

5-19-2017

The Imaginary Worlds of ISPOR: Modeled Cost-Effectiveness Claims Published in Value in Health from January 2016 to December 2016

Paul C. Langley

University of Minnesota, langley@maimonresearch.com

Taeho Greg Rhee

University of Minnesota, Twin Cities, rhee0041@umn.edu

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

Recommended Citation

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): Article 14. <http://pubs.lib.umn.edu/innovations/vol8/iss2/14>

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

The Imaginary Worlds of ISPOR: Modeled Cost-Effectiveness Claims Published in *Value in Health* from January 2016 to December 2016

Paul C Langley, PhD and Taeho Greg Rhee, PhD
College of Pharmacy, University of Minnesota

Abstract

In 2016, a review of modeled cost-effectiveness studies published in *Value in Health* between January 2015 and December 2015 was presented. The purpose of the review was to consider whether these modeled claims for cost-effectiveness met the standards of normal science: were the claims made credible, evaluable and replicable? The review concluded that none of the 16 studies assessed met this standard. They should be seen as thought experiments; the construction of imaginary worlds which should be categorized as pseudoscience. The reader, or health care decision maker, would have had no idea, and would never know, whether the claims were right, wrong or misleading. Similar reviews were undertaken in *Pharmacoeconomics* and the *Journal of Medical Economics* and came to the same conclusion. The purpose of this second review is to consider the modeled claims published in *Value in Health* between January 2016 and December 2016, applying the same criteria. Unfortunately, for those who subscribe to the standards of normal science, we must come to the same conclusion. Of the 13 economic evaluations reviewed, 12 simulated claims that were immune to failure. The model structures ensured that the claims were neither evaluable nor replicable. They were categorized as pseudoscience; they failed to meet the standards of normal science. Five of these studies were supported by manufacturers and all supported the manufacturer's product. Three systematic reviews were also evaluated. Once again, there was a failure to consider meeting the standards of normal science in presenting modeled claims for cost-effectiveness.

Keywords: ISPOR, Markov, modeled claims, imaginary worlds, pseudoscience, red flag

Introduction

In mid-2016, a commentary was published in *INNOVATIONS in Pharmacy* reviewing, from the perspective of the standards of normal science, modeled technology assessment claims published in *Value in Health* from January 2015 to December 2015¹. This systematic review concluded that of the 16 identified papers, 14 presented a cost-per-QALY analysis, with 9 presenting their claims in a lifetime cost-per-QALY framework. The technology assessments presented, while conforming to ISPOR recommended standards, failed to meet the standards of normal science: the claims were neither credible, nor were they evaluable and replicable. They were best seen as imaginary claims created by imaginary modeled worlds. Recipients of these claims had no idea whether they were right or whether they were wrong, and they would never know as the claims were immune to failure. Reviews of modeled studies published in *Pharmacoeconomics* and the *Journal of Medical Economics* over the same time period came to the same conclusion^{2 3}.

These reviews of cost-effectiveness models are part of a series of commentaries published in *INNOVATIONS in Pharmacy* over the last 12 months that have focused on the evidentiary standards for claims assessment. These standards are 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)^{4 5 6 7 8 9 10}. The common theme in these commentaries has been that the standards proposed and accepted in health technology assessment in the construction of non-evaluable modeled claims for pharmaceutical products and devices should be seen as pseudoscience (a.k.a. pure bunk); as intelligent design rather than natural selection^{11 12}.

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 been applied in 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 their technology assessment framework is not to be judged by the standards of normal science¹³. The guidelines are designed to set criteria for the construction of imaginary simulations to support cost-outcomes claims, to

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
Web: www.maimonresearch.net

'inform' health system decision makers, not to test hypotheses. Supporting claims that are credible, evaluable and replicable are put to one side by CADTH in favor of the construction of imaginary worlds. Typically an imaginary simulation of benefits and harms, focused on health related quality of life (HRQoL) over the lifetime of a hypothetical target population.

The purpose of this second review of health technology assessment studies published in *Value in Health* 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 to December 2016. It will be followed by reviews over the same time period for *Pharmacoeconomics* and the *Journal of Medical Economics*.

