

9-26-2017

## More Imaginary Worlds: A Systematic Review of the Status of Modeled Cost-Effectiveness Claims Published in the Journal of Medical Economics from January 2016 to December 2016

Paul C. Langley

*University of Minnesota*, [langley@maimonresearch.com](mailto:langley@maimonresearch.com)

Taeho Greg Rhee

*Yale University*, [tgrhee.research@gmail.com](mailto:tgrhee.research@gmail.com)

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

### Recommended Citation

Langley PC, Rhee TG. More Imaginary Worlds: A Systematic Review of the Status of Modeled Cost-Effectiveness Claims Published in the Journal of Medical Economics from January 2016 to December 2016. *Inov Pharm.* 2017;8(3): Article 13.  
<http://pubs.lib.umn.edu/innovations/vol8/iss3/13>

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

## More Imaginary Worlds: A Systematic Review of the Status of Modeled Cost-Effectiveness Claims Published in the *Journal of Medical Economics* from January 2016 to December 2016

Paul C. Langley, PhD<sup>1</sup> and Taeho Greg Rhee, PhD<sup>2</sup>

<sup>1</sup>College of Pharmacy, University of Minnesota

<sup>2</sup>Department of Internal Medicine, School of Medicine, Yale University

### Abstract

In 2016, a review of modeled cost-effectiveness studies published in the *Journal of Medical Economics* between January 2015 and December 2015 was presented in *INNOVATIONS in Pharmacy*. The purpose of this review, together with similar reviews for studies published in calendar 2015 in *Value in Health* and *Pharmacoeconomics*, was to consider whether these modeled claims for cost-effectiveness met the standards of normal science: were the claims made credible, evaluable and replicable? A total of 32 studies were identified. None of the studies presented their claims or projections in an evaluable form and none suggested how they might be evaluated. None met the standards of normal science. The claims made for cost-effectiveness were either impossible to verify, or if potentially verifiable, were not presented in an evaluable form. The studies lacked credibility. There was no basis for assessing whether the claims were right or even if they were wrong. The purpose of the present review which covers cost-effectiveness studies published in the *Journal of Medical Economics* between January 2016 and December 2016 is to revisit this question of the credibility of the claims made against the standards of normal science. A total of 40 cost-effectiveness studies were identified. Although 14 had a timeframe of 5 years or less and had the potential to provide short-term evaluable claims, none addressed the issue of claims evaluation and the possible protocols that would support empirical assessment. Of the balance, 19 presented results as unevaluable lifetime modeled claims. The conclusions from the 2016 review remain unchanged.

**Keywords:** *Journal of Medical Economics*, systematic review, cost-outcomes, imaginary worlds, evaluation

### Introduction

Some 12 months ago a commentary was published in *INNOVATIONS in Pharmacy* reviewing, from the perspective of the standards of normal science, modeled technology assessment claims published in the *Journal of Medical Economics* from January 2015 to December 2015<sup>1</sup>. This systematic review concluded that of the 32 cost-effectiveness studies reviewed, none presented their claims or projections in an evaluable form and none suggested how they might be evaluated. None met the standards of normal science. The claims made for cost-effectiveness were either impossible to verify, or if potentially verifiable, were not presented in an evaluable form. The studies lacked credibility and were best seen as constructed imaginary worlds. There was no basis for assessing whether the claims were right or whether they were wrong, and in the majority of cases they would never know, as

the timeframe for the analysis guaranteed that the claims were immune to failure. Reviews of studies published in *Pharmacoeconomics* and *Value in Health* over the same calendar 2015 period came to the same conclusion<sup>2,3</sup>.

The *Journal of Medical Economics* review pointed out that this lack of scientific credibility is a major concern. If medical economics is to advance through the formulation and testing of hypotheses, then editors of journals should consider whether or not to set standards for the acceptance of publications to include the requirement for evaluable claims and the results of claims assessment. If this is not acceptable, then it should be made clear that published papers are simply imaginary worlds or thought experiments.

These reviews of published studies are part of a series of commentaries in *INNOVATIONS in Pharmacy* over the last 12 months that have focused on the evidentiary standards for claims assessment required or recommended by technology assessment agencies such as the National Institute for Health and Care Excellence (NICE) in the UK and the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia, professional groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the Academy of Managed Care Pharmacy (AMCP) in the US and independent research groups such as the Institute for Clinical and Economic Review (ICER)<sup>4,5,6,7,8,9,10</sup>. The common theme in these commentaries, and papers published earlier in 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  
Email: [langley@maimonresearch.com](mailto:langley@maimonresearch.com)  
Web: [www.maimonresearch.net](http://www.maimonresearch.net)

*Journal of Medical Economics* in a special supplement, is that the standards proposed and accepted in health technology assessment in the construction of non-evaluable claims should be seen as pseudoscience; as intelligent design rather than natural selection<sup>11 12 13</sup>.

In case this characterization might appear as an unnecessarily harsh judgement on standards that have been in place for 30 years or more and which have generated literally thousands of published, peer review studies and evaluations by technology assessment groups, the commentaries recently pointed to the latest version of the guidelines released in March 2017 by the Canadian Agency for Drugs and Technologies in Health (CADTH). The CADTH guidelines made it quite clear that the technology assessment framework is not to be judged by the standards of normal science<sup>14 15</sup>. Rather, CADTH guidelines are designed to set criteria for the construction of imaginary worlds or simulations to support cost-outcomes claims; to 'inform' health system decision makers, not to test hypotheses. Presumably, within disease areas and for specific product comparisons, this additional modeled information will be added to the imaginary modeled claims already published or submitted to other agencies for the same or similar products.

The purpose of this second review of cost-effectiveness studies published in the *Journal of Medical Economics* is to consider whether the standards of normal science continue to be put to one side, with the Journal continuing to accept simulated, non-evaluable claims generated by imaginary worlds. The period covers January 2016 December 2016. This review follows on one recently published which revisited cost-effectiveness claims in *Value in Health* for studies published between January 2016 and December 2016<sup>16</sup>. This will be followed in a forthcoming commentary that will revisit studies published in the same time period in *Pharmacoeconomics*.

### Methods

A systematic review, following the PRISMA-P checklist (MeSH terms 'cost', 'cost effectiveness', 'Markov', 'QALY') of all papers published in *the Journal of Medical Economics* in calendar 2016 was undertaken<sup>17</sup>. In order to judge whether the modeled claims presented met the standards of normal science four questions were considered:

- Is the study model capable of generating testable claims?
- Did the study attempt to generate testable claims?
- Did the study suggest how the claims might be evaluated?
- Did the study caution readers as to the implications of generating non-testable claims?

Each author independently reviewed the selected studies with consensus agreement reached on the assessment.

A testable claim was defined as one that could be evaluated empirically in a timeframe relevant to the needs of a formulary committee (ideally a period of 2 to 3 years). This period was chosen because a testable claim was seen as provisional. A product or device could, in this context, be accepted by a committee for formulary listing, but subject to an agreement with the manufacturer to report back to the committee with evidence to support the claims made. These claims could be for product comparative effectiveness, for the impact of the product on resource utilization or some combination of these to support a claim for incremental cost-effectiveness. The claim for comparative effectiveness could encompass clinical endpoints as well as those captured as patient reported outcomes (PROs).

In judging whether or not a model might support testable (1.e., falsifiable) claims, even if the possibility was not considered by the authors(s), three characteristics of the model are important. These are (i) the modeling framework, (ii) the choice of primary outcome measure; and (iii) the time frame for the model. A Markov or discreet event simulation model with a lifetime perspective and with discounted cost per QALY claims as the primary endpoints would be one where comparative claims would be impossible to evaluate. There is no chance of falsification, feedback to decision makers or replication. It would be assessed as immune to failure. Against this, a simple, trial-based decision model with a timeframe of 12 to 18 months with claims expressed in clinical (including PRO) and resource utilization endpoints would be open to hypothesis testing and feedback to a formulary committee. Even with a short-term time horizon, however, the choice of outcome may not be evaluable outside of a protocol-driven observational study. If health care systems do not collect specific QALY measures on an ongoing basis then it is impossible to evaluate cost-per-QALY willingness-to-pay threshold claims from integrated data bases. This assumes, of course, that the QALY measure that might be collected is consistent with the measure utilized in the simulation model.

While claims may be potentially evaluable, what is typically missing in modeled claims is any direction as to how the claims might be assessed in treatment practice. Presenting the same model that has been re-formulated for different markets and different countries, with the model consistently generating positive claims for a sponsor's product seems rather pointless in the absence of a protocol that proposes how the modeled claims can be evaluated. This is a feature conspicuous by its absence in technology assessment submission guidelines.

At the same time, this evaluation also searched for systematic reviews published in calendar 2016. Two questions were considered relevant:

- Did the systematic review of economic evaluations address the issue of the credibility, evaluation and replication of clinical claims in the respective modeled economic evaluations?
- Did the systematic review recommend (or caution against) accepting the claims from the modeled economic evaluations as the basis for formulary decisions?

Finally, the review considered whether or not the cost-effectiveness study was funded or supported by a pharmaceutical or device manufacturer. The question addressed was whether or not the results of the cost-effectiveness modeling supported the manufacturer's product.

### Results

The review identified 40 cost-effectiveness studies (Table 1) together with one systematic review of economic models in moderate-to-severe asthma and COPD <sup>18</sup>.

### Economic Models

None of the studies met the standards of normal science in providing evaluable and replicable claims in a form that could be supported by a claims assessment protocol. Overall, 19 of the 40 studies (47.5%) presented their claims in a lifetime model framework <sup>19 22 24 26 27 28 29 30 31 32 34 35 36 40 42 44 45 46 53 56</sup>. In addition, if we take a cutoff of 6 years or more in the model timeframe, a further seven models would be considered to have non-evaluable claims <sup>37 38 49 50 51 54 57</sup>. The balance of the models with a timeframe of 5 years or less had the potential to generate or did produce claims that were potentially evaluable <sup>20 21 23 25 33 39 41 43 46 47 48 52 55 58</sup>.

Among those models that presented a lifetime framework for generating imaginary claims, 12 utilized a Markov framework <sup>24 29 30 31 32 36 35 40 44 45 46 56</sup>. A further 3 studies utilized the lifetime IMS CORE diabetes model <sup>22 26 42</sup>. In addition, 4 studies utilized a Markov framework for timelines of 10, 25 and 70 years together with two studies utilizing the Cardiff Diabetes Model for 40 year time horizons <sup>51 38 50 49 54 57</sup>. In total, therefore, 26 of the 40 economic evaluations (65.0%) presented modeled claims that were immune to failure given the timeframe of the model.

In terms of the questions raised:

- Two thirds of the published economic evaluations failed to present credible claims that met the expected standards of normal science in presenting claims for cost-effectiveness that were potentially evaluable and replicable

- None of the studies attempted to present evaluable claims, although by default a handful of studies presented claims that could be evaluated
- None of the studies considered how their claims, even if potentially evaluable, could be assessed in treatment practice
- None of the studies cautioned the reader that their claims might not be evaluable and that the reader would have no idea if they were right, wrong or potentially misleading

### Systematic Review

The Einarson et al systematic review of models in moderate-to-severe asthma and COPD identified 53 articles, 14 of these were for patients with asthma and 39 in COPD. Models were defined under three heads: (i) prospective trial based models (8); (ii) predictive decision models (27); and (iii) retrospective patient record models.

Key points noted in summarizing the results were:

- Markov models accounted for 22 of the 27 decision models
- Among the Markov asthma models:
  - only 3 reported validation of some sort, but most provided no details
  - two used a time period of 12 weeks, seven used 1 year, one used 10 years, six used a lifetime
  - ten examined inhaled corticosteroids and 9 omalizumab
  - nine were sponsored by pharmaceutical companies, 3 were not stated and 2 were public sector (NCE, National Heart and Blood Institute)
- Among the asthma related pharmaco-economic studies 11 used RCTs as primary source for clinical data
- Overall, among the asthma models results varied widely: omalizumab was found to be cost-effective in 9 out of 14 studies but not cost-effective in 5
- Among the COPD models:
  - Thirty out of 39 were sponsored by the pharmaceutical industry
  - Nineteen models did not mention any validation process
  - Thirteen of the studies used a time frame between 3-5 years, 10 used a time frame of 1 year
  - A Markov model was used in 19 studies

The key point noted in the discussion was that study parameters for the most part were not validated: face validity

was most common while internal, external and predictive validity were largely absent.

In terms of the two questions raised in respect of this systematic reviews of modeled claims:

- The review did not address the question of the credibility of the modeled claims, et alone issues of whether the claims were evaluable and replicable
- The review did not caution the reader that in accepting the claims made that they did not meet the standards expected of normal science as inputs to the formulary evaluation process

## Discussion

The fact that in those studies funded or sponsored by a pharmaceutical or device manufacturer in the *Journal of Medical Economics* support the manufacturer's product is, perhaps, unsurprising. After all, the previous review of modeled studies in the *Journal of Medical Economics* made the same observation as has the more recent review of studies in *Value in Health*<sup>13</sup>. Indeed, the more cynical reader may wonder why there is any surprise at all. After all, does the apparent lack of any challenge to these published claims simply reflect the difficulty, or more appropriately, the impossibility of challenging the construction of imaginary worlds? Or, does the lack of challenge reflect a lack of interest by decision makers in imaginary worlds and their imaginary claims? Rather than taking claims at face value, holding to the belief that the standards of normal science can be put to one side by rejecting hypothesis testing in favor of 'information provision', decision makers may see these publications as marketing exercises that they can put to one side.

### Discretion in Modeled Claims

It is worth recalling the 1994 Editorial in the *New England Journal of Medicine* regarding the discretionary nature of cost-effectiveness studies and the Journal's policy on accepting such studies for review<sup>19</sup>. The position taken was that because of the discretionary nature of the methods used to analyze cost-effectiveness it is 'incumbent on authors journal editors and the funders of these studies to minimize any source of bias'. Two conditions were proposed: (i) *any study supported by industry must be funded by a grant to a not-for-profit entity such as a hospital or a university, not to an individual or group of individuals*; and (ii) *the Journal must receive written assurance that the agreement between the authors and the funding company ensures the authors' independence in the design of the study, the interpretation of data and writing of the report, and decisions regarding publication, regardless of the results of the analysis*. Studies will not be reviewed by the Journal if *any of the authors is receiving a direct salary from*

*the sponsoring study or a competing company, or if any author has an equity interest in, an ongoing consultancy with, or membership on the scientific advisory board of such a company, or a related patent pending.*

The discretionary nature of both clinical and cost-effectiveness claims has continued to be of concern, notably in the wider context of integrity in research and publications, including the issues surrounding the replication of randomized clinical trial (RCT) results. Addressing the question of integrity (How do you know it is true?) Buckwalter et al argue that *The current high-stakes research environment has been characterized by an increase in plagiarism, falsification or manipulation of data, selected presentation of results, research bias, and inappropriate statistical analyses* where research findings can be biased by 'falsification or manipulation'<sup>20</sup>. A recent Cochrane review of interventions to prevent misconduct and promote integrity in research and publication found that 'Overall, there is very low quality evidence that various methods of training in research integrity had some effect on participants' attitudes to ethical issues but minimal (or short-lived) effects on their knowledge'<sup>21</sup>.

At the same time, as noted in previous commentaries in *INNOVATIONS in Pharmacy* have been increasing concerns over the inability of researchers to replicate the results of RCTs. A *Nature* survey, reported in May 2016, found that over 70% of researchers had tried and failed to reproduce other experiments and more than half failed to reproduce their own experiments<sup>22</sup>. At the moment we do not seem to have struck a balance between tolerating tentative conclusions and honest errors and efforts to improve reproducibility in biomedical claims<sup>23</sup>.

Questioning the integrity of published clinical research is reinforced by more fundamental claims that most published research findings can be shown to be false<sup>24</sup>. If this is accepted then further doubt is cast on those modeled claims that rely on one or two clinical trials (typically the sponsors) or on network indirect comparisons between (possibly false?) comparator trials to support the models clinical assumptions. In the absence of claims replication the clinical evidence base may be considered insufficient to support modeled claims for formulary listing and pricing. Health care decision makers would, as detailed below, be justified in asking for the clinical outcomes assumptions to be evaluated and reproduced, together with claims made for cost-effectiveness. In the case where the modeled claims are immune to failure, those sponsoring such claims run the risk of them being ignored and rejected out of hand as they fall at the first hurdle for establishing their credibility.

### Discretion and Imaginary Worlds

One aspect of modeled claims for cost-effectiveness that was not addressed by the Editorial in the *New England Journal of*

*Medicine* is the apparent disregard for the standards of normal science; the acceptance of modeled claims for cost-effectiveness that are non-evaluable. This is not a question of replication, but rather one, as noted in the introduction to this review, of immunity to failure. This rejection of these standards, as noted in the commentaries disregards standards that have been in place for 350 years. The motto of the Royal Society (founded 1660; Royal Charter 1662) clearly expresses this: nullius in verba ('take no man's word for it'). If a modeled claim is unable to generate, credible, evaluable and replicable claims then it should be rejected. The claims should be seen as pseudoscience. A conclusion that is reinforced given the discretion and latitude that is open to model builders in constructing long-term or lifetime cost-outcomes and cost-utility models; time horizons that may stretch for decades in modeled simulations. Examples detailed in this review would include claims generated by the Cardiff Diabetes Model with time horizons of that effectively preclude any empirical challenge.

Unfortunately, in health technology assessments, as detailed again in previous commentaries in *INNOVATIONS in Pharmacy*, agencies and professional organizations such as ISPOR and the AMCP, seem determined to keep the door open to standards for technology assessment that encourage the construction of imaginary worlds. The NICE reference case has been widely accepted as the 'gold standard'. It is, therefore, not surprising that journal editors are quite prepared to accept modeled claims for competing products and devices that are outside the standards of normal science.

Indeed, model builders have considerable latitude even within the so called 'reference case' to construct modelled lifetime comparative claims. Professional groups and technology assessment agencies ask the reader to take at face value extrapolations beyond the remit of clinical trials. In the present review 19 of the 40 studies (47.5%) present their claims in a lifetime model<sup>19 22 24 26 27 28 29 30 31 32 35 36 40 42 44 45 46 53 56</sup>. Of these, 12 utilized a lifetime Markov framework<sup>24 29 30 31 32 36 35 40 44 45 46</sup><sup>56</sup>. A further 3 studies utilized the lifetime IMS CORE diabetes model<sup>22 26 42</sup>. In addition, a further 5 studies utilized a Markov framework for timelines of 5, 7, 10, 25 and 70 years together with two studies utilizing the Cardiff Diabetes Model for 40 year time horizons<sup>39 51 38 50 49 54 57</sup>. None of these studies presented claims that could be evaluated.

### Threshold Values

Irrespective of the choice of imaginary world, the primary purpose of the majority of these studies would appear to be to demonstrate that, given the choice of model, its structure and input variables, that the claims made for the cost-effectiveness of the sponsor's product can be considered because, unless they are shown to be dominant in respect of comparators, the

cost-per-QALY claims falls with 'accepted' or 'recognized' willingness-to-pay thresholds. The thresholds deemed to be appropriate in the various studies vary widely. Some studies adopted notional willingness-to-pay thresholds with US\$50,000 being popular, although, if considered appropriate, this could be extended to US\$100,000 or US\$150,000 to claim cost-effectiveness and appropriateness for formulary listing. Other studies adopted the NICE thresholds of £20,000 and £30,000. For studies based in European treating environments a popular threshold value was €30,000 with a few studies applying the World Health Organisation (WHO) proposed per capita GDP 3-times multiple. It is not clear whether the model outcomes determined the threshold value chosen or the threshold value determined the model. In any event, the argument is moot as the claims for threshold compatibility, including applications of probabilistic sensitivity analysis, were unevaluable. The focus in all cases appeared to support claims for the cost-effectiveness of the sponsor's product. Whether this reflects publication bias is unknown. We have no idea of how many other, potentially competing, modeled claims were developed which have not been or have yet to be published.

Given the popularity of published constructed claims that conclude that a product is cost-effective, there is surprisingly perhaps no consensus on the use of threshold values, what thresholds represent and the construction of threshold values<sup>25</sup>. Questions on how we might value health improvements and the opportunity cost of budgetary decisions might be more illuminating if we put to one side constructed and unevaluable incremental cost per QALY claims and focused instead on claims that are evaluable and can be reported on to health decision makers in a meaningful timeframe. As it stands, and has been pointed out again in previous *INNOVATION in Pharmacy* commentaries, there is no agreed standard for measuring utilities for cost-per-QALY claims and there is unlikely to be one. Although there are a number of generic and disease-specific instruments that have been promoted as the basis for utility metrics, clinical trials are typically not powered for utility endpoints. As a result the imaginary lifetime models will go to the literature to justify the choice of utility metric. As well, at least in the US, there is little interest in QALY claims and willingness to pay thresholds. The exception here being the modeled (yet unevaluable) claims, typically from a long-term or lifetime perspective, that have been promoted by the ICER to support recommendations for the 'appropriate' pricing of pharmaceutical products.

It is also worth noting that NICE, in attempting to move to a more 'flexible' version of the EQ-5D (the 'reference case' mandated instrument), adopting the EQ-5D-5level instrument as a successor to the EQ-5D-3level version has found that the two generate different utility profiles and it is not possible to crosswalk from the EQ-5D-3L to the EQ-5D-5L. This illustrates a more general point, and a strong argument for abandoning cost-per-QALY claims, that different instruments generate

different results. Achieving a willingness-to-pay cost per QALY threshold with one instrument is no guarantee that an alternative, yet equally valid, instrument will come to the same conclusion.

### **Comorbidities**

As noted also in previous commentaries, the presence of comorbidities and their potential impact on outcomes, costs and compliance is typically overlooked in modeled long-term claims for cost effectiveness. Given the likelihood of older patient groups presenting with one or more comorbidities this is a significant oversight by model builders. This is recognized, for example, in the Lobotesis et al assessment of combination stent-retriever thrombectomy in acute ischemic stroke<sup>59</sup>. The authors caution that there may be limitations in relation to resource use data where acute and long term costs are taken from a study based on patients with atrial fibrillation and an average age of 80 years. Apparently, these patients are older than those in the base case clinical study used; they also suffer from atrial fibrillation with possible additional comorbidities.

### **Assistance and Persistence**

A further issue with modeled claims is the treatment of adherence and persistence with therapy. This is seldom acknowledged in claims made, although the fact that in many disease areas only a minority of patients are compliant after two to three years from index prescription will presumably impact long-term claims. One exception here is the qualification in the application of the CORE diabetes model that the model does not take into account 'potential differences between the respective treatments in adherence and persistence that can influence both effects and costs'. As noted in previous commentaries, it may seem pointless to model over (in this case) 50 years when patients have possibly long-ceased to persist with the therapy<sup>26</sup>. Also, it should be noted, the possible impact of new products entering the type 2 diabetes market and consequent therapy switching is also not considered.

### **Peer Reviews of Imaginary Worlds**

The resources devoted by NICE and other agencies to the assessment of a manufacturer's modeled reference case cost-utility claims stand in contrast to the support given to peer reviewers who are asked to evaluate modeled claims for journals such as the *Journal of Medical Economics* and *Value in Health*. Given time constraints for peer review and claims by journals that they can offer 'rapid publication' and a short turnaround of papers (and you can often pay a 'quick expedition fee' to guarantee an even faster turnaround) suggests that, at best, the evaluation of a constructed imaginary world will be limited. This does not imply any dereliction on the part of journal editors or unpaid peer reviewers, but is simply a fact of life. Journals do not have the luxury of investing resources on a model review to the standards of a NICE or PBAC review.

One solution to the peer review question is to establish standards for accepting or rejecting modeled claims for cost-effectiveness: to ask that the paper meet two acceptance criteria: (i) that the claims are evaluative in a timeframe that is meaningful for feedback to health system decision makers; and (ii) that the model claims are supported by a protocol that details how the claims are to be evaluated in treatment practice. Perhaps as a first step journal editors (and certainly formulary committees) should reject any manuscript that begins with 'A lifetime Markov model was ....'.

### **Qualified Support**

In a few cases modeled claims comparisons provide only qualified support for the sponsor's product. This typically occurs when the modeled lifetime cost-per-QALY estimates exceed notional willingness-to-pay thresholds. Consider, for example, the Goeree et al modeled case for nivolumab in non-small cell lung cancer (NSCLC) in Canada<sup>50</sup>. Application of a lifetime Markov model or a progression-free survival model generated cost-per-QALY estimates versus the comparators docetaxel and erlotinib in excess of \$140,000. In this case, while pointing out that funding has been approved in Canada for oncology drugs where the \$100,000 willingness to pay threshold has been exceeded, the authors emphasize the unmet need within NSCLC and that given the perceived NSCLC disease burden in Canada, as evidenced by its epidemiological profile, nivolumab 'has the potential to bring significant health benefits to patients in comparison to standard chemotherapy options (p. 641)'. Nivolumab is seen to 'represent(s) a major advance in disease management', 'provides unprecedented survival benefits' and a more favorable adverse event profile. These aspects were not, apparently, captured within the model framework but should, it can be argued, be factored into a willingness to accept cost-per-QALY projections which, although immune to failure, are in excess of accepted notional thresholds.

### **Comparing Imaginary Worlds**

It is commonplace for the authors of one long-term imaginary world to compare their world with 'competitor' imaginary worlds. It is not clear what this achieves given that the various competitor models also yield claims that are typically immune to failure. All that can be achieved are claims that one model is more 'representative' of an unknown future than another model. The potential readers, if they are interested, have no basis in the absence of feedback from the evaluation of modeled claims to judge whether one model is in any way 'superior' to another. It seems such a pointless exercise. An exercise that could, presumably, continue indefinitely as manufacturer's contract with consultants to publish even more product-supporting, non-evaluative modeled claims in disease and therapy areas. While it might be unreasonable to draw to close an analogy with an imaginary world such as that represented by Gormenghast, the description of the novels in

the series as representing 'a sprawling, gothic structure' has a certain resonance<sup>27</sup>. But perhaps, as noted in a previous commentary, we could attempt to close the wardrobe door on Narnia and return to the real world of normal science<sup>28</sup>.

Discussions by authors of model 'limitations' also point to a number of common 'elements' that may limit or relegate the claims generated by the model. Apart from assumptions (or their absence) regarding compliance behavior and comorbidities, typical qualifications would include: (i) the use of clinical trials to support input values where the trials may lack external validity for input values; (ii) the short-term nature of the trial data; (iii) reliance on indirect comparisons for clinical endpoints; (iv) costs sourced from the literature which may not be those faced by payers; (v) absence of non-medical or indirect costs; (vi) arbitrary choice of health states for treatment progression; (vii) cycle length; (viii) estimation of transition probabilities for health states; (ix) assumed future patterns of therapy switching; (x) health state utility values; (xi) choice of utility measure; and (xii) survivorship profiles. Even with this checklist, there is still no objective basis, other than evaluating claims made, for declaring the superiority of one imaginary construct from another.

### **Systematic Reviews**

Given the dominance of constructed imaginary worlds in cost-effectiveness claims, it is reasonable to ask what role systematic reviews actually play. Apart, obviously, from the review pointing to the diversity, latitude and incompatibility of modeled claims within disease areas, it is far from clear that such reviews actually perform a useful service. This is in contrast to the Cochrane reviews, for example, which are anchored in the scientific method to support evidence-based medicine. To argue, as the CADTH guidelines so eloquently put it, that economic models are not intended to test hypotheses but merely to 'provide information', raises the question of how a formulary committee is to make sense of the diversity and incompatibility of the information presented. While it can be appreciated that sponsoring economic models to support a marketing strategy is an entirely legitimate exercise, the presence of highly technical competing yet incompatible messages is hardly a positive step forward. Are formulary committees expected to undertake a further review of systematic reviews? Perhaps, in such a review, a formulary committee following advice above to journal editors, should immediately reject any study that began 'A lifetime Markov model ...'.

The overall impression from the Einarson et al systematic review is the difficulty of comparing model results and coming to any general conclusions on the imaginary claims for cost-effectiveness of therapies, either against specific comparators or the 'standard of care' within either asthma or COPD<sup>18</sup>. In the case of omalizumab, the results varied widely, in some models supporting a cost-effectiveness case with others

coming to the opposite conclusion. The results for COPD presented such variation in the choice of the primary drug and its comparators, that it is also difficult, if not impossible, to come to any meaningful conclusions that might guide formulary decision making.

While the difficulties attendant upon any attempt to 'herd the cats' in such a variable modelling environment can be appreciated, the more substantive issues of credibility, evaluation and replication are not addressed in the review. On the information presented, there is no evidence that the issue of presenting evaluable claims was addressed in any of the studies. While the issue of the face validity of the model was apparently addressed in a few studies, internal external and predictive validity were largely ignored. Unfortunately, in setting the standards for validation in terms of those proposed by ISPOR and the Society for Medical Decision Making (SMDM) in their Good Practices Research reports on constructing imaginary worlds, the authors overlooked the importance of predictive validation as standing outside face validity, internal and external validity<sup>29 30</sup>. A modeled claim is not to be judged on the apparent 'validity' of its structure or input assumptions. This point was made in a recent critique of modeled claims<sup>11 12</sup>. Predictive validity is not an option in 'validation'. Failure to consider predictive validity puts the standards of normal science to one side. Irrespective of any conclusions regarding the extent to which a model is considered to have face, internal and external validity, the only test (unless we take the claims at face value) is empirical: can the claims be validated (or at least not falsified) in treatment practice. Otherwise, all that Einarson et al have accomplished is to point to the variability in constructed imaginary worlds. The paradox is that, at least in COPD and to a lesser extent in the asthma models, the time frame for many of the models presented is one that could support evaluable claims. The possibility that the claims presented could be supported by protocols and evaluated in clinical practice was not considered.

### **Potentially Evaluable Claims**

This review has identified a number of studies that have the potential to generate evaluable claims. These include the Einarson et al 1-year models for paliperidone in schizophrenia<sup>21 47</sup>, the Pettigrew et al colorectal cancer model<sup>23</sup>, the Qin et al overactive bladder model<sup>25</sup>, metastatic colorectal cancer model, the Xuan et al duodenal ulcer model<sup>52</sup> and the Goeree et al model of opioid-induced constipation<sup>58</sup>. While there is the potential to generate claims for a range of outcomes and summary ICER claims (including cost-per-QALY), none of the studies considered how these claims might be evaluated in treatment practice. This raises the issue of the role claims assessment protocols might have in supporting sponsored and other comparative claims for cost-effectiveness. If a formulary committee, as proposed in the Minnesota guidelines could insist that manufacturers underwrite protocols to support, and provide feedback on, clinical and cost-effectiveness claims,



then we would go a long way towards, not only abandoning the construction of imaginary worlds but addressing issues of the replication RCT-based clinical claims<sup>31</sup>.

If journals adopted a policy (i) of requiring cost-effectiveness claims to be accompanied by an assessment protocol and (ii) requiring manufacturers to report on the results of the evaluation in a future issue of the journal, this may go some way towards alleviating the concerns expressed by the Editors of the *New England Journal of Medicine*. Instead of attempting to enforce review ('entry') requirements, which would still allow the submission of claims based on imaginary worlds (e.g., a recent modeled imaginary world review in the *Journal of the American Medical Association* of PCSK9 inhibitors<sup>32</sup>), the Journal could adopt a protocol policy which would allow for manufacturer' sponsored submissions, but with the 'integrity' of the model claims checked by the need to feed back the results of a effectiveness trial or prospective observational study.

As an additional check, the Journal could require peer reviews of both the initial modeled claims and the report of claims evaluation to be published. The former would meet any concerns as to the quality and feasibility of the claims assessment protocol while the latter would hopefully ensure that once published the claims could be assessed in other treatment environments, providing an accumulating body of evidence to support clinical and outcomes claims for the product.

### Conclusions

In the recently published review of modeled cost-effectiveness claims in *Value in Health*, the discussion concluded on a pessimistic note: it is unlikely that present practice would be overturned. Rather, we could expect a continuation of the commitment to creating and publishing claims that are 'immune to failure'. It was pointed out that Richard Dawkins, in *Unweaving the Rainbow*, recognizes our willingness to feed on 'superstition, the paranormal and astrology'. Or, as he labels this, our continuing appetite for being 'Hoodwink'd with faery fancy'<sup>33</sup>. We might extend the reference and argue that the 'imaginary world' meme is so well entrenched in health technology assessment that the biologically valuable tendency of individuals to imitate (c.f., the Dawkins' example of the opening of milk bottle tops among European tits or American Chickadees) perpetuates the acceptance of lifetime Markov models, conferring a survival advantage to those who subscribe to the construction of imaginary worlds, supporting professional advancement.

**Table 1**  
**Imaginary Worlds: Modeled Economic Evaluation Studies in the**  
***Journal of Medical Economics* January 2016 to December 2016**

Paper (author)	Target Population or Intervention	Sponsor or Funding(if any)	Modeling Technique and Claims Status	Did the study support the sponsor/funders product?	Claims Assessment and Credibility
Chan et al <sup>34</sup>	Invasive fungal disease in acute myeloid leukemia or myelodysplastic syndrome	No funding declared; one author an employee of Merck	Lifetime decision model of posaconazole versus fluconazole/itraconazole in preventing fungal disease in a Hong Kong hospital. Despite higher cost, posaconazole prophylaxis cost-effective over a lifetime horizon by reduction in costs of treating invasive fungal diseases as costs incurred to avoid one IFD well under Hong Kong GDP per capita.	The study supported the Merck product posaconazole	Claims presented are potentially evaluable. While the authors recommend future research no recommendation for protocols to support claims assessment to meet required standards of normal science.
Wan et al <sup>35</sup>	Nosocomial pneumonia caused by methicillin resistant staphylococcus aureus	Pfizer Investment Co. Ltd	Trial-based cost-effectiveness modeled assessment utilizing bootstrap methods to evaluate success with linezolid versus vancomycin in four Chinese cities. Linezolid dominated vancomycin in 1/3 total cases with vancomycin dominating in only 2% of cases.	Study supported the Pfizer product linezolid	Claims presented are potentially evaluable. While the authors recommend future research no recommendation for protocols to support claims assessment to meet required standards of normal science.
Einarson et al <sup>36</sup>	Long-acting atypical antipsychotics in acutely relapsed schizophrenia	Janssen-Cilag	A 1-year decision sequential therapy relapse/non-response switching model to assess cost-effectiveness of aripiprazole (ARI-LAI), paliperidone (PP-LAI), olanzapine (OLZ-LAI) and risperidone (RIS-LAI). Outcomes: QALYs, hospitalization rates and relapse not requiring hospitalization. Probability sensitivity analysis projected PP-LAI dominated ARI-LAI in 75.8% of 100,000 iterations, RIS-LAI in 83.1% and OLZ-LAI in 95.7%.	Study supported the Janssen-Cilag product paliperidone	Claims presented are potentially evaluable. While the authors recommend future research no recommendation for protocols to support claims assessment to meet required standards of normal science.

Roussel et al <sup>37</sup>	Type 2 diabetes patients failing glycemic control with metformin monotherapy	Novo Nordisk	Application of IMS CORE diabetes model to project lifetime clinical outcomes and direct costs in a French treating environment. Comparisons liraglutide vs sitagliptin and glimepride. Projected ICER of €10,275 per QALY gained vs. sitagliptin and €20,709 per QALY gained vs. glimepride (both under €30,000 threshold).	Study supported the Novo Nordisk product liraglutide	Claims presented are non-evaluable and do not meet required standards of normal science.
Pettigrew et al <sup>38</sup>	Metastatic colorectal cancer	Amgen Canada	A net cost impact analysis to determine total patient cost of panitumumab vs. cetuximab in Canada (5 provinces) in 14-20 week timeframe. Limitations noted: no efficacy or QALY estimates, and health resource utilization from expert opinion. Panitumumab cost-saving in all scenarios.	Study supported the Amgen product panitumumab	Claims presented are potentially evaluable. While the authors recommend future research no recommendation for protocols to support claims assessment to meet required standards of normal science
Moshyk et al <sup>39</sup>	Chronic hepatitis C genotype 3	Bristol-Myers Squibb (BMS)	A Markov lifetime cost-utility model replicating the natural history of hepatitis C and complications. Treatments: daclatasvir (DCV) + sofosbuvir (SOF) vs. DCV + ribavirin (RBV) and pegalated interferon (pINF) + RBV in treatment naïve and treatment experienced populations. Conclusion DCV+SOF safe and effective option and could be considered cost-effective following pINF+RBV treatment.	Study supported the BMS product daclatasvir	Claims presented are non-evaluable and do not meet required standards of normal science.
Qin et al <sup>40</sup>	Overactive bladder with urge urinary incontinence	Pfizer Inc	Decision model to estimate 52-week costs of initiating treatment with fesoterodine vs. no pharmacotherapy in vulnerable elderly population in a hypothetical health plan. OAB related costs included fesoterodine drug acquisition costs, healthcare resource use and OAB related co-morbidities. Estimated cost saving of \$1,616 per patient in US.	Study supported the Pfizer product fesoterodine	Claims presented are potentially evaluable. While the authors recommend future research no recommendation for protocols to support claims assessment to meet required standards of normal science

Roze et al <sup>41</sup>	Type 1 diabetes	Medtronic International Trading	Application of IMS CORE diabetes model to project lifetime cost-effectiveness of sensor-augmented pump therapy (SAP) with a low glucose-suspend (LGS) feature vs. continuous subcutaneous infusion (CSII) plus self-monitoring in a UK treating environment (SMBG). Study concludes that SAP+LGS likely to be cost-effective compared to CSII+SMBG.	Study supported the Medtronic sensor augmented pump (SAP) product	Claims presented are non-evaluable and do not meet required standards of normal science.
Usmani et al <sup>42</sup>	Transplant ineligible patients with newly diagnosed multiple myeloma	Celgene Corporation	Lifetime partitioned survival model with 3 health states progression-free survival, post-progression survival and dead to project lifetime, total direct costs, total life years (LY) and QALYs. Treatment: lenalidomide + dexamethasone (Rd) vs. bortezomib +melphalan + prednisone (VMP) as initial treatment. Incremental cost per QALY and LY gained with Rd vs. VMP estimated \$53,826 and \$35,552 respectively. Rd may be a cost-effective alternative to VMP given oncology willingness-to-pay thresholds.	Study supported the Celgene product lenalidomide	Claims presented are non-evaluable and do not meet required standards of normal science.
Vandewalle et al <sup>43</sup>	Myelofibrosis	Novartis	A discrete state lifetime cohort model to compare impact on overall survival of ruxolitinib versus best available therapy (BAT) in a Portuguese treatment setting. Applying a Weibull parametric survival model a projected discounted (at 5%) median survival for ruxolitinib of 5.7 years vs 3.3 years for BAT. Discounted incremental lifetime healthcare costs €97,052. At willingness to pay of €50,000 per life year probability of ruxolitinib being cost-effective against BAT was > 95%.	Study supported the Novartis product ruxolitinib	Claims presented are non-evaluable and do not meet required standards of normal science.

Whalen et al <sup>44</sup>	Chronic myelogenous leukemia	Bristol Myers-Squibb (BMS)	Lifetime Markov model comparing dasatinib-initiated treatment sequences for patients failing imatinib compared to high dose imatinib and nilotinib. Dasatinib is associated with modeled increased survival and quality of life compared to high dose imatinib and to a smaller extent nilotinib. Respective ICER per QALY \$170,00 and \$160,000.	Study supported the BMS product dasatinib	Claims presented are non-evaluable and do not meet required standards of normal science.
Francois et al <sup>45</sup>	Neurogenic orthostatic hypotension	Lundbeck	A lifetime Markov model to predict the number of falls and treatment responses in neurogenic orthostatic hypotension (nOH) and the predicted impact of droxipoda to treat symptomatic nOH in patients with Parkinson’s disease vs. standard of care. Outcomes: projected number of falls, QALYs and direct costs. Droxipoda was cost-effective vs. standard of care with ICERs of \$47,001/QALY gained, \$24,866 per fall avoided with moderate/major industry and \$1,559 per avoided fall with no/minor injury. Droxipoda a cost-effective option.	Study supported the Lundbeck product droxipoda	Claims presented are non-evaluable and do not meet required standards of normal science
Quon et al <sup>46</sup>	Recurrent venous thromboembolism	Pfizer Canada Inc and Bristol-Myers Squibb (BMS) Canada	A lifetime Markov model following venous thromboembolism (VTE) patients in a Canadian treating environment. Extended treatment with apixaban compared to enoxaparin/warfarin. Modeled impact of apixaban associated with fewer recurrent VTEs, VTE-related deaths and bleeding events and lowest overall cost compared to other NOACs, but at slightly increased cost. Apixaban can offer substantial clinical benefits and is a cost-effective alternative to enoxaparin/warfarin.	Study supported the Pfizer/BMS product apixaban.	Claims presented are non-evaluable and do not meet required standards of normal science

Lopez-Belmonte et al <sup>47</sup>	Herpes zoster vaccination	Sanofi Pasteur-MSD	Lifetime Markov model applied to Spanish population cohort 50 years and over comparing vaccinated and non-vaccinated tracks. Vaccinating 30% of population resulted in QALY gains in range €30,000 to €50,000 range with new vaccine zostavax.	Study supported the Sanofi Pasteur-MSD product zostavax	Claims presented are non-evaluable and do not meet required standards of normal science
Perez-Ruiz et al <sup>48</sup>	Hyperuricemia in gout	Menarini Group	A five-year Markov model of febuxostat vs allopurinol in sequential treatment of hyperuricemia in Spain. Addition of febuxostat to any therapeutic strategy was an efficient option with ICERs ranging from €5,268 to €9,737 and below €30,000/QALY threshold.	Study supported the Menarini Group product febuxostat	Claims presented are potentially evaluable in a shorter timeframe
Beauchemin et al <sup>49</sup>	Advanced breast cancer	Canadian Institutes for Health Research	Application of Global Pharmacoeconomics of Metastatic Breast Cancer (GPMBC) Markov lifetime model of interventions in breast cancer.	Not applicable	Claims presented are non-evaluable and do not meet required standards of normal science
Goeree et al <sup>50</sup>	Second-line advanced squamous non-small cell lung cancer (NSCLC)	Bristol-Myers Squibb	Lifetime progression free survival and Markov models applied to evaluating expected costs, outcomes and incremental cost-utility (QALYs) of docetaxel and erlotinib vs nivolumab for NSCLC in a Canadian treating environment. Nivolumab cost utility estimates were \$152,229 and \$141,836 respectively vs. docetaxel and erlotinib.	Study provided qualified support for BMS product nivolumab	Claims presented are potentially evaluable in a shorter timeframe
Haig et al <sup>51</sup>	Diabetic macular edema	Novartis Pharma AG	Lifetime Markov model to evaluate the impact of ranibizumab (RAN) monotherapy of combination therapy (RAN + laser photocoagulation) versus laser monotherapy for visual impairment due to diabetic macular edema (DME) in a Canadian treating environment. Building on a previous model long-term costs and outcomes followed 3 years of treatment with patients cycling between 8 health states best-corrected visual acuity (BCVA) status. Given a Canadian ICER	Study supported the Novartis AG product ranibizumab	Claims presented are non-evaluable and do not meet required standards of normal science

			threshold of C\$50,000 both RAN monotherapy and combination therapy were cost-effective compared to laser.		
Bruhn et al <sup>52</sup>	Type 2 diabetes	GlaxoSmith Kline (GSK)	Cost-utility assessment with the CORE diabetes model in a 50 year timeframe to assess albiglutide against insulin lispro, insulin glargine and sitagliptin	Study supported the GSK product albiglutide	Claims presented are non-evaluable and do not meet required standards of normal science
Hernandez et al <sup>53</sup>	Relapsing-remitting multiple sclerosis	Biogen	A 10-year Markov cohort cost-utility model to evaluate peginterferon beta-1a vs. interferon beta-1a and glatiramer acetate in a US treating environment. Over 10 years peginterferon beta-1a was dominant when compared with interferon beta-1a and glatiramer.	Study supported the Biogen product peginterferon beta-1a	Claims presented are non-evaluable and do not meet required standards of normal science
Guerin et al <sup>54</sup>	Mitral regurgitation	Abbott Vascular	A 5-year 4-state Markov model to evaluate efficacy of a MitraClip strategy vs. medical management in a French treating environment. Primary endpoint deaths avoided. Over 5 years among a hypothetical cohort of 1000 patients, 276 deaths avoided. Cost per dewath avoided €20,720.	Study supported the Abbott Vascular product MitraClip	Claims presented are potentially evaluable in a shorter timeframe
Su et al <sup>55</sup>	Relapsing-remitting multiple sclerosis	Biogen	A lifetime Markov model with 21 health states to evaluate delayed-release dimethyl fumarate (DMF) as first line treatment in a Canadian treatment environment vs. DMTs glatiramer acetate and rebif 44mcg. On traditional Canadian willingness-to-pay thresholds DMF a cost-effective option compared to other first line DMTs.	Study supported the Biogen product DMF	Claims presented are non-evaluable and do not meet required standards of normal science

Rønborg et al <sup>56</sup>	Dust mite respiratory allergic disease	ALK	Cost minimization comparison between Alutard SQ as standard of care and SQ SLIT-tablet ACARIZAX. Direct and indirect annual cost savings with SQ-SLIT of approximately €1,000.	Study supported the ALK product SQ-SLIT	Claims presented are evaluable and replicable in a Danish treating environment and other countries (recommended as a therapy by Global Initiative for Asthma[GINA])
Roze et al <sup>57</sup>	Type 2 diabetes	Medtronic International Trading	Application of IMS CORE diabetes model to project lifetime cost-effectiveness of continuous subcutaneous insulin infusion (CSII) to achieve improved glycemic control vs. optimization of basal bonus MDI therapy where patients still unable to achieve good glycemic control. ICER estimated at €62,895 per QALY gained and €60,474 per QALY with indirect costs included. As ICER below willingness-to-pay threshold of €80,000, CSII likely to represent good 'value for money' in poorly controlled type 2 diabetes patients.	Study supported the Medtronic product CSII	Claims presented are non-evaluable and do not meet required standards of normal science
Lee et al <sup>58</sup>	Hepatocellular carcinoma	Bayer Healthcare	Decision model driven cost comparison of Gd-EOB-DTPA MRI vs., ECCM-MRI and multi-detector computed tomography detection and characterization of liver lesions in Thai and Korean treating environments from payer and hospital perspectives. Input data estimated and validated by experts. Costs of diagnostic alternatives and related treatment options from literature.. Conclusion: Gd-EOB-DTPA-MRI gives better diagnostic certainty and cost savings.	Study supported the Bayer contrast agent product Gd-EOB-DTPA MRI	Claims presented are potentially evaluable and replicable in other treating environments and countries



Labotesis et al <sup>59</sup>	Acute ischemic stroke	Medtronic	A lifetime Markov model to evaluate cost-effectiveness in neurothrombectomy of stent-retriever thrombectomy in combination with intravenous tissue-type plasminogen activator (IV t-PA) vs. IV t-PA alone. Combination therapy associated with improved quality of life, overall cost-savings with higher treatment costs offset by long-term cost savings. In a UK treatment environment net benefit of £79,402. Conclusion: combined therapy cost-effective in acute ischemic stroke.	Study supported the Medtronic stent retriever product Solitaire	Claims presented are non-evaluable and do not meet required standards of normal science
Saab et al <sup>60</sup>	Genotypes 1 or 4 hepatitis C	AbbVie	A lifetime Markov model of chronic hepatitis C for both treatment naïve and experienced patients evaluated the cost-effectiveness of direct-acting antiviral therapies in a US treating environment. Outcomes modeled included rates of compensated/decompensated cirrhosis, hepatocellular carcinoma, liver related deaths, total costs, life-years and quality adjusted life years. Among the currently recommended treatments ombitasvir/paritaprevir/ritonavir + dasabuvir +/- ribavirin (3d+/-R) for GT1 and ombitasvir/paritaprevir/ritonavir + ribavirin (2D+R) for GT4 had a favorable cost-effectiveness profile.	Study supported the AbbVie combination therapy products	Claims presented are non-evaluable and do not meet required standards of normal science
Bianic et al <sup>61</sup>	Liver transplants	Novartis Pharma	A lifetime Markov model with flexibility to allow for results over shorter time scales to compare cost-effectiveness of everolimus (EVR) plus reduced tacrolimus (TAC) vs. standard dose TAC in an Italian treating environment. Modeled ICERs favored the cost-effective combination EVR+TAC for both survival and quality of life.	Study supported the Novartis product everolimus	Claims presented are potentially evaluable in the short-term. Evaluable claims not presented. No recommendation for protocols to support claims assessment to meet required standards of normal science

Einarson et al <sup>62</sup>	Chronic schizophrenia	Janssen Pharmaceuticals	A 1-year decision model adapted to a Portuguese environment to compare paliperidone palmitate (PP-LAI) to risperidone (RIS-LAI), haloperidol (HAL-LAI) and olanzapine (OLZ). Clinical and cost data from the literature. Outcomes: INCERs for copst-utility and cost-effectiveness assessments. Results: PP-LAI dominated HAL-LAI and RIS-LAI and cost-effective over OLZ against NICE/Portuguese willingness-to-pay thresholds.		Claims presented are potentially evaluable in the short-term. Evaluable claims not presented. No recommendation for protocols to support claims assessment to meet required standards of normal science
Hurry et al <sup>63</sup>	Anaplastic lymphoma kinase positive (ALK+) non-small cell lung cancer	Novartis Pharmaceuticals	A 4-year partitioned survival model with three health states for cost-effectiveness of ceritinib vs pemetrexed, best supportive care and historical control (HC) in a Canadian treating environment for patients previously treated with crizotinib. Over 4 years ceritinib associated with gain 0.86 QALYs and total direct costs of \$89,740. ICER vs BSC \$149,117, \$80,100 vs pemetrexed and \$104,436 vs. HC. Ceritinib may be considered a cost-effective option.	Study supported Novartis product ceritinib. Crizotinib also a Novartis product.	Claims presented are potentially evaluable in the short-term. Evaluable claims presented over 4-years.
Johnson et al <sup>64</sup>	Genotype 1 hepatitis c virus infection	AbbieVie	A 70-year Markov model with 1-year cycle and 6 health states. Modeled comparison ombitasvir/paritaprevir/ritonavir and dasabuvir+/-ribavarin (OMB/PTRV/r + DSV+/-RBV) versus regimens including pegylated interferon for treatment naïve/experienced in a UK treatment setting. Comparator regimens other than PegINF/RBV where ICERs per QALY were £13,864 and £10,258 respectively. Conclusion: product a cost-effective oral treatment compared to standard treatment regimens in the 70-year timeframe.	Study supported AbbieVie combination therapy products	Claims presented are non-evaluable and do not meet required standards of normal science

Muduma et al <sup>65</sup>	Liver transplant	Astellas Pharma Europe Limited	A 25-year Markov model to evaluate cost-utility of immunosuppressive regimens prolonged release (PR) tacrolimus, immediate release (IR) tacrolimus and ciclosporin in liver transplants in a UK treating environment. PR tacrolimus resulted in increased life expectancy and quality adjusted life expectancy relative to IR tacrolimus and ciclosporin. Overall, ICER of £1,889 per QALY gained.	Study supported Astellas product tacrolimus PR	Claims presented are non-evaluable and do not meet required standards of normal science
Solon et al <sup>66</sup>	Characterization of diminutive colon polyps	Olympus Europa	Decision model with 7-year time horizon to model a cost-consequences and budget impact analysis of narrow band imaging (NBI) vs. white light endoscopy (WLE) in a UK treating environment. NBI offered cost-savings within minimal impact on health outcomes and associated adverse events with annual cost savings of £141,192,057.	Study supported Olympus NBI product	Potential to reformulate claims for cost savings to make them evaluable in a shorter time frame. As stated do not meet required standards of normal science.
Xuan et al <sup>67</sup>	Newly diagnosed duodenal ulcer	Livzon Pharmaceuticals Group Inc	A literature-based one-year decision tree model to assess cost-effectiveness of 10mg once daily ilaprazole vs. once daily 20mg omeprazole in a China treatment environment. Ilaprazole was cost-effective with an ICER of ¥132,056 which is within the WHO threshold of 3 times per capita GDP.	Study supported Livzon product ilaprazole	Claims presented are potentially evaluable given timeframe of study
Jakubowiak et al <sup>68</sup>	Relapsed multiple myeloma	Amgen Inc	A lifetime partitioned survival model (the Kyprolis Global Economic Model-K-GEM) to assess the cost-effectiveness of carfilzomib-lenalidomide-dexamethasone (KRd) vs. lenalidomide-dexamethasone (Rd). KRd was estimated to be more effective than Rd in QALY gains and with an ICER of \$107,520 per QALY. Assuming a willingness-to-pay threshold of \$150,000 KRd may represent an efficient allocation of resources.	Study supported Amgen product carfilzomib	Claims presented are non-evaluable and do not meet required standards of normal science

Chuang et al <sup>69</sup>	Type 2 diabetes	AstraZeneca Pharmaceuticals	Application of the Cardiff Diabetes Model with a 40-year time horizon to evaluate cost-effectiveness of exenatide (EQW) vs. dulaglutide, liraglutide and lixisenatide in patients not adequately controlled on metformin. QALY gains ranged from 0.043 to 0.102 against comparators. EQW dominated liraglutide. Cost per QALY gained £596, £1,004 and respectively. Probability EQW cost effective ranged from 76-99%.	Study supported AstraZeneca product exenatide	Claims presented are non-evaluable and do not meet required standards of normal science
Wielage et al <sup>70</sup>	Overactive bladder	Astellas Pharma Global Development	Markov state-transition model with 3-year time horizon to assess mirabegron vs. six antimuscarinics in US Medicare Advantage and commercial perspectives. Primary outcome cost per QALY. Costs per QALY respectively \$66,347 and \$59,690. Product judged cost-effective.	Study supported AstraZeneca product exenatide	Potential to reformulate claims to make them evaluable in a shorter time frame. As stated do not meet required standards of normal science.
Virabhak et al <sup>71</sup>	Chronic hepatitis C virus genotype 1b	AbbVie	Lifetime Markov state transition model to compare ombitasvir/paritapovir/ritonavir (OBV/PTV/r) vs. daclatasvir+asunaprevir (DVC/ASV) and no treatment in patients without cirrhosis and OBV/PTV/r vs. DCV/ASV and sofosbuvir/ledipasvir (SOF/LDV) in Y93H mutation negative patients with and without cirrhosis in treatment naïve/experienced in a Japanese treatment environment. Results: OBV/PTV/r appears to be a cost-effective treatment against DCV/ASV and dominates no treatment without cirrhosis	Study supported AbbVie combination therapy products	Claims presented are non-evaluable and do not meet required standards of normal science

Gordon et al <sup>72</sup>	Type 2 diabetes	AstraZeneca	Application of the Cardiff Diabetes Model with a 40 year time horizon to evaluate cost effectiveness of exenatide BID vs. bolus insulin lispro (TID) as add on when glycemic control sub-optimal with titrated basal insulin glargine and metformin from a Swedish healthcare perspective. Results: exenatide BID associated with incremental cost of €1,270 and QALY increase of 0.64 compared to lispro TID over 40 years. Cost per QALY gained €1,971.	Study supported AstraZeneca product exenatide	Claims presented are non-evaluable and do not meet required standards of normal science
Goeree & Goeree <sup>73</sup>	Opioid induced constipation	Purdue Pharma	One-year cost-utility decision model comparing combination oxycodone with naloxone versus oxycodone in managing moderate-to-severe pain in Canadian patients with opioid-induced constipation. Costs offsets to analgesic costs resulted in 1-year cost utility of naloxone ranging from C\$2178-C\$7,732 per QALY gained in base case analysis.	Study supported Purdue Pharma product Targin	Claims presented are potentially evaluable although no protocol presented for assessment in Canadian treating environment

References

<sup>1</sup> Langley PC, Rhee T. 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

<sup>2</sup> Langley PC. Imaginary Worlds: Modeled claims for cost-effectiveness published in *PharmacoEconomics* January 2015 to December 2015. *Inov Pharm.* 2016;7(2): No. 9

<sup>3</sup> Langley PC, Rhee T. Imaginary Worlds: The Status of Modeled Economic Evaluation Claims Published in *Value in Health* January 2015 to December 2015. *Inov Pharm.* 2016;7(2): No. 18

<sup>4</sup> Langley PC. Sunlit Uplands: The Genius of the NICE Reference Case. *Inov Pharm.* 2016;7(2): No. 12

<sup>5</sup> 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. 5.

<sup>6</sup> Langley PC. Intelligent Design and Imaginary Worlds in Cost-Effectiveness Claims: An Overview of Commentaries in INNOVATIONS in Pharmacy from July 2016 to February 2017. *Inov Pharm.* 2017;8(1): No. 22

<sup>7</sup> Langley PC. Modeling Imaginary Worlds: Version 4 of the AMCP Format for Formulary Submissions. *Inov Pharm.* 2016;7(2): No. 11

<sup>8</sup>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. 5

- <sup>9</sup> 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): No.12
- <sup>10</sup> Langley PC. Imaginary Worlds and the Institute for Clinical and Economic Review (ICER) Evidence Report: Targeted Immune Modulators for Rheumatoid Arthritis. *Inov Pharm*. 2017;8(2): No.10
- <sup>11</sup> Pigliucci M. Nonsense on stilts: How to tell science from bunk. Chicago: University of Chicago Press., 2010
- <sup>12</sup> Langley P. The status of modeled claims (Editorial). *J Med Econ*. 2015;18(12):991-92
- <sup>13</sup> Langley P. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ*. 2015;18(12):993-99
- <sup>14</sup> CADTH. Guidelines for the economic evaluation of health technologies: Canada. 4<sup>th</sup> ed. Ottawa: CADTH, 2017 Mar
- <sup>15</sup> Langley PC. True North: Building Imaginary Worlds with the Revised Canadian (CADTH) Guidelines for Health Technology Assessment. *Inov Pharm*. 2017;8(2): No. 9
- <sup>16</sup> Langley PC, Rhee T. The Imaginary Worlds of ISPOR: Modeled Cost-Effectiveness Claims Published in Value in Health from January 2016 to December 2016. *Inov Pharm*. 2017;8(2): No. 14
- <sup>17</sup> Moher D, Shamseer L, Clarke E et al. Preferred reporting items for systematic reviews and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*. 2015;4:1 DOI: 10.1186/2046-4053-4-1
- <sup>18</sup> Einaron T, Bereza B, Nielsen A et al. Systematic review of models used in economic analyses in moderate-to-severe asthma and COPD. *J Med Econ*. 2016;19(4):319-55
- <sup>19</sup> Kassirer J, Angell M. The Journal's policy on cost-effectiveness analysis. *New Eng J Med*. 1994;331:669-700
- <sup>20</sup> Buckwalter J, Tolo V, O'Keefe R. How do you know it is true? Integrity in research and publications: AOA critical issues. *J Bone Joint Surg Am*. 2015; 97(1):e2(1-8)
- <sup>21</sup> Marusic A, Wager E, Utrobicic A et al. Interventions to prevent misconduct and promote integrity in research and publication. *Cochrane Database Syst Rev*. 2016;4:MR000038.
- <sup>22</sup> Baker M, 1,500 scientists lift the lid on reproducibility *Nature News Feature*. 25 May 2016. <http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>
- <sup>23</sup> Flier J. Irreproducibility of published bioscience research: Diagnosis, pathogenesis and therapy. *Molecular Metabolism*. 2016;6:2-9
- <sup>24</sup> Ioannidis J. Why most published research findings are false. *Plos Medicine*. 2005;2(8):e124
- <sup>25</sup> Vallejo-Torres L, García-Lorenzo B, Castilla I et al. On the estimation of the cost-effectiveness threshold: Why, What, How? *Value Health*. 2016;19(10): 558-66
- <sup>26</sup> 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
- <sup>27</sup> Peake N, Gormenghast. London: Eyre & Spottiswoode, 1950
- <sup>28</sup> Lewis C. Narnia. The Lion, the Witch and the Wardrobe. London: G Bles, 1950
- <sup>29</sup> 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

- <sup>30</sup> 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
- <sup>31</sup> Langley PC. Nullius in verba: Version 2.0 of the University of Minnesota, School of Social and Administrative Pharmacy Program, Proposed Guidelines for Formulary Evaluation. *Inov Pharm*. 2016;7(4):No. 16
- <sup>32</sup> Kazi DS, Moran A, Coxson P et al. Cost-effectiveness of PCSK9 inhibitor therapy in patients with heterozygous familial hypercholesterolemia or atherosclerotic cardiovascular disease. *JAMA*. 2016;316(7):743-753
- <sup>33</sup> Keats J. *The Eve of St Agnes* in Dawkins R. *Unweaving the Rainbow*. New York: Houghton Mifflin, 1998
- <sup>34</sup> Chan S, Marcella S, Gill H et al. Posaconazole vs fluconazole or itraconazole for prevention of invasive fungal diseases in patients with acute myeloid leukemia or myelodysplastic syndrome: a cost-effectiveness analysis in an Asian teaching hospital. *J Med Econ*. 2016;19(1):77-83
- <sup>35</sup> Wan Y, Li Q, Chen Y et al. Economic evaluation among Chinese patients with nosocomial pneumonia caused by methicillin-resistant staphylococcus aureus and treated with linezolid or vancomycin: a secondary, post-hoc analysis based on a Phase 4 clinical trial study. *J Med Econ*. 2016;19(1):53-62
- <sup>36</sup> Einarson T, Pudas H, Goswami P et al. Pharmacoeconomics of long-acting atypical antipsychotics for acutely relapsed chronic schizophrenia in Finland. *J Med Econ*. 2016;19(2):121-130
- <sup>37</sup> Roussel R, Martinez L, Vandebrouck T et al. Evaluation of the long-term cost-effectiveness of liraglutide therapy for patients with type 2 diabetes in France. *J Med Econ*. 2016;19(2):131-44
- <sup>38</sup> Pettigrew M, Kavan P, Surprenant et al. Comparative net cost impact of the utilization of panitumumab versus cetuximab colorectal cancer in Canada. *J Med Econ*. 2016;19(2):145-57
- <sup>39</sup> Moshyk A, Martel M-J, Tahami A et al. Cost-effectiveness of daclatasvir plus sofosbuvir-based regimen for treatment of hepatitis C virus genotype 3 infection in Canada. *J Med Econ*. 2016;19(2):191-202
- <sup>40</sup> Qin L, Luo X, Zou K et al. Economic impact of using fesoterodine for the treatment of overactive bladder with urge urinary incontinence in a vulnerable elderly population in the United states. *J Med Econ*. 2016;19(3):229-35
- <sup>41</sup> Roze S, Smith-Palmer J, Valentine W et al. Long-term economic benefits of sensor-augmented pump therapy vs continuous subcutaneous insulin infusion alone in type 1 diabetes: a UK perspective. *J Med Econ*. 2016;19(3):236-42
- <sup>42</sup> Usmani S, Cavenagh J, Belch A et al. Cost-effectiveness of lenalidomide plus dexamethasone vs bortezomib plus melphalan and prednisone in transplant-ineligible US patients with newly-diagnosed multiple myeloma. *J Med Econ*. 2016;19(3):243-58
- <sup>43</sup> Vandewalle B, Andreozzi V, Almeida J et al Pharmacoeconomics of ruxolitinib therapy in patients with myelofibrosis. *J Med Econ*. 2016;19(4):424-31
- <sup>44</sup> Whalen J, Stillman I, Ambavane A et al. Cost-effectiveness analysis of second-line tyrosine kinase inhibitor treatment for chronic myelogenous leukemia. *J Med Econ*. 2016;19(5):445-61
- <sup>45</sup> Francois C, Hauser R, Aballéa et al. Cost-effectiveness of droxidopa in patients with neurogenic orthostatic hypotension: post-hoc economic analysis of Phase 3 clinical trial data. *J Med Econ*. 2016;19(5):515-25
- <sup>46</sup> Quon P, Le H, Raymond V et al. Clinical and economic benefits of extended treatment with apixaban for the treatment and prevention of recurrent venous thromboembolism in Canada. *J Med Econ*. 2016;19(6):557-67
- <sup>47</sup> Lopez-Belmonte J, Cisterna R, de Miguel A et al. The use of Zostavax in Spain: the economic case for vaccination of individuals aged 50 years and older. *J Med Econ*. 2016;19(6):576-86

- <sup>48</sup> Perez-Ruiz F, Díaz-Tomé C, Carcedo D. Cost-effectiveness analysis of febuxostat in patients with gout in Spain. *J Med Econ.* 2016;19(6):604-10
- <sup>49</sup> Beaucheminb C, Letarte N, Mathurin K et al. A global economic model to assess the cost-effectiveness of new treatments for advanced breast cancer in Canada. *J Med Econ.* 2016;19(6):619-629
- <sup>50</sup> Goeree R, Villeneuve J, Goeree J et al. Economic evaluation of nivolumab for the treatment of second-line advanced squamous NSCLC in Canada: a comparison and extrapolate survival outcomes. *J Med Econ.* 2016;19(6):630-44
- <sup>51</sup> Haig J, Barbeau M, Ferreira A. Cost-effectiveness of ranibizumab in the treatment of visual impairment due to diabetic macular edema. *J Med Econ.* 2016;19(7):663-71
- <sup>52</sup> Bruhn D, Martin A, Tavares R et al. Cost-utility of albiglutide versus insulin lispro, insulin glargine, and sitagliptin for the treatment of type 2 diabetes in the US. *J Med Econ.* 2016;19(7):672-83
- <sup>53</sup> Hernandez L, Guo S, Kinter E et al. Cost-effectiveness analysis of perinterferon beta-1a compared with interferon beta-1a and glatiramer acetate in the treatment of relapsing-remitting multiple sclerosis in the United States. *J Med Econ.* 2016;19(7):684-95
- <sup>54</sup> Guerin P, Bourguignon S, Jamet N et al. MitraClip therapy in mitral regurgitation: a Markov model for the cost-effectiveness of a new therapeutic option. *J Med Econ.* 2016;19(7):696-701
- <sup>55</sup> Su W, Kansal A, Vicente C et al. The cost-effectiveness of delayed-release dimethyl fumarate for the treatment of relapsing-remitting multiple sclerosis in Canada. *J Med Econ.* 2016;19(7):718-27
- <sup>56</sup> Rønborg S, Johnsen C, Theilgaard S et al. Cost-minimization analysis of sublingual immunotherapy versus subcutaneous immunotherapy for house dust mite respiratory allergic disease in Denmark. *J Med Econ.* 2016;19(8):735-41
- <sup>57</sup> Roze S, Duteil E, Smith-Palmer J. Cost-effectiveness of continuous subcutaneous insulin infusion in people with type 2 diabetes in the Netherlands. *J Med Econ.* 2016;19(8):742-49
- <sup>58</sup> Lee J-M, Kim M-J, Phongkitkarun S et al. Health economic evaluation of Gd-EOB-DTPA MRI vs. ECCM-MRI and multi-detector computed tomography in patients with hepatocellular carcinoma in Thailand and South Korea. *J Med Econ.* 2016;19(8):759-68
- <sup>59</sup> Lobotesis KJ, Veltkamp R, Carpenter I et al. Cost-effectiveness of stent-retriever thrombectomy with IV t-PA alone for acute ischemic stroke in the UK. *J Med Econ.* 2016;19(8):785-94
- <sup>60</sup> Saab S, Parisé H, Virabhak S et al. Cost-effectiveness of currently recommended direct-acting antiviral treatments in patients infected with genotypes 1 or 4 hepatitis C virus in the US. *J Med Econ.* 2016;19(8):795-805
- <sup>61</sup> Bianic F, Campbell R, Damera V et al. Cost-effectiveness of everolimus plus reduced tacrolimus in de novo-recipients in the Italian setting. *J Med Econ.* 2016;19(9):866-73
- <sup>62</sup> Einarson T, Maia-Lopes S, Goswami P et al. Economic Analysis of paliperidone long-acting injectable for chronic schizophrenia in Portugal. *J Med Econ.* 2016;19(9):913-21
- <sup>63</sup> Hurry M, Zhou Z-Y, Zhang J et al. Cost-effectiveness of ceritinib in patients previously treated with crizotinib in anaplastic lymphoma kinase positive (ALK+) non-small cell lung cancer in Canada. *J Med Econ.* 2016;19(10):936-44
- <sup>64</sup> Johnson S, Parisé H, Virabhak S et al. Economic evaluation of ombitasvir/paritaprevir/ritonavir and dasabuvir for the treatment of chronic genotype 1 hepatitis c virus infection. *J Med Econ.* 2016;19(10):983-94
- <sup>65</sup> Muduma G, Odeyemi I, Pollock R et al. A cost-utility analysis of prolonged-release tacrolimus relative to immediate-release tacrolimus and ciclosporin in liver transplant recipients in the UK. *J Med Econ.* 2016;19(10):995-1002



- <sup>66</sup> Solon C, Klausnitzer R, Blissett D et al. Economic value of narrow band imaging versus white light endoscopy for the characterization of diminutive polyps in the colon: systematic literature review and cost-consequence model. *J Med Econ.* 2016;19(11):1040-48
- <sup>67</sup> Xuan J, Song R, Xu G et al. Modeling the cost-effectiveness of ilaprazole versus omeprazole for the treatment of newly diagnosed duodenal ulcer patients in China. *J Med Econ.* 2016;19(11):1056-60
- <sup>68</sup> Jakubowiak A, Campioni M, Benedict Á et al. Cost-effectiveness of adding carfilzomib to lenalidomide and dexamethasone in relapsed multiple myeloma from a US perspective. *J Med Econ.* 2016;19(11):1061-74
- <sup>69</sup> Chuang L, Verheggen B, Charokopou M et al. Cost-effectiveness analysis of exenatide once-weekly versus dulaglutide, liraglutide, and lixisenatide for the treatment of type 2 diabetes mellitus: an analysis from the UK NHS perspective. *J Med Econ.* 2016;19(12):1127-34
- <sup>70</sup> Wielage R, Perk S, Campbell N et al. Mirabegron for the treatment of overactive bladder: cost-effectiveness from US commercial health-plan and Medicare Advantage perspectives. *J Med Econ.* 2016;19(12):1135-43
- <sup>71</sup> Virabhak S, Yasui K, Yamazaki K et al. Cost-effectiveness of direct-acting antiviral regimen ombitasvir/paritaprevir/ritonavir in treatment-naïve and treatment-experienced patients infected with chronic hepatitis C virus genotype 1b in Japan. *J Med Econ.* 2016;19(12):1144-56
- <sup>72</sup> Gordon J, McEwan P, Sabale U et al. The cost-effectiveness of exenatide twice daily (BID) vs insulin lispro three times daily (TID) as add-on therapy to titrated insulin glargine in patients with type 2 diabetes. *J Med Econ.* 2016;19(12):1167-74
- <sup>73</sup> Goeree R, Goeree J. et al. Cost-effectiveness analysis of oxycodone with naloxone versus oxycodone for the management of moderate-to-severe pain in patients with opioid-induced constipation in Canada. *J Med Econ.* 2016;19(3):277-91