INNOVATIONS in pharmacy

Volume 7 | Number 2 Article 18

7-12-2016

Imaginary Worlds: The Status of Modeled Economic Evaluation Claims Published in Value in Health January 2015 to December 2015

Paul C. Langley
University of Minnesota, langley@maimonresearch.com

Taeho Greg Rhee
University of Minnesota, Twin Cities, rhee0041@umn.edu

Imaginary Worlds: The Status of Modeled Economic Evaluation Claims Published in Value in Health January 2015 to December 2015

Cover Page Footnote
Note: An earlier version of this paper was presented as a poster at the May 2016 ISPOR Washington DC meeting: Langley PC, Schommer JC, Rhee TG. Imaginary Worlds: The status of modeled claims published in Value in Health in 2015. The authors would like to than Dr Schommer for his contribution.

Imaginary Worlds: The Status of Modeled Economic Evaluation Claims Published in *Value in Health* January 2015 to December 2015

Paul C Langley, PhD and Taeho Greg Rhee, PhD(c)
College of Pharmacy University of Minnesota, Minneapolis MN

Note: An earlier version of this paper was presented as a poster at the May 2016 ISPOR Washington DC meeting: Langley PC, Schommer JC, Rhee TG. Imaginary Worlds: The status of modeled claims published in Value in Health in 2015. The authors would like to thank Dr Schommer for his contribution.

Abstract

The purpose of this paper is to assess the extent to which modeled or simulated cost-effectiveness claims published in Value in Health in 2015 meet the standards of normal science. To meet these standards, modeled or simulated claims must be credible. They must be capable of empirical evaluation and replication. If these standards are not met then such claims run the risk of being labeled as pseudoscience. Following a systematic review of all publications in Value in Health, 16 papers were identified. Of these 14 presented a cost-per-QALY analysis, with 9 presenting their claims for comparative effectiveness in a lifetime cost-per-QALY framework. With the focus of the assessment on whether or not these studies generated testable claims, none of the studies met this standard. They were best seen as thought experiments or imaginary worlds. Recipients of such claims can of course, on the one hand, reject them outright as not meeting accepted standards in normal science. After all, QALYs are never collected by health care systems and are unlikely to be collected. This means that the outcome metrics are untestable and may never have been intended to be tested. On the other hand, if the recipient believes that the model or simulation provides a sufficient correspondence to reality then the claims made are necessarily entailed. The issue of testing is irrelevant to the belief in the credibility of the claims. The modeled or simulated claims are immune to failure. The review concluded that none of the claims presented were in a testable form and that while 7 (at most) of the studies had the potential to generate testable claims, the rest were immune to failure. In the absence of testable claims, the studies reviewed are most appropriately characterized as imaginary worlds or thought experiments.

Keywords: Value in Health, pseudoscience, cost-effectiveness modeling, credibility, imaginary worlds, scientific method

Introduction

In a recent supplement to the *Journal of Medical Economics (JME)*, the case was put forward that if claims for the impact of products and devices on costs and outcomes in health care systems are to be accepted then they should meet the standards expected in 'normal science' [1 2 3 4]. The argument was made that if modeled or simulated claims are to be credible, practical and useful in formulary decisions then the only acceptable claims are those that are potentially falsifiable and replicable in a timeframe relevant to the needs of a formulary committee. If claims do not meet this standard they should be rejected.

The present review of cost-effectiveness studies published in *Value in Health* in the period January 2015 to December 2015 follows from two previous assessment of modeled or simulated cost-effectiveness studies published in *PharmacoEconomics (PECON)* and the JME in the same time period [5 6]. Both reviews concluded that the majority of studies published did not meet the standards of normal science. The studies were best characterized as thought experiments or imaginary worlds; as pseudoscience rather than science in putting to one side the construction of empirically evaluable theories and hypotheses. This is exemplified in the choice of cost-per-QALY endpoints. Apart from the fact that there is no accepted standard for a QALY metric, this is an immediate barrier to generating testable hypotheses as QALYs are not only not collected by health systems but no health system appears to be interested in collecting them [7].

With the exception of a handful of single payer health care systems that have embraced the National Institute for Health and Care Excellence (NICE) reference case, to include New Zealand (with the exceptions of cost-per-QALY thresholds) and Ireland, there seems little interest in adopting the reference case constructed model or simulation format to support pharmaceutical product decisions [8 9 10 11]. Mention should, however, be made of the Academy of Managed Care Pharmacy (AMCP) Format for Formulary Submissions in the US, which recommends long-term models and simulations in chronic disease [12]. This Format, unfortunately, also fails to meet the standards of normal science [13].

Corresponding author: Paul C Langley, PhD
Adjunct Professor
College of Pharmacy, University of Minnesota
Email: langley@maimonresearch.com

Methods

A systematic review, following the PRISMA-P checklist (MeSH terms 'cost', 'cost effectiveness' , 'QALY') of all papers published in *Value in Health* in the period January 2015 to December 2015, identified 16 economic evaluation studies [14]. 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 author(s) 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 up to 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 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.