Methods

A systematic review, following the PRISMA-P checklist (MeSH terms 'cost', 'cost effectiveness', 'Markov', 'QALY') of all papers published in *Value in Health* in 2016 was undertaken¹⁴. In order to judge whether the modeled claims presented met the standards of normal science four questions were considered:

- Is the model capable of generating testable claims?
- Did the author(s) attempt to generate testable claims?
- Did the authors suggest how the claims might be evaluated?
- Did the author(s) 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 formulary 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 anticipated 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) such as health related quality of life (HRQoL) and quality adjusted life years (QALYs).

In judging whether or not a model might support testable (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, 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.

Apart from evaluating the published economic evaluations, attention was also given to systematic reviews published in *Value in Health*. Two questions are 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; and
- Did the systematic review recommend (or caution against) accepting the claims from the modeled economic evaluations as the basis for formulary decisions

Results

A total of 13 papers classified by *Value in Health* as either economic evaluations or comparative effectiveness research were identified (Table 1). As well, three systematic reviews were identified as these were directly related to the issues raised in this review (Table 2).

Modeled Cost-Effectiveness Claims

Key findings are:

- Only one of the studies provided comparative claims that met the standards of normal science¹⁶
- None of the claims presented in other 12 papers were credible, evaluable and replicable
- Modeled time horizons for the claims presented in these 12 papers ensured the claims were immune to failure
- Five of these economic evaluations were funded directly by pharmaceutical manufacturers and all simulated claims that supported the manufacturer's product^{17 18 20 24 25}

- Two more of the economic evaluations had manufacturer links and both supported the manufacturer's product^{19 27}
- None of the papers considered how the claims might be evaluated in treatment practice

Modeled Claims and Systematic Reviews

In terms of the questions raised, none of the reviews addressed the issue of the credibility, evaluation and replication of modeled claims neither did they explicitly recommend or caution against accepting the modeled claims as the basis for formulary submissions. The only qualification here is from the Brilleman et al study which, in addressing claims for cost-effectiveness modeled on clinical trials, points to the lack of standards potentially impacting model-based claims for cost-effectiveness and the need for reporting on cost-allocation²⁸. A key point to note in the Brilleman et al study is that: 'None of the articles included in our review used statistical methods to adjust for non-adherence or incorporate non-adherence information directly into the economic evaluation' (p. 103).

Discussion

As noted above, the common theme in the commentaries published in *INNOVATIONS in Pharmacy* since mid-2016 is that the recommended standards and their application in health technology assessments fail to meet the standards of normal science⁴. Rather than focusing on generating testable hypotheses to assess anticipated comparative product performance, the studies fall back on creating modeled imaginary worlds in which the comparative claims are immune to failure. The exemplars here are the lifetime cost-per-QALY willingness-to-pay 'reference case' model structures, which support 'information only' formulary submissions where the claims are non-evaluable⁶. This review of economic evaluations in *Value in Health* for calendar 2016 demonstrates that, despite criticisms raised against the construction of simulated imaginary worlds and their characterization as 'pseudoscience', this acceptance of imaginary constructs continues.

Consider the question, for example, of cost-per-QALY estimates and willingness-to-pay thresholds. Imaginary model simulations that yield lifetime claims that propose that market entry prices of a product are not cost-effective are, as noted in reviews of ICER evidence reports, immune to failure^{8 9 10}. As such they fail the standards of normal science and should be rejected.

Modeled Claims

Once a study design commits to a long-term or lifetime Markov (or similar) model, there is no chance that the evaluation will generate claims that are credible and evaluable. Irrespective of the technical appeal in constructing multi-health state treatment pathway models that may extend for decades into

an unknown future, with assumptions justified to appeals to literature or short term RCTs, the reader has to take the claims at face value. The Gregory et al paper, to give one example, projects on a yearly cycle from age 10 years to death or 100 years in the risks from computed tomography in Ireland¹⁵. Other models in *Value in Health* have lifetime horizons, horizons of 25 years and horizons of 20 years.

Certainly, if we are prepared to suspend our disbelief in these constructs, we could point to a number of ways in which the 'realism' of the imaginary world could be enhanced. These have been detailed in previous commentaries and could include assumptions capturing anticipated patterns of persistence and adherence, presence and impact of comorbidities, anticipated pricing policies for annualized whole sale acquisition cost (WAC) price increases, possible WAC discounting to target patient groups, entry of pipeline competitor products and assumptions as to unknown but possibly defensible therapy switching patterns following initial response. Unfortunately, 'improving' the appeal of an imaginary construct merely increases the number of competing scenarios. Rather than providing 'more information' to support decision making, the more likely result is information overload and the rejection of the model. Consider, for example the issues of adherence and persistence. These are typically overlooked (or ignored) in constructing lifetime models. This is an odd decision given that there is now ample evidence for limited persistence with the majority of patients in disease areas abandoning a therapy within two to three years of an index prescription. If claims for persistence (let alone adherence) are based upon RCTs then as Brilleman et al point out, the evidence base may be somewhat questionable²⁸.

The belief in the credibility of modeled claims may be further tested by manufacturer sponsored models, which readers might regard as 'marketing exercises'. In this review, five of the economic evaluations were sponsored or funded by manufacturers. As noted, all of these supported non-evaluable claims for the manufacturer's product. The absence of an evaluable claims means that there is no basis for assessing the 'validity' of the model claims. Competing models sponsored by manufacturers in the same class of comparator products might, not surprisingly, come to quite different conclusions. If journals such as *Value in Health* are prepared to publish manufacturer support models the, at least, there could be a 'red flag' annotation with the peer review process assessing (and reporting) on the model presented in the context of other models in the therapy area. Even so, we will still have competing claims where the authors of the systematic review will attempt to explain that variation in terms of model structures, assumptions and presumptions of willingness-to-pay; an exercise which is pointless in the absence of modeled credible and evaluable claims with possible feedback from target populations.

Systematic Reviews

Even if the authors of imaginary worlds claim to be adherent to a 'reference case' mandated by a technology assessment group, there is ample opportunity to propose alternative model structures and input assumptions to make the case for comparative efficacy or effectiveness. It might be argued that subjecting a submitted model to the ministrations of an evidence review group would resolve this issue. Unfortunately, it does not solve the problem. Blinded assessments by competing review groups could come to quite different conclusions as to the 'preferred' model structure and the 'appropriate' input assumptions.

The Kirsch review of Markov modeling in disease management programs, following the ISPOR good practice guidelines for decision-analytic modeling, reviewed 16 studies³⁰. In chronic heart disease the results ranged from cost savings of \$657 and an increase of 0.0051 QALYs to an increase in costs of \$4,607 per life year gained (LYG) and \$146,544 per QALY. The two asthma studies also yielded cost savings or additional costs, with the five diabetes studies reported again both cost savings and additional costs. A detailed comparison of the various models was presented pointing to varying quality ('far from perfect') and coverage in terms of, for example, model structure, model cycle length, time horizon, utility weights, costs and parameter sensitivity. The author found it difficult to determine whether funders tried to influence study results or prevent the publication of unfavorable results.

Kirsch concluded that if the problems identified in the review are addressed then 'Markov models should be more suitable to evaluate economic effects of multicomponent interventions, and provide helpful information for decision makers' (p. 1052)³⁰. Apart from the question of what is considered 'helpful' as opposed to possibly misleading information, this conclusion misses the point. At no stage in the review is the issue of establishing credible, evaluable and replicable claims for disease management programs considered. Even if model builders 'improved' the quality of the Markov model decision makers would be no further ahead. They would still be faced with imaginary claims and the possibility of an endless procession of diverse claims. We are still in the CADTH-mandated realm of pseudoscience, hoping that our non-evaluable claims will 'inform' decision makers.

The same objections apply to the Nunes et al review of long-term mechanical circulatory support²⁹. Following a literature search a total of 11 country-specific cost-effectiveness analyses were identified. Once again the object is seen to be to 'inform' decision makers 'such that societal benefit is maximized'. The focus is on non-evaluable ICER thresholds rather than on presenting claims for devices that are credible, evaluable and replicable. The non-evaluable modeled results, not unexpectedly, vary widely. In mechanical circulatory support as a bridge to transplantation. ICERs between this

support and medical management ranged from C\$85,025 and C\$200,166 per QALY and for destination between C\$87,622 and C\$1,257,946 per QALY. The authors concluded that the adoption of mechanical circulatory support has occurred despite, apparently, not achieving stated or implied cost-effectiveness thresholds. Indeed, no study in the review 'concluded that mechanical circulatory support is cost-effective with respect to optimal medical management'. While an assessment of possible contributing factors is apparently outside the scope of the review, one possible reason is that decision makers are not interested in being merely 'informed' through imaginary ICER models, presenting with such diverse claims, but are looking to evaluable claims and feedback from those claims to support purchasing decisions.

A major criticism of modeled claims for cost-effectiveness is that the models typically ignore issues of adherence and persistence with therapy. In previous commentaries it has been pointed out that if patients are non-persistent with therapy then if the majority of patients are non-compliance within two to three years of an index prescription⁴. It makes little sense to advocate formulary acceptance from a lifetime high-compliance cost-utility model. The importance of the Brilleman et al review is to emphasize how poorly adherence and persistence behavior are captured in clinical trial protocols and the importance of this for economic evaluations alongside clinical trials²⁸. To which might be added the implications of taking trial end points as inputs to the creation of modeled imaginary worlds when these are potentially impacted by compliance behavior.

Conclusions

In a recent commentary on the ICER evidence review for PCSK9 inhibitors, the question was raised as to the likelihood that authors of health technology assessments could put what can be seen as the pointless construction of evidence and comparative claims from imaginary worlds behind them¹⁰. Committing themselves instead to a research program for comparative claims assessment that embraced the standards of normal science and not its explicit rejection as evidenced by the latest CADTH guidelines¹³. In support of this, it was also proposed that journal editors could make it clear to their readership that published modeled claims, if they persisted in focusing on non-evaluable claims, could caution that the results presented did not meet the standards of normal science. A 'red flag' warning that would clarify circumstances where the modeled claims could be interpreted as a marketing exercise.

The PCSK9 commentary concluded that this 'reformation' was unlikely to occur¹⁰. There were too many vested interests in supporting the status quo. After all, with over 30 years acceptance and publication of literally thousands of non-evaluable claims in the leading journals, and acceptance of this commitment to constructed evidence by agencies such as

NICE, CADTH and the PBAC, the discomfiture would be (to say the least) embarrassing. Let alone, it should be added, the impact on thousands of graduate students and others who have been trained to construct imaginary worlds and put the standards of normal science to one side. Richard Dawkins, in *Unweaving the Rainbow*, recognizes our willingness to feed on ‘superstition, the paranormal and astrology’³¹. Or, as he describes it, our continuing appetite for being ‘Hoodwink’d with faery fancy’. Perhaps we could recognize our appetite for ‘faery fancy’ and put the endorsement and publication of modeled imaginary worlds behind us.

Table 1: Imaginary Worlds: Modeled Economic Evaluation Studies in *Value In Health* January 2016 to December 2016

Paper (author)	Target Population and Intervention	Sponsor (if any)	Modeling Technique and Claims Status	Claims Assessment and Credibility
Gregory et al ¹⁵	Protocol for diagnosis of appendicitis in children	None stated	Study objective to apply a decision model to quantify the benefits, costs and harms for children with suspected appendicitis considering a validated clinical decision rule and a staged ultrasound and computed tomography imaging protocol. Markov model applied to estimate long-term clinical and economic outcomes. In absence of empirical data the Markov model was used to estimate radiation-induced cancer risks resulting from computed tomography. Simulation of 100,000 hypothetical population on a yearly cycle from age 10 years to death or 100 years. Strategies compared on basis of estimated health benefits and QALYs. With a discount rate of 3% cost-effectiveness assessed against willingness-to-pay thresholds of \$50,000 and \$100,000. QALY gains minimal; for both boys and girls willingness-to-pay thresholds exceeded.	No consideration of modeled credible, evaluable and replicable claims. No possibility of meeting standards of normal science given Markov framework and timelines.
Johnson et al ¹⁶	Increasing physical activity and maintaining weight in sedentary African American Women	National Institute for Nursing Research	To evaluate the 48-week Women’s Lifestyle Physical Activity Program trial. Outcomes reported included physical outcomes, weight stability and marginal cost effectiveness ratios for each outcome with bootstrap confidence intervals.	Absence of consideration of possible protocol(s) to evaluate the introduction of this program in treatment practice and cost/physical activity outcomes criteria for continuing support.
Tilden et al ¹⁷	Management of blepharospasm	Merz Pharmaceuticals	Markov state transition model to support a cost-utility analysis comparing incobotulinumtoxin-A (the sponsor’s product)	No consideration of modeled credible, evaluable and

	and cervical dystonia		to onabotulinumtoxin –A (Allergan) in an Australian modeled environment. Three health states modeled and patients followed on weekly cycles up to 5-years. HRQoL assessed via EQ-5D instrument. Results supported incobotulinumtoxin-A with with incremental cost per QALY of A425,588 and A423794 respectively in blepharospasm and cervical dystonia. QALYs gained for incobotulinumtoxin-A due to earlier alleviation of symptoms in model.	replicable claims. No recommendation for meeting required standards of normal science in protocol-driven claims assessment outside of model.
Heijnsdijk et al ¹⁸	Prostate cancer detection with Prostate Health Index	Beckman Coulter Inc	Modeled assessment of Beckman Coulter Prostate Health Index (PHI - the sponsor’s product) based on European Randomized Study of Screening for Prostate Cancer Trial. Object to evaluate effects of prostate-specific antigen (PSA) screening with PHI vs. PSA-only. Model predicted number of prostate cancers, negative biopsies, deaths, QALYs of PSA and PHI in a hypothetical population aged 50 to 75 years tested at 4-year intervals. Conclusion: compared to PSA, PHI reduced number of negative biopsies and more cost-effective.	Although authors suggest additional research, no consideration of modeled credible, evaluable and replicable claims. No recommendation for protocols to support claims assessment and meeting required standards of normal science.
Becker et al ¹⁹	Options in patients with previously untreated chronic lymphocytic leukemia	Support for third party writing assistance from F Hoffmann-La Roche Ltd	Markov model with a base-case time horizon of 20 years (considered equivalent to lifetime). Assessed the cost-effectiveness of six alternative first-line treatments, to include anti-CD20 monoclonal antibody obinutuzumab plus chlorambucil (GClb – Hoffman La Roche) in untreated patients with chronic lymphocytic leukemia unsuited for full-dose fludarabine therapy from a UK NHS perspective. Three health states: progression free survival, progression and death. In base-case GClb cost-effective against all comparators in the model under a range of plausible modeled scenarios.	Claims presented are non-evaluable No recommendation for protocols to support claims assessment to meet required standards of normal science.
Bijlani et al ²⁰	Prostate surgery costs with robotic-assisted versus retropubic radical prostatectomy	Intuitive Surgical	Modeled care pathway analysis, including multiple iterations with pathways, from hospital and payer perspectives comparing robotic assisted laparoscopic prostatectomy (RALP – the sponsor’s product) versus retropubic radical prostatectomy (RRP). Clinical outcomes modeled from systematic literature review. Monte-Carlo probabilistic sensitivity analysis generated claim that RALP had a 38% to 99% probability of cost-savings.	Claims presented are non-evaluable No recommendation for protocols to support claims assessment to meet required standards of normal science.
Rafia et al ²¹	Alternative upper age limits for breast cancer screening	UK National Institute for Health Grant	To assess whether the extension of breast screening to women older than 70 years would be cost-effective from a UK NHS perspective. A natural history model breast cancer progression model simulated time at which breast cancer presents and its characteristics	Claims presented are non-evaluable No recommendation for protocols to support claims assessment to meet required

			at time of detection in absence of screening vs. characteristics of cancer if detected earlier through screening. Six health states identified with utilities determined by expert opinion. Outcomes were modeled number of cases detected/ 10 ⁵ screened, incremental discounted life years and quality of life years/ 10 ⁵ screened, incremental discounted costs / 10 ⁵ screened, discounted costs per life years gained and QALYs gained.	standards of normal science.
Hinde S et al ²²	Advanced ovarian cancer	Funded as part of ICON7, a UK Medical Research Council sponsored, academic-led and Roche supported trial.	A partitioned lifetime survival model with three disease states (preprogression, postprogression and death) extrapolating beyond the 5-year end point of the ICON7 trial for bevacizumab in advanced ovarian cancer. The study followed on the NICE decision not to recommend bevacizumab (a Roche product) for advanced ovarian cancer which was not based on the unlicensed lower dosage of the drug despite being used in the NHS and the ICON7 trial. The analysis, given NICE willingness-to-pay thresholds, was to consider if the lower dose was cost effective versus chemotherapy alone. The base-case analysis demonstrated that in none of the scenarios was bevacizumab cost-effective at the NICE conventional willingness-to-pay thresholds. Price reductions of 46% and 67% would be required to meet NICE thresholds. In a long-term scenario price reductions would be 21% and 45%.	Claims presented are non-evaluable. While the authors recommend future research no recommendation for protocols to support claims assessment to meet required standards of normal science.
Altawalbeh et al ²³	Older adults with persistent cancer	University of Pittsburgh School of Pharmacy	Markov model to estimate incremental costs and quality-adjusted life expectancy associated with inhaled corticosteroid treatment (ICS) with long-acting beta agonists (LABA) compared to ICS with leukotriene receptor agonists (LTRA) in older adults with persistent asthma. Simulated cohort of persons 65 years of age or older ICS+LABA vs. ICS+LTRA. Model transition on one-month cycles through 5 clinical health states followed over 20 years. Five health states: healthy without any exacerbation, post-asthma exacerbation, post-cardiovascular (CV) exacerbation, post asthma/CV exacerbation and death. EQ-5D utilities from literature or for younger adults with asthma. Results: ICS-LABA costs \$5,823 more than ICS + LTRA while gaining 0.03 QALYs or \$209,090 per QALY. ICS + LABA not recommended in this group.	Claims presented are non-evaluable. While the authors recommend future research no recommendation for protocols to support claims assessment to meet required standards of normal science.
de Boer et al ²⁴	Quadrivalent vs. trivalent influenza vaccine	Sanofi Pasteur	Dynamic transition model to estimate age-stratified numbers of symptomatic influenza cases under quadrivalent vaccine (QIV – sponsor’s product) and trivalent (TIV)	Claims presented are non-evaluable and do not meet required

			strategies. Model to estimate impact of QIV over TIV in clinical outcomes, costs and health effects. Over a 20 year timeframe replacing TIV with QIV projected to reduce influenza B cases by 27.2% (16.0 million), prevent 137,000 hospitalizations and 16,100 deaths, with a gain of 212,000 QALYs. ICER projected as \$27,411/QALY.	standards of normal science.
Allen R et al ²⁵	Chronic immune thrombocytopenia	Glaxo Smith Kline	Markov cohort model comparing response after initial treatment with either eltrombopag (sponsor's product) or romiplostin in treatment of chronic immune thrombocytopenia. Patients modeled to receive a sequence of treatments, time spent in each of six health states in each treatment arm and lifetime long-term outcomes. Modeled results claimed eltrombopag dominated romiplostin in both splenectomised and nonsplenectomized patients. In UK treatment practice there was a claimed 99% and 92% chance of eltrombopag being cost-effective respectively at the NICE willingness to pay threshold of £20,000.	Claims presented are non-evaluative and do not meet required standards of normal science.
Moran et al ²⁶	Screening for atrial fibrillation	No funding	A 25 year Markov model to simulate costs and clinical outcomes with and without screening for atrial fibrillation in Ireland in persons 65 years of Age and over in primary care. Assuming those detected through screening have a comparable stroke risk profile to those detected in routine practice, the model claimed ICER €23,004/QALY compared to routine care.	Claims presented are non-evaluative and do not meet required standards of normal science.
Bohensky et al ²⁷	BRAF wild-type melanoma	None stated but two of the five authors are employees of Bristol-Myers Squibb (BMS) in Australia	A state-transition Markov model simulated over a 10 year time horizon the to project the cost-effectiveness of nivolumab (a BMS product approved by PBAC in Australia May 2016) versus Ipilimumab for previously untreated patients with BRAF-advanced melanoma. Outcomes modeled were progression free survival and overall survival. Quality of life data were from the nivolumab trial. Over 10 years nivolumab compared to ipilimumab yielded 1,58 life-years and 1.30 QALYS per person at a net cost of US\$39,039 per person. ICER US\$25,101 per year of life saved and \$30,475 per QALY saved. Nivolumab was claimed to be cost-effective at a willingness to pay threshold of US\$35,000.	Claims presented are non-evaluative and do not meet required standards of normal science.

Table 2: Imaginary Worlds: Systematic Reviews of Modeled Economic Evaluation Studies
Value in Health January 2016 to December 2016

Paper (author)	Systematic Review Focus	Sponsor (if any)	Modeling Technique and Claims Status	Claims Assessment and Credibility
Brilleman et al ²⁸	Treatment non-adherence and impact on economic evaluations alongside clinical trials	National Institute for Health Research (UK)	To review economic evaluations alongside clinical trials from the perspective of how lack of adherence is accommodated and interpreted. Adherence, defined as the degree of correspondence between a trial participants intended and actual treatment, was reviewed in 96 eligible trials. None of the studies used statistical methods to adjust for non-adherence in the economic evaluation. Lack of standard practice for allocating intervention costs to patients on basis of degree of adherence. Reporting was not comprehensive with the potential for biased cost-effectiveness results in nontrivial proportion of studies.	In recommending standards for reporting and accommodation adherence (a clearer distinction needs to be drawn between adherence and persistence) behavior the review also points to a need (unfulfilled) for addressing the issue in models developed to support evaluable and replicable claims.
Nunes et al ²⁹	Cost-effectiveness of long-term mechanical circulatory support	Alberta Heritage Foundation	Systematic review of 11 studies analyzing cost or cost-effectiveness of mechanical circulatory support in end-stage heart failure. Standard of care medical management. Concluded that mechanical circulatory support in a Canadian environment not cost effective. Nevertheless, widespread adoption.	No consideration of modeled credible, evaluable and replicable claims. No recommendation for meeting required standards of normal science.
Kirsch ³⁰	Multicomponent disease management models with Markov models	None	Systematic review of 16 studies of disease management studies utilizing variable timeline Markov framework. Quality of models and individual study characteristics assessed for relevance and credibility. Major limitations noted included bad reporting practice, variation in selection of input parameters, number of Markov states modeled and time horizons. No consensus in outcomes modeled.	No consideration of modeled credible, evaluable and replicable claims. No recommendation for meeting required standards of normal science.

References

- ¹ Langley PC, Rhee TG. 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
- ² Langley PC. Imaginary Worlds: Modeled claims for cost-effectiveness published in *PharmacoEconomics* January 2015 to December 2015. *Inov Pharm.* 2016;7(2): No. 9
- ³ 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
- ⁴ 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
- ⁵ Langley PC. Modeling Imaginary Worlds: Version 4 of the AMCP Format for Formulary Submissions. *Inov Pharm.* 2016;7(2): No. 11
- ⁶ Langley PC. Sunlit Uplands: The Genius of the NICE Reference Case. *Inov Pharm.* 2016;7(2): No. 12
- ⁷ 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
- ⁸ 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
- ⁹ 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
- ¹⁰ Langley PC. Rush to Judgement: Imaginary Worlds and Cost-Outcomes Claims for PCSK9 Inhibitors. *Inov Pharm.* 2017;8(2): No. 11
- ¹¹ Wootton. *The Invention of Science*. New York:
- ¹² Pigliucci M. *Nonsense on stilts: How to tell science from bunk*. Chicago: University of Chicago Press., 2010
- ¹³ Langley PC. True North: Building Imaginary Worlds with the Revised Canadian (CADTH) Guidelines for Health Technology Assessment. *Inov Pharm.* 2017;8(2): No. 9
- ¹⁴ 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
- ¹⁵ Gregory S, Kuntz K, Sainfort F et al. Cost-effectiveness of integrating a clinical decision rule and staged imaging protocol for diagnosis of appendicitis. *Value Health*. 2016. 19:28-35
- ¹⁶ Johnson T, Schoeny M, Wilbur J. The cost of increasing physical activity and maintaining weight for midlife sedentary African American women. *Value Health*. 2016; 19:2-27
- ¹⁷ Tilden D, Guarnieri C, Cost-effectiveness of incobotulinumtoxin-A with flexible treatment intervals compared to onabotulinumtoxin-A in the management of blepharospasm and cervical dystonia. *Value Health*. 2016;19:145-52
- ¹⁸ Heijnsdijk E, Denham D, de Koning H. The cost-effectiveness of prostate cancer detection with the use of Prostate Health Index. *Value Health*. 2016;19:153-57
- ¹⁹ Becker U, Griggs A, Moreno S et al. Cost-effectiveness model for chemoimmunotherapy options in patients with previously untreated chronic lymphocytic leukemia unsuitable for full-dose fludarabine-based therapy. *Value Health*. 2016;19:374-82

- ²⁰ Bijlani A, Hebert A, Davitan M et al. A multidimensional analysis of prostate surgery costs in the United States: Robotic-assisted versus retropubic radical prostatectomy. *Value Health*. 2016;19:391-403
- ²¹ Rafia R, Brennan A, Madan J et al. Modeling the cost-effectiveness of alternative upper age limits for breast cancer screening in England and Wales. *Value Health*. 2016;19:404-12
- ²² Hinde S, Epstein D, Cook A et al. The cost-effectiveness of bevacizumab in advanced ovarian cancer using evidence from the ICON7 trial. *Value Health*. 2016;19:431-39
- ²³ Altawalbeh S, Thorpe J, Thorpe C et al. Cost-utility analysis of long-acting beta agonists versus leukotriene receptor antagonists in older adults with persistent asthma receiving concomitant inhaled corticosteroid therapy. *Value Health*. 2016;19:537-543
- ²⁴ de Boer P, Crépey P, Pitman R et al. Cost-effectiveness of quadrivalent versus trivalent influenza vaccine in the United States. *Value Health*. 2016;19:964-975
- ²⁵ Allen R, Bryden P, Grotzinger K et al. Cost-effectiveness of eltombopag versus romiplostim for the treatment of chronic immune thrombocytopenia in England and Wales. *Value Health*. 2016;19:614-622
- ²⁶ Moran P, Teljeur C, Harrington P et al. Cost-effectiveness of a national opportunistic screening program for atrial fibrillation in Ireland. *Value Health*. 2016;19:985-95
- ²⁷ Bohensky M, Pasupathi K, Gorelik A et al. A cost-effectiveness analysis of nivolumab compared with ipilimumab for the treatment of BRAF wild-type advanced melanoma in Australia. *Value Health*. 2016;19:1009-1015
- ²⁸ Brilleman S, Metcalfe C, Peters T et al. The reporting of treatment non-adherence on economic evaluations conducted alongside randomized trials: A systematic review. *Value Health*. 2016;19:99-108
- ²⁹ Nunes A, MacArthur R, Kim D et al. A systematic review of the cost-effectiveness of long-term mechanical circulatory support. *Value Health*. 2016;19:494-504
- ³⁰ Kirsch F. Economic evaluations of multicomponent disease management programs with Markov models: A systematic review. *Value Health*. 2016;19:1039-54
- ³¹ Keats J. *The Eve of St Agnes* quoted in Dawkins R. *Unweaving the Rainbow*. New York: Houghton Mifflin, 1998