

2-1-2017

Multiple Sclerosis and the Comparative Value Disease Modifying Therapy Report of the Institute for Clinical and Economic Review (ICER)

Paul C. Langley

University of Minnesota, langley@maimonresearch.com

Follow this and additional works at: <http://pubs.lib.umn.edu/innovations>

Recommended Citation

Langley PC. Multiple Sclerosis and the Comparative Value Disease Modifying Therapy Report of the Institute for Clinical and Economic Review (ICER). *Inov Pharm*. 2017;8(1): Article 12. <http://pubs.lib.umn.edu/innovations/vol8/iss1/12>

INNOVATIONS in pharmacy is published by the University of Minnesota Libraries Publishing.

Multiple Sclerosis and the Comparative Value Disease Modifying Therapy Report of the Institute for Clinical and Economic Review (ICER)

Paul C Langley, PhD

Adjunct Professor, University of Minnesota

Abstract

On January 26, 2017, the Institute for Clinical and Economic Review (ICER) posted its final report on treatments for multiple sclerosis (MS) with disease modifying therapies (DMTs). The objective was to provide a modeled assessment of the effectiveness and value of the various DMTs against each other and supportive care. The model considered both relapsing-remitting and primary-progressive MS with hypothetical patient cohorts tracked from therapy initiation to death in a lifetime cost-utility framework. Recommendations were made for possible DMT package price discounts given benchmark willingness-to-pay cost per QALY thresholds. The purpose of this commentary is to assess this modeled analysis from the criteria of normal science: are the claims presented for the competing DMTs credible, evaluable and replicable? The review concludes that the ICER model does not meet required standards. The claims made for comparative effectiveness and value are non-evaluable. They are immune to failure. The review concludes that if models are to contribute to improving our understanding of the effectiveness and costs of DMTs then they should be evaluable in the short-term to allow feedback to formulary committees in a meaningful timeframe.

Keywords: Multiple sclerosis, ICER, simulations, credibility, replication, daclizumab, alemtuzumab, natalizumab

Introduction

Over the past 12 months a number of papers and commentaries have been published, principally in *INNOVATIONS in Pharmacy*, which have questioned the methodological basis of much of what we know as cost-effectiveness or cost-utility modeling^{1 2}. These include critiques of established formulary submission guidelines by agencies including the National Institute for Health and Care Excellence (NICE) in the UK, the Pharmaceutical Management Agency (PHARMAC) guidelines in New Zealand, the Health Information and Quality Authority (HQIA) guidelines for Ireland, the European Network for Health Technology Assessment (EUnetHTA), Recommendations for Health Economic Evaluations and the Academy of Managed Care Pharmacy (AMCP) Format for Formulary Submissions in the US^{3 4 5 6 7}. At the same time, commentaries have addressed the role of lifetime quality adjusted life year (QALY) claims in the US and the limitations implicit in such claims^{8 9}. Critiques have also been presented on the policy of journal editors to publish lifetime cost per QALY claims and reviews of claims in selected disease areas^{10 11 12 13 14}. Attention has been given to the likely disruptive impact of next generation sequencing

on cost effectiveness modeling claims and the release of the second version of the Minnesota proposed Guidelines for Formulary Evaluation^{15 16}. Most recently, commentaries have been published on the latest versions of guidelines for the Netherlands, Australia and France^{17 18 19}. In each case, the focus of the guideline has been on the construction of imaginary worlds. Indeed, in the case of Australia the requirement by the Pharmacy Benefits Advisory Committee (PBAC) for constructing non-evaluable cost-effectiveness claims has been in place for over 25 years. None of the guidelines addressed the issue of claims evaluation and replication.

The theme that underpins these commentaries and papers is that health technology assessments, as characterized by current standards and textbooks, have missed the point in developing models to support cost-outcomes claims²⁰. Rather than focusing on modeled claims that are credible, evaluable and replicable in a timeframe that provides rapid feedback to formulary committees, the focus has been on the development of modeled claims that have no chance of ever being evaluated. The view seems to be that the models are merely to 'inform' decision makers rather than supporting hypothesis testing or systematic observation. It is also unclear as to what 'informed' actually means and how decision makers are expected to factor this information into healthcare choices.

The purpose of this commentary is to evaluate the latest report from the Institute for Clinical and Economic Review (ICER) on the effectiveness and value of disease modifying therapies (DMTs) in multiple sclerosis (MS)²¹. Following from an earlier assessment of the ICER review of Entresto in heart failure, the

Corresponding author: Paul C Langley, PhD

Adjunct Professor, College of Pharmacy

University of Minnesota

Director, Maimon Research LLC

5061 North Apache Hills Trail, Tucson, AZ 85750

Tel: 520-577-0436; Email: langley@maimonresearch.com

Web: www.maimonresearch.net

case is made that in persisting with a reference case methodology to support cost-effectiveness claims the review is of limited value⁹. While the intent may be to 'inform' decision makers, the claims made for the cost-effectiveness of the competing therapies are not only non-evaluable but also non-replicable. They should not be considered as guides to formulary decisions and pricing. Rather, the claims are constructed from a simulation represented by a lifetime cost-per-QALY framework that takes its inspiration from the NICE reference case. Unfortunately, in constructing simulations, regardless of appeals that a chosen model is 'sufficient' to represent a future reality, any model can be challenged and competing models constructed to give quite disparate claims. The only basis for differentiating models is through assessing their claims or, more simply, hypothesis testing. As it stands, if cost-utility or associated lifetime claims are non-evaluable then we don't know whether they are right or even if they are wrong. Indeed, we will never know. The claims are immune to failure.

The ICER Blueprint

If the primary objective in health technology assessments of competing pharmaceutical products is to inform formulary committees through the construction of long-term or lifetime models or simulations, then the ready acceptance of the NICE reference case by groups such as ICER, even with minor modifications, is quite understandable. Indeed, there could be any number of competing reference cases mandated by various health assessment agencies. The nature of the reference case perspective is made quite clear in the latest draft (Version 4) of the Canadian guidelines: *Economic evaluations are designed to inform decisions. As such they are distinct from conventional research activities, which are designed to test hypotheses*²². Presumably it is up to the decision makers as to which reference or other model they believe best 'informs' them in the absence of any attempt to evaluate the merits of competing claims empirically.

A reference case model has been used extensively by NICE over the last 18 years to support reviews of products within disease areas as well as single product reviews²³. It has been endorsed by professional groups, notably the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) in both recommended standards for model building as well as by extensive educational programs^{24 25}. At the same time journal editors have been prepared to accept claims based on reference case models with, literally, thousands of claims presented through various papers yet with no apparent attempts to evaluate those claims or even to recognize that the claims are non-evaluable.

In the case of ICER, the procedures for model development and review follow closely the NICE reference case. The key elements

of the model, which in the case of MS was developed by the University of Washington (UW) School of Pharmacy Modeling Group, are:

- defining the decision problem, through a stakeholder scoping exercise
- expressing health effects as QALYs within a modeled cost-utility framework with a fully incremental analysis for the technologies being compared
- focusing on direct medical costs
- ensuring that the model time horizon is long enough to reflect all important differences in costs and outcomes between the technologies being compared
- application of threshold willingness-to-pay values to support recommendations for product access

The ICER reference case (following NICE) requires the decision model time horizon to be sufficiently long 'to reflect all important differences in costs and outcomes between the technologies being considered'. For NICE and the ICER, as many technologies have an impact over the lifetime of a patient, a lifetime horizon is usually considered appropriate. This is the case for MS. Although not explicitly stated by ICER, NICE makes the point that 'analyses that limit the time horizon to periods shorter than the expected impact of treatment do not usually provide the best estimates of benefits and costs'.

Application of NICE reference case standards, in the case of chronic disease interventions, results in a model or simulation that attempts to mimic the natural progression of the disease and the impact of competing interventions, over the patient's lifetime or similar long-term time horizon. Stages of the disease progression are captured by, for example, a Markov process which tracks the hypothetical cohort of patients through the disease stages. Each health state is defined in terms of associated utilities and costs. This results in scenario driven claims expressed in cost-per-QALY terms. By application of a willingness-to-pay threshold cost-per-QALY by NICE, products are judged acceptable, rejected or accepted after agreement on a discounted price. ICER, following NICE, also applies willingness-to-pay thresholds to support recommendations for product acceptance as well as discounting. Also, following NICE, there appears to be no interest to place the model in the public domain so that alternative simulations can be run. This has been proposed by the US Academy of Managed Care Pharmacy (AMCP)²⁶.

The ICER multiple sclerosis model

The analytic framework for the MS modeling tracked a hypothetical cohort of patients 18 years of age and over. The framework tracked the patient cohort from initial diagnosis with an index DMT to possible relapse and therapy switching to

a second line DMT and, ultimately, supportive care. The interventions for the review of RRMS, by route of administration, were:

- injectable (daclizumab, glatiramer acetate, interferon β – 1a, peginterferon β – 1a, interferon β – 1b)
- oral (dimethyl fumarate, fingolimod, teriflunomide)
- infused (alemtuzumab, natalizumab, ocrelizumab, rituximab)

All these agents were compared across route of administration using both head-to-head and placebo controlled trial data, as well as to placebo and to each other through a network meta-analysis. For the PPMS review, the interventions included ocrelizumab and rituximab, with primary comparator best supportive care.

The principal outcome for the ICER comparative value assessment of the various DMTs was their projected lifetime cost-effectiveness in treating RRMS and PPMS. The modeled intervention constructed outcomes were: (i) drug costs; (ii) adverse event costs; (iii) total costs; (iv) quality adjusted life years; (v) life years; and (vi) relapses. Two Markov models were developed, one for relapsing remitting MS (RRMS) the other for progressive MS (PPMS) and with health states based on the Expanded Disability Status Scale (EDSS). As RRMS patients may progress to secondary progressive MS (SPMS) over their lifetime, these health states were included in the RRMS model.

A natural history of MS transition matrix described the patient experience of MS with an estimate of the relative risk for each DMT therapy to derive transition probabilities between the health states of the model. The RRMS model consisted of 20 health states: EDSS 0-9, EDSS 1 – 9 and death. Patients were distributed at baseline between the 10 RRMS health states according to their initial diagnosis of MS, transitioning between health states on a one-year cycle until death. First line treatment was initiated with one of the DMTs and after discontinuation progressed to second-line treatment and then supportive care. Patients could progress to death or relapse from any health state. As it was not feasible to model every possible combination of DMTs over time, an average second-line approach was applied that aggregated second-line treatment over all patients. The PPMS model consisted of 10 health states (EDSS 1- 9 including death). As in the case of the RRMS model, treatment naive patients were distributed among the health states, transitioned over a one-year cycle over a lifetime horizon. Discontinuation leads to best supportive care. Annual discontinuation probabilities were estimated from 25 of the 27 studies included in the base case network meta-analysis for EDSS progression. The annual clinical study discontinuation probabilities ranged from 1.9% (alemtuzumab 10 mg) to 15.5%

(teriflunomide 7/14 mg) with 10 of the 14 DMTs estimated to have an annual discontinuation rate under 7%. Mortality rates were based on US life tables and adjusted for MS-specific EDSS state mortality multiplier.

Annual utility values for each EDSS state were based on previously published estimates and scored for the EQ-5D instrument. Each mild/moderate relapse event was associated with a one-cycle disutility of 0.091 and a severe event with a one-cycle disutility of 0.302 (on a range of 0 – 1, where 1 is perfect health). Annual utility rates were estimated for RRMS and SPMS/PPMS separately. In the former case they ranged from 0.875 (EDSS = 0) to -0.1701 (EDSS = 9). In the latter they ranged from 0.791 (EDSS = 1) to -0.2138 (EDSS = 9). A negative sign indicates a utility score that is worse than death.

Both models were subject to a one-way sensitivity analysis to capture the key drivers in the model. The models were also subject to a probabilistic sensitivity analysis jointly varying the model parameters. Possible 95% range estimates for each model outcome were then generated. Model validation involved feedback from manufacturers on the choice of model assumptions, comparing the model results with other independently developed models and sensitivity analyses.

Model results were presented for (i) total discounted costs; (ii) relapses; (iii) life years; and (iv) QALYs over the lifetime time horizon. The base case lifetime model costs for RRMS were approximately \$333,000 for supportive care, ranging from \$572,000 for alemtuzumab to \$1.5 million for daclizumab. Number of projected relapses was 16.4 for supportive care, ranging from 10.8 for alemtuzumab to 15.6 for interferon β -1a to. Discounted life expectancy from age of DMT initiation was 21.4 years for supportive care, ranging from 21.9 years for teriflunomide 7mg and interferon β -1a 22cg to 23.1 years for alemtuzumab. Discounted QALYs were 5.7 for supportive care, with a range from 7.8 for teriflunomide 7mg to 12.6 for alemtuzumab.

The modeled results for PPMS patients yielded \$264,300 for supportive care with 15.6 life years and 2.7 QALYs, compared to 16.1 years and 3.3 QALYs for ocrelizumab (not yet approved by the FDA).

All scenarios were subject to one way and probabilistic sensitivity analyses (PSAs) to capture the impact of parameter uncertainty on costs and health outcomes. The PSAs involved jointly varying all model parameters over 5,000 simulations and calculating 95% credible range estimates for each model outcome. This yielded a probability estimate that each DMT was notionally cost-effective at the \$150,000 per QALY

threshold compared to supportive care and glatiramer acetate 20mg.

ICER Conclusions

In the case of supportive care, the threshold probabilities ranged from 0.0% to 99.9%, with the majority of the 15 DMTs reporting $\leq 5\%$. In the case of glatiramer acetate 20 mg the probabilities ranged from 1.5% to 99.3% and with 7 of the DMTs $\leq 10\%$.

Estimates were also provided of the prices for each drug that would achieve cost-effectiveness thresholds ranging from \$50,000 to \$150,000 per QALY. Because of the accrued costs of second-line drugs and other care it was not possible to calculate a threshold price for ten of the DMTs and \$50,000 and for two of them at a \$100,000 threshold. The wholesale acquisition cost could not be raised enough to reach these so-called cost-effectiveness thresholds.

On the basis of these modeled results, ICER presents value-based benchmark prices. This is the price range that would achieve cost-effectiveness ratios between \$100,000 and \$150,000 per QALY gained. Apart from alemtuzumab, whose price could be increased substantially, for most DMTs the discounts required to achieve both willingness to pay thresholds would be greater than the current average estimated discounted WAC except for glatiramer acetate 20mg and interferon β -1b 250 mcg. Average discounts from WAC for DMTs to achieve a \$100,000 threshold ranged from 25% to 75% and for a \$150,000 threshold a range from 94% to 58%. Finally, as an example, daclizumab the newest approved agent generates an estimate of \$223,000 per QALY gained, which implies a discount from WAC to reach a willingness to pay threshold of 35% to 58%.