In judging whether or not a model might support testable (falsifiable) claims, even if the possibility was not considered by the author(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 that 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 PROs) and resource utilization endpoints would, given access to readily available data sources in the US, be open to hypothesis testing and feedback to a formulary committee.

The important point to note is that the modeled claim was not to be judged on the reasonableness or otherwise of the assumptions of the model. Certainly the model would be expected to cover comparator products, or least the key comparators, and to identify the target population for the claims. Beyond this, there was no attempt to evaluate whether or not the model necessarily complied with ISPOR recommended standards for good practice, although given that V*alue in Health* subscribes to these standards, notably the CHEERS format, it was assumed that these criteria would have been addressed as part of the peer review process [15].

Results

Table 1 summarizes the results of the review of the economic evaluation studies. Each paper is assessed under four headings:

- Target population and intervention (product, comparator products, devices)
- Sponsor (financial support if known)
- Modeling technique and major claims (simple decision model, Markov or cohort model, discrete event simulation)
- Claims assessment

Of the 16 articles reviewed:

- 14 of the 16 papers presented a modeled cost-per-QALY analysis [17 19 20 21 22 23 24 25 26 27 28 29 30 31]
- 9 of the 14 cost-per-QALY papers presented lifetime modeled claims [19 20 22 23 24 25 26 27 28 29]
- One paper presented the analysis as lives saved and costs [16]
- One paper presented results as costs per major complication/death avoided [18]
- Six papers were supported or funded by manufacturers and all modeled claims supported the manufacturers' product [19 25 28 29 30 31]

Lifetime modeled cost per-QALY claims were not expected to generate evaluable claims. The only exception was the analysis by Carlson et al of the cost-effectiveness of tocilizumab versus adalimumab for patients with rheumatoid arthritis for whom methotrexate is inappropriate [29]. In this case a six month initiation phase was modeled as well as the patient's lifetime.

Of the 5 cost-per-QALY models that did not take a lifetime perspective, the timeframes varied from 2 to 7 years. There was no suggestion in any of these papers as to how testable claims might be generated by the model and how these claims might be evaluated in a treating environment. The same conclusions apply to the remaining two papers.

In respect of the four questions addressed above, the conclusions are: of the 16 economic evaluations presented in

2015 by *Value in Health*, only 7 (at most) had the potential to develop testable claims, although none did so.

None of the economic evaluations addressed the issue of testable claims (and, of course, did not address the issue of how the claims might be evaluated and replicated)
None of the evaluations cautioned readers as to the implications of generating non-testable claims as potential inputs to health care decisions

Discussion

Given the increased emphasis in the pharmacoeconomics literature on the potential role of QALYs as the preferred endpoint in cost-effectiveness claims, the fact that 14 of the 16 economic evaluations utilized a cost-per-QALY framework should come as no surprise. Possibly more surprising is the fact that 9 of these 14 papers used a lifetime cost-per-QALY framework. Perhaps this can put be down to standards (including mandated and recommended guidelines for formulary submissions) for modeling the 'natural' course of a chronic disease. Even so, the framework and endpoints adopted puts these studies in the category of pseudoscience. The reader is asked to take (or leave) these conclusions at face value. It is unclear how these claims could be factored into formulary decisions if the recipient raised the issues of assessment and replication.

Overall, however, the most interesting question that emerges from this review is the fact that none of the evaluations raised any concerns as to what is best described as the imaginary nature of the models presented. There is no attempt to address the issue of sufficient correspondence in constructing the model or simulation in generating predictive claims. Instead, the reader is asked to take the inherent reasonableness of the simulation or model at face value. With due regard to the limitations pointed to in these evaluations, the claims for cost-effectiveness, for the probable superiority of one product or course of treatment over another, the authors clearly believe that their models should be taken seriously as a quantitative contribution to informing health decision makers.

This does not mean that, at least for a few of the evaluations presented, there is not the potential to generate testable predictions. The Vertuani et al study, for example, points out that QALYs as a measure of effectiveness may not fully represent the patient's perspective [17]. The authors note that as the QALYs were EQ-5D based, relatively small differences in pain may not be fully captured. In addition, the full treatment pathway after surgical complications was not modeled due to the absence of data, nor was there data to capture post-operative work loss. Even so, in focusing on a two year timeframe, it would have been possible to generate modeled claims and to propose how these might be validated in treatment settings, possibly including complementary measures of HRQoL as well as claims for resource utilization, reduction in length of hospital stay, reduced blood loss and fewer complications. This was not attempted.

Another cost-per-QALY study that had the potential to generate testable claims over the short-term, is the Nguyen et al model of treatment resistant depression interventions[19] . The 3-year Markov model yields the unequivocal claim that rTMS dominated pharmacotherapy for patients with treatment resistant depression, generating (admittedly minimal) QALYs gained (1.25 vs. 1.18) but at a slightly lower cost with the 73% probability, at a threshold of A$50,000, that it is cost-effective. Given this it would surely have been possible to propose an assessment protocol that would have tested these claims, with possibility that with such minimal benefits and cost savings the claim for cost-effectiveness may have been overturned.

The Legrand study, although based on an acute model of hemodynamic monitoring and fluid therapy strategies gave no indication of how this might be translated to evaluate claims for cost and clinical outcomes such as hospital mortality and major complications [18].

Unfortunately, from the perspective (as discussed in detail below) of normal science, the preference for lifetime cost-per-QALY models in 9 of the 16 papers points to the acceptance of a methodology that, while conforming to ISPOR recommended standards, is at odds with that of normal science. The models and claims are immune to failure and, as such, if these standards are accepted, should be put to one side. [19 20 22 23 24 26 27 28 29] . While there may be an audience for this type of model, there was no attempt to caution health decision makers that, even if the limitations to the model noted by the respective authors were overcome, the claims would still lack credibility and should be put to one side as speculative imaginary worlds. While they may appear persuasive in their simulation framework and the data sources utilized to populate the simulation, in the last resort we have no idea whether the claims made are right or even if they are wrong. Appeals to sensitivity analysis, the application of cost-effectiveness acceptability analysis to support thresholds and statements that there is an x% likelihood the product is cost-effective are simply speculation. They do nothing to change the fact that the claims are impossible to falsify or replicate.

Indeed, what is missing in all of the models reviewed is any concept of feedback from claims, their replication and the potential for revising and enhancing models to accommodate new therapies as they enter the market place, competing against existing therapies.

There has been increasing concern expressed in the last few years over the presence of repetitive flaws and the need for guidelines to improve experimental reproducibility. This applies equally well to simulations and models in pharmacoeconomics. As noted in a recent editorial in *Nature*, applicants to the National Institutes of Health (NIH) are now required 'to explain the scientific premise behind their proposals and defend the quality of their experimental design' [16]. More recently, Camerer et al in their evaluation of laboratory experiments in economics find, of the 18 studies considered, an effect size in the same direction in only 11 replications with on average a replicated effect size of 66% of the original[17] . As the authors point out 'the deepest trust in scientific knowledge comes from the ability to replicate empirical findings', although rarely carried out in the social sciences.

Informing Decision Makers

If these models and their claims are intended to inform decision makers, as presumably is the case for those sponsored by manufacturers, then the effort is probably wasted. In each case we could, presumably, engage other research groups to develop competing models and come up with competing claims. This possibility was recognized over 20 years ago by the *New England Journal of Medicine* (*NEJM*) when the journal set out, in an Editorial, its policy on cost-effectiveness studies [18]. The case put forward was that because of the discretionary nature of cost-effectiveness methods it was incumbent upon authors, journal editors and funders of such studies to minimize any source of bias. In consequence, the policy of the *NEJM* on publication was (and remains) not only to ask for an author's financial connections with a company, but to reject for consideration any article, to include reviews and editorials, in which an author has a personal financial conflict of interest. In the present case, as noted above, 6 of the 16 papers had funding or other support from the manufacturer whose product, on the modeled assessment, was claimed to be cost-effective.

There is, however, a more substantive reason than the potential for bias for journals such as the *NEJM* to establish standards in accepting cost-effectiveness studies: the fact that simulated or modeled claims for cost-effectiveness, unless they generate testable hypotheses and meet the standards for normal science are simply imaginary worlds or thought experiments. As discussed below: simulations, even if there is a claim for similarity, are not experiments. Unless there is the ability to generate testable claims (and repeatedly test those claims) and provide guidance as to how these claims could be assessed, the cost-effectiveness analysis lacks credibility. The analysis may be aimed at informing and influencing decision makers but there is no way of judging whether the information provided is relevant to decision making. The claim may be quite misleading, indeed harmful, although we have no idea as to how misleading it might be.

Testable Hypotheses

The requirement for testable hypotheses in product and device impacts is unexceptional. Since the 17th century it has been accepted that if a research agenda is to advance, if there is to be an accretion of knowledge and if models are to generate meaningful hypotheses, then these hypotheses must be such that they can be empirically evaluated. This position is made abundantly clear by Wootton in his reassessment of the use of language in the idea of the scientific revolution in *The Invention of Science* [19] and by Grayling in his assessment of the intellectual history of the 17th century *The Age of Genius* [20].

If a simulation or model fails to generate testable or measurable hypotheses, then it should be seen as simply a construct to support the exploration of imaginary worlds or thought experiments and not part of a meaningful research program; a program that underpins the notion of progress in the accumulation of knowledge. Both Wootton and Grayling point to the motto of the Royal Society, first meeting in 1660 and a Royal Charter in 1662, *nullius in verba* as evidