3%) appear to be best supported by the evidence and would seem reasonable to include in a base-case CEA, with the evidence of more favorable reductions in relapse rates associated with IM IFNβ-1a reserved for sensitivity analyses.

Moving Beyond Modeling and Towards Real-World Effectiveness

While CEA can be a powerful analytic tool (if used appropriately), have we moved beyond CEA based on the RCTs for assessing the value of DMDs for the management of RRMS? The majority of CEA, especially those detailed in Table 2, rely on efficacy data reported in the RCTs and have consistently resulted in rather large ICERs, with the majority of CUA ICERs exceeding the commonly accepted threshold of $50,000 per QALY gained. Furthermore, the models have all been
sensitive to changes in a core set of parameter inputs/assumptions, including the following: time horizon, efficacy (disease progression or relapses), drug acquisition costs, and discount rate. Has the time come to augment our decision making with real-world effectiveness data and subsequently evaluate the economic implications based on these observations? The first DMD for the treatment of RRMS was approved in 1993 (IFNβ-1b) with 2 DMDs approved in 1996 (IM IFNβ-1a and GA) and 2 others approved in 2002 (SC IFNβ-1a) and 2004 (natalizumab). Thus, there is approximately 18 years worth of data for the DMDs from which to assess outcomes and inform decision making (yes, the author acknowledges the challenges to identification/analysis of the data, and the limitations associated with such data are substantial).

Conclusions

Overall, Becker and Dembek should be applauded for developing an accurate replicate of the Goldberg CEA model. The ability to replicate such a model from the original publication speaks to the relative transparency in which Goldberg et al. present their methodology. Becker and Dembek should also be applauded for challenging readers to critically evaluate all aspects of the analysis being presented. They accomplish this by demonstrating via example the implications of changing a single parameter input in an economic evaluation on the subsequent interpretation of results. Furthermore, Becker and Dembek highlight the need to pressure test our analyses with consideration of the appropriate range of parameter input values to test in sensitivity analyses.

However, the vehicle chosen by Becker and Dembek to deliver their message may impede its successful delivery. Economic modeling of the DMDs for MS is fraught with challenges that place even the best CEA practice guidelines to the ultimate stress test. Foremost are the numerous methodological shortcomings of the RCTs, which require a number of simplifying assumptions for modeling purposes. Next, the selection of a single parameter input to test without careful consideration of the methodology, assumptions, and robustness of the model does not increase the transparency of the analysis. This is further confounded by the fact that the article by Becker and Dembek was sponsored by the manufacturer of IM IFNβ-1a, the arm of the model being changed, thus, potentially contributing to the perception of distrust of CEAs among U.S. decision makers. Finally, in light of the evidence for the relative efficacy of IM IFNβ-1a, the justification of using parameter inputs estimated from the 2-year completers cohort as the base-case analysis is lacking. It is imperative that those responsible for developing economic evaluations such as CEAs and CUAs carefully design the analysis with transparency in mind and that the analyses are tested in a rigorous manner to ensure the robustness of results and the subsequent interpretation of said results.

The author reports no financial or other conflicts of interest related to the subject of this commentary. Bell is employed by GlaxoSmithKline, which is engaged in the development of drugs to treat multiple sclerosis.

REFERENCES