Discussion

It is not clear how these model results, with the various scenarios and comparisons to establish the impact of uncertainty, would be evaluated by a formulary committee. None of the various outcomes are presented in terms that are evaluable either as stand-alone or comparative claims. Although it is extremely doubtful if a manufacturer would underwrite a submission to ICER to give their 'version' of a reference case claim, there is no doubting the popularity of lifetime cost-utility models in the literature and the willingness of manufacturers to support such models. With the obvious caveat that we seldom find manufacturer sponsored claims that are not supportive of their product, there is no 'standard' reference model. Unsurprisingly, perhaps, a significant number of published sponsored models point to the cost per-QALY estimate coming in under the \$50,000 threshold with the attendant claim that they 'meet' modeling standards. Again, it

is difficult to see how a further investment in non-evaluable modeled claims is expected to influence formulary committees. Of course, formulary committee members could always attend a pre-conference ISPOR workshop on lifetime models to appreciate better the role of non-evaluable lifetime modeled claims in formulary decision making. Not surprisingly perhaps, the extent to which health technology assessments, including modeled claims, are factored into health system formulary decisions in the US seems limited. The awareness of comparative effectiveness research does not translate to application, with concerns expressed over the need for timely information to support decisions^{27 28 29}.

Models and Assumptions

It is important to remember that, in reviewing the ICER model, it is not the question of whether the model structure or the assumptions appear 'reasonable, but the acceptance of reference case approach itself. Better or more 'realistic' assumptions do not lead to a better projection; rather, different assumptions lead to different projections. Comparisons are irrelevant (e.g., comparing the final version of the ICER MS model with earlier versions)³⁰. The fundamental issue is one of whether or not the model generates evaluable claims. If it does not, then it fails the standards of normal science. The exercise may be interesting, but it does nothing to further our understanding of the impact of competing DMTs in MS patients. Even so, it is of interest to consider a number of the key assumptions driving the model. This does not mean the model can be 'improved' but that these issues should be addressed if, as one might hope, the focus could shift to developing models with evaluable claims.

Lifetime Markov Frameworks

There is no denying the mathematical appeal of decision models and their more complex offspring as Markov and related processes. Putting on one side the issue of how the models are intended to generate evaluable claims, the question is, first, the choice of health states to map the natural course of the disease and, second, the transition probabilities between health states and, the absorbing state of death. Even within the reference case structure, there is no unique standard, even within a specific disease state, for the choice of health state, the time interval for the health state or the transition probabilities for that choice.

Clearly, QALY estimates will be sensitive to the number of health states, the time spent in each health state over the projected lifetime and the utilities assigned to the time spent in each health state. But this is not the point. Irrespective of potential challenges to the model structure, the fact that a lifetime framework is common to all models, means they will all lack credibility in the absence of evaluable and replicable claims

for competing DMT choice and the outcomes from supportive care. As lifetime claims are, by construct, impossible to evaluate, competing models can only be defended by perceptions that one is more 'realistic' or 'sufficient' than another. A comparison that has more in common with epistemological debates than more practical issues of their relevance for formulary decisions.

It is worth noting that criticisms as to the limited evidence base on which modeled claims in MS are based are not the first time these issues have been raised. Almost ten years ago, an Editorial in the *Journal of Managed Care Pharmacy* asked whether pharmacoeconomic modeling in MS was 'building houses on sand'³¹. Identified weaknesses in the evidence base included: (i) the short duration of RCTs to support the various DMTs; (ii) lack of follow-up from the RCTs; (iii) the high drop-out rate; (iv) incomplete reporting of dropouts; and (v) potential elimination of treatment effect by including missing data.

Model Input Pricing

A weakness of cost-effectiveness modeling is that it is always possible to demonstrate that a product is cost-effective by a judicious choice of input prices, in particular unit drug costs and the annualized cost of treatment. An initial market entry wholesale acquisition cost (WAC) may be selected to establish a cost-effectiveness claim. Once on formulary or accepted as part of a pharmacy benefit management (PBM) package – which may also involve unknown discounting and kickbacks – the manufacturer is then free to engage in a policy of regular price increases. Indeed, a common gambit in pharmaceutical pricing is, once a product has been approved for formulary listing, to increase WAC on a regular basis over the life of the product. If sales are underperforming, the price can be raised to meet revenue targets.

As the claims for cost-effectiveness that initially 'convinced' the formulary committee or PBM to accept the product are seldom (if ever) revisited over the product life cycle, the manufacturer has significant scope for price increases. If there is a public outcry, then the manufacturer can fall back on the well-tried defense of offering co-payment coupons. As detailed in the recent report from the Office of Inspector General (OIG) on the prevention of copayment coupon use for Medicare Part D drugs, the increasingly widespread use of coupons could impose significant costs on the part D program by encouraging use of more expensive brand name drugs³². The motivation for pharmaceutical manufacturers is not entirely benevolent: coupons are a well-tried tactic to retain market share in the face of generic competition and to encourage greater adherence and persistence to enhance the revenue stream.

Pricing behavior for the various DMTs in MS is no exception to this policy of regular price increases. Indeed, in this class of drugs, the pricing trend over the last 20 years has been particularly egregious. A recent study by Hartnung et al, provides estimates of the trend in annual drug costs for nine DMTs from 1993 to 2014³³. Apart from the fact that DMT costs are two to three times bigger in the US than other countries, the principal finding is that DMT costs have accelerated well beyond inflation and substantially above rates for drugs observed in a similar biologic class. Annualized change in the cost of the nine DMTs in the evaluation ranged from 35.7% for glatiramer acetate to 7.9% for fingolimod. Four of the DMTs had annualized cost increases above 20% and four with annualized price increases between 13.0% and 16.8%. Natalizumab, for example, although being withdrawn briefly from the market between February 2005 and June 2006, increased in cost from \$25,850 in 2004 to \$64,233 in 2013 or an annualized increase of 16.2%. In terms of the overall 'costs' of care for commercially insured MS patients, a comparison of charges between 2006 to 2011 pointed to the continuing significant impact on total costs of the charges associated with drug costs (52.6% in 2011)³⁴. At the same time, the charge increase for DMTs far outstripped the charges for other medical services (95.7% vs. 32.4%).

As far as can be ascertained, the ICER cost-utility model takes no account of the likely increase in DMT costs over the lifetime of the model cohort. DMT annual acquisition costs were estimated through the third quarter of 2016 as a discounted WAC price. If, however, we consider the policies followed by manufacturers over the past ten, or even 20 years, it would be expected that the drug costs (and hence WAC discounted annualized costs) would be expected to be at least 10% and probably by more than 15% on an annualized basis. These alternative annualized cost assumptions would, presumably, have an impact of the cost-per-QALY estimates in the model and any proposals for discounting versus willingness-to-pay thresholds. Conversely, if, with the same willingness-to-pay threshold, the ICER model had been applied five years ago, the various DMTs would have been judged highly 'cost-effective' in terms of their annualized costs (which, again, would not have been adjusted for annualized cost increases).

Adherence and Persistence

The ICER model says nothing about adherence with therapy in MS. This is a surprising omission as the evidence from retrospective studies suggests that over the period to medication discontinuation the average medication possession ratio or proportion of days covered is < 0.8. This is considered sub-therapeutic. From a study of three western Canadian provinces, Evans et al report the proportion of days covered by DMTs ≥ 0.8 for only 76.4% (95%CI: 69.1 – 82.4%) of subjects at one year³⁵. Over the period of the study (1996-2012) the

combined proportion of subjects with optimal adherence was 42.4% (95% CI: 32.5% - 52.9%). The medication possession ratios ranged from 0.70 to 0.77.

Unadjusted estimates of MS medication persistence in the US by Halpern et al for the period 2000 – 2008 also point to sub-optimal behavior. With a patient follow-up period of up to 36 months (with patients enrolled for at least 12 months following the index DMT prescription) the results for four DMT therapies (IM interferon beta-1a, subcutaneous IFN β -1a, IFN β -1b and glatiramer acetate) give adherence percentages ranging from 48.5 to 62.3%. Again, these are well below those rates considered therapeutically acceptable. Similar patterns were observed from a retrospective study of DMT in Germany³⁶. The study found, for a two-year observation period an MPR of 39.9% for four DMTs (Avonex, Betaferon, Copaxone and Rebif).

Although adherence behavior is ignored, the ICER model attempts to capture discontinuation or persistence behavior. As noted, the discontinuation rates are based on the studies utilized in the clinical network meta-analysis. As such, they understate by a significant margin the MS DMT persistence rates reported in the literature. The Evans et al study referenced above estimates that, for the four DMTs average persistence with therapy from index prescription ranged from 471 to 508 days (1.29 to 1.39 years) while 15.2% (CI:9.7 – 23.1%) discontinued initial DMT therapy within 6 months and 27.1% (CI:19.7 - 36.2%) at 12 months. In Germany, to give a further example, overall persistence with DMTs at the end of 24 months was 32.3%.

The conclusion for persistence, therefore, is that only a minority of MS patients are persistent with their therapy at the end of 18 months to two years, with the overall proportion likely to have discontinued therapy before 30 months. These patterns are not uncommon in a number of disease areas. That being said, it seems odd that the ICER would assume discontinuation behavior that is clearly overly optimistic. While this may be a reasonable assumption to ensure the model continues to be populated with adherent and persistent imaginary patients in the long term, it points to the irrelevance of lifetime models when the overwhelming majority of patients discontinue therapy in the short term or switch, for an unknown period, to second line DMT therapy and then supportive care). A model where patients switch between DMTs over their lifetime (for which there is no evidence), would be the only justification for the ICER lifetime assumptions. Alternatively, after a few months of sub-optimal initial DMT therapy and early discontinuation the patient is presumably locked into the default condition of 'best supportive care'.

Replication of Claims

Overall, the evidence base for populating the ICER MS model is not only limited in the number of published studies (none of which attempt to replicate claims) but they are further constrained by their observation period. None of the phase 3 studies of the impact on competing DMTs extends for more than 2 years. Rather than taking at face value claims for comparative DMT performance drawn from a handful of clinical trials, a more substantive approach would be to focus on the question of replication: can the claims from trials based on a one to two-year time horizon be replicated with feedback to formulary committees? Instead, we are asked to buy into a framework that takes these claims at face value and builds a non-evaluable claims framework that extends for decades into the future. Replication of claims is a key issue as there is ample evidence for the difficulty, if not impossibility, of replicating phase 2 and 3 randomized clinical trial (RCT) claims³⁷.

There are a number of claims that could be assessed within a 2-year timeframe. These would relate not only to relapse and side effects, where the claims are developed, possibly in a short term model framework, from the DMT specific RCTs. These would establish a base for comparative claims as well as presenting the opportunity to assess in more detail the determinants of adherence and persistence behavior. Given the lifetime experience of MS, it seems odd that from the evidence presented, the majority of that time is apparently spent, not on a DMT or succession of DMTs, but on 'supportive' care.

Willingness to Pay Thresholds

Although the ICER approach places emphasis on willingness to pay thresholds as a guide to decisions as to whether or not manufacturers should offer discounts or reduce prices, there is no evidence to suggest that in the US formulary committees apply such thresholds in pricing negotiations. Two points should be emphasized. First, in the US, no one has the slightest interest in cost-per-QALY lifetime claims let alone cost-effectiveness thresholds; these are an existential *weltanschauung* which, while possibly resonating with a few academic groups, together with single payer systems who have embraced the NICE reference case to support resource allocation, are irrelevant to treating environments and pricing. It might be noted, in passing, that PHARMAC in New Zealand while adopting a reference case approach to product assessment, does not apply willingness-to-pay thresholds. Second, even if there were a modicum of belief in thresholds, they face the same objection as lifetime cost per QALY estimates. They are based on a model which lacks scientific credibility.

Which Simulation?

To illustrate the concern some formulary committees might have in accepting results from the ICER model, a reasonable question is how the model outcomes might vary with changes in assumptions? To illustrate this possibility, it is worth noting the differences in estimated costs per QALY and cost per relapse avoided, pairwise results for DMTs compared to supportive care in RRMS, between the draft and final ICER reports. These results are shown in Table 1.

The differences in the estimates from the two reports are in many cases quite substantial with the percentage differences in cost per additional QALY ranging from -44.0% for teriflunomide 7mg to 42.4% for alemtuzumab. The range for percentage differences for modeled claims where the cost per QALY has fallen between the draft and final report range from -1.8% to -44.0%; in those cases, where the cost has increased the range is from 16.3% to 42.4%. A similar picture emerges for the estimates cost per relapse avoided. The overall range here is from 30.4% for interferon β -1b 250 mcg (betaseron) to -43.7% for alemtuzumab. Where the cost per relapse avoided has increased the range is from 0.9% to 30.4%; where the cost has fallen the range is from -3.2% to 43.7%.

Presumably, future modifications to the model would result in revised estimates with each subject to a sensitivity analysis. A situation that could continue indefinitely with competing claims that 'their model is the one that should 'inform' decision makers. Access to the model would facilitate such experimentation which is one reason why countries that have adopted a reference case model typically claim commercial confidentiality as the reason for not allowing public access. The point remains: how are we to judge one version of an MS model from another and how much faith do we have in the 'final' version presented in the ICER report.

A Black Box Warning

If health technology assessment groups are to persist in developing lifetime cost-per-QALY models, then it might be advisable to issue these models with the equivalent of a black box warning. Readers, including formulary committee members evaluating a reference case submission, would be advised: (i) that lifetime (or long-term) modeled claims are non-evaluable; (ii) that it is entirely possible to construct a competing lifetime model to generate alternative claims; (iii) that if the lifetime model is intended to 'inform' decision makers, we have no idea if the claims are right or even if they are wrong; (iv) that claims for the superiority of one lifetime model over another based upon a review of assumptions regarding model structure or input parameters is a pointless exercise.

Although not necessarily part of a black box warning, it should be advised that the lack of scientific credibility that attaches to lifetime cost-utility models also extends (i) to projections based on the model for the budget impact of therapies and (ii) to claims for possible pricing discounts based on modeled cost-per-QALY estimates set against willingness-to-pay thresholds. If, of course, the model generates short term budget impact claims, then these can certainly be evaluated.

Conclusions

Perhaps this commentary has been too harsh in judging the ICER MS model by the standards of normal science. It may be that the ICER has never intended that claims made should be evaluable or replicable. Perhaps they should be seen as 'our best non-evaluable projections'? Unfortunately, if this model for evaluating competing therapies in MS is seen as one among many possible projections, then the exercise loses credibility. Why should this model be superior to other, as yet unknown, projections? Are we to judge this model on its correspondence to what its authors and stakeholders see as the reality of treating MS? Is the model sufficient in its representation of reality such that the conclusions are necessarily entailed? A correspondence with a future, yet unknown, reality that is sufficient to support its claims? A model that assumes manufacturers will maintain present DMT price levels? A model which offers a 'superior' vision compared to other cost-effectiveness modeled claims for DMTs? A model which claims to look 20 or 30 years ahead from an index DMT prescription, putting to one side the pipeline of new entrant and competitor therapies?

The fundamental issue is that in accepting a US version of the NICE reference case with willingness-to-pay thresholds, the possibility that short term models would make more sense to formulary committees is ignored. Rather than focusing on credible, evaluable and replicable claims for the DMTs, the model looks to an imaginary future where there is no chance of the claims being evaluated. There is no feedback to formulary committees and no appreciation of the pressing needs in MS to improve the evidence base for decision making. There is no concept of progress in understanding the impact of DMTs in target MS populations through hypothesis testing and systematic observational studies. A more fruitful line of inquiry, if ICER wished to take up the challenge, would be to abandon the reference case. Focus instead on generating short term evaluable claims for competing DMTs, basing any recommendations for comparative product pricing and discounts on modeled claims that meet the standards of normal science: claims that are credible, evaluable and replicable.

Table 1
Cost Per Additional QALY and Cost Per Relapse Avoided, Draft and Final ICER Modeled
Results for Relapsing-Remitting Multiple Sclerosis

Drug	Cost per Additional QALY (draft)(\$)	Cost per additional QALY (final)(\$)	Percentage difference	Cost per relapse avoided (draft)(\$)	Cost per relapse avoided (final)(\$)	Percentage difference
Teriflunomide 7mg	517,764	289,970	-44.0	482,101	414,754	-14.0
Interferon β -1a 22mcg (Rebif)	432,633	341,359	-21.1	508,612	430,998	-15.3
Interferon β -1a 30 mcg (Avonex))	352,949	331,381	-6.1	946,187	954,935	0.9
Teriflunomide 14 mg	320,696	236,954	-26.1	432,990	400,198	-7.6
Glatiramer acetate 20 mg (Copaxone)	308,808	303,302	-1.8	421,467	407,877	-3.2
Glatiramer acetate 20mg (Glatopa)	204,268	194,253	-4.9	278,789	261,230	-6.3
Interferon β -1a 44 mcg (Rebif)	314,864	284,135	-9.8	445,299	418,760	-6.0
Dimethyl fumarate	260,631	211,444	-18.9	283,560	332580	17.3
Fingolimod	270,066	238,970	-11.5	253,402	276,100	9.0
Interferon β -1b 250 mcg (Betaseron)	183,617	214,355	16.7	359,069	468,100	30.4
Interferon β -1b 250 mcg (Extavia)	159,412	185,369	16.3	311,736	404,801	30.0
Peginterferon B-1a	256,255	238,321	-7.0	478,512	514,656	7.6
Natalizumab	215,180	208,987	-2.9	205,442	228,597	11.3
Daclizumab	270,373	222,782	17.6	302,209	344,719	14.1
Alemtuzumab	60,209	34,659	42.4	76,655	43,178	-43.7

Source: Table 22 draft ICER report; Table 22 final ICER report

References

- ¹ Langley PC. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ.* 2015;18(12):993-99
- ² Langley PC. Supporting formulary decisions: The discovery of new facts or constructed evidence? *Inov Pharm.* 2016; 7(2): No.15
- ³ Langley PC. Sunlit uplands: the genius of the NICE reference case *Inov Pharm.* 2016;7(2): No.12
- ⁴ Langley PC. He ao pohewa: The PHARMAC Prescription for Pharmacoeconomic Analysis in New Zealand and the standards of normal science. *Inov Pharm.* 2016;7(2): No. 13
- ⁵ Langley PC. Na domhain shamhlaíochta: Formulary submission guidelines in Ireland and the standards of normal science. *Curr Med Res Opin.* 2016;32(5):
- ⁶ Langley PC. Imaginary Worlds: The European Network for Health Technology Assessment (EUnetHTA) Recommendations for Health Economic Evaluations. *Inov Pharm.* 2016;7(3): No. 3
- ⁷ Langley PC. Modeling imaginary worlds: Version 4 of the AMCP Format for Formulary Submissions. *Inov Pharm.* 2016;7(2): No. 11.
- ⁸ Langley PC. Great expectations: Cost-utility models as decision criteria. *Inov Pharm.* 2016;7(2): No.14
- ⁹ Langley PC. Cost-Effectiveness and Formulary Evaluation: Imaginary Worlds and Entresto Claims in Heart Failure. *Inov Pharm.* 2016;7(3): No. 6
- ¹⁰ Langley PC. Imaginary worlds: Modeled claims for cost-effectiveness published in *PharmacoEconomics* January 2015 to December 2015. *Inov Pharm.* 2016;7(2): No. 9
- ¹¹ Langley PC, Rhee TG. Imaginary worlds: A systematic review of the status of modeled cost-effectiveness claims published in the *Journal of Medical Economics* from January 2015 to December 2015. *Inov Pharm.* 2016;7(2): No. 16
- ¹² Langley PC, Rhee TG. Imaginary worlds: The status of modeled economic evaluation claims published in *Value in Health* January 2015 to November 2015. *Inov Pharm.* 2016;7(2): No. 18
- ¹³ Langley PC, Rhee TG. Imaginary Worlds: The status of simulation modeling in claims for cost-effectiveness in diabetes mellitus. *Inov Pharm.* 2016;7(2): No. 17
- ¹⁴ Langley PC, Rhee T. Imaginary Worlds: The Status of Modeled Quality Adjusted Life Year Claims for New Oral Anticoagulants in Atrial Fibrillation Published Between January 2012 and February 2016. *Inov Pharm.* 2016;7(3): No. 7
- ¹⁵ Langley PC. Next Generation Sequencing Claims in Precision Medicine: Questions a Formulary Committee Should Ask. *Inov Pharm.* 2016;7(4): No. 11
- ¹⁶ Langley PC. Creative Destruction: Next Generation Sequencing in Drug Development, Formulary Evaluations and Pricing. *Inov Pharm.* 2016;7(4): No. 13.
- ¹⁷ Langley PC. Denkbeeldige wereld: The new Dutch guideline for economic evaluations in healthcare. *Inov Pharm.* 2017;8(1): No.
- ¹⁸ Langley PC. Dreamtime: Version 5.0 of the Australian guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). *inov pharm.* 2017;8(1): no.

- ¹⁹ Langley PC. HAS should not be NICE: Rejecting imaginary worlds in the French Technology Assessment Guidelines. *Inov Pharm.* 2017;8: No. 9
- ²⁰ Drummond M, Sculpher M, Torrance G et al.; Methods for the economic evaluation of health care programmes 3rd Ed. Oxford: University Press, 2005
- ²¹ Institute for Clinical and Economic Review. Disease-modifying therapies for relapsing-remitting and primary-progressive multiple sclerosis: Effectiveness and Value. Evidence Report. ICER: January 26, 2017.
- ²² Canada. Canadian Agency for drugs and Technologies in Health. Guidelines for the Economic evaluation of Health Technologies. 4th Ed. Draft Report. Ottawa: CADTH, October 2016
- ²³ National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: Nice, April 2013
- ²⁴ Caro JJ, Briggs A, Siebert U et al. Modeling good research practices – Overview: A report of the ISPOR-SMDM Modeling good practices task force – 1. *Value Health.* 2012 (15):796-803
- ²⁵ Eddy DM, Hollingworth W, Caro JJ et al. Model transparency and validation: A report of the ISPOR-SMDM modeling good research practices Task Force – 7. *Value Health.* 2012;15:843-850
- ²⁶ Correspondence: AMCP to ICER, September 12, 2016 www.amcp.org/WorkArea/DownloadAsset.aspx?id=21410
- ²⁷ Augustine J, Warholak TL, Hines LE et al. Ability and use of comparative effectiveness research by P&T Committee members and support staff. *J Man Care Spec Pharm.* 2016;22(6):618-625
- ²⁸ Tang DH, Warholak TL, Hines LE et al. Evaluation of pharmacy and therapeutic (P&T) committee member knowledge, attitudes and ability regarding the use of comparative effectiveness research (CER) in health care decision-making. *Res Social Adm Pharm.* 2014;10(5):768-80
- ²⁹ Villa L, Warholak TL, Hines LE et al. Health care decision makers' use of comparative effectiveness research: report from a series of focus groups. *J Manag Care Pharm.* 2013;19(9):745-54
- ³⁰ Institute for Clinical and Economic Review. Disease-modifying therapies for Relapsing-Remitting and Primary-Progressive Multiple sclerosis: Effectiveness and Value. Draft Evidence Report. ICER: November 22, 2016
- ³¹ Curtiss FR. Pharmacoeconomic modeling of drug therapies for drug therapies for multiple sclerosis = Are we building houses on sand? *J Man Care pharmacy.* 2007;13(3):287-289
- ³² Office of Inspector General. Department of Health and Human Services. *Manufacturer safeguards may not prevent copayment coupon use for Part D drugs.* September 201 (OEI-05-12-00540).
- ³³ Hartung D, Bourdette D, Ahmed S et al. The cost of multiple sclerosis drugs in the US and the pharmaceutical industry. Too big to fail? *Neurology.* 2015;84:2185-2192
- ³⁴ Carroll C, Fairman K, Lage MN. Updated cost-of-care estimates for commercially insured patients with multiple sclerosis: retrospective observational analysis of medical and pharmacy claims data. *BMS Health Services Res.* 2014;14:286
- ³⁵ Evans C, Marrie R, Zhu F et al. Adherence and persistence to drug therapies for multiple sclerosis: A population-based study. *Multiple Sclerosis Related Disorders.* 2016;8:78-85

³⁶ Hansen K, Schussel K, Kieble M et al. Adherence to disease modifying drugs among patients with multiple sclerosis in Germany: A retrospective cohort study. *PlosOne*. 2015. DOI:10.1371/journal.pone.0133279

³⁷ Baker M. Is there a reproducibility crisis? *Nature*. 2016;533:452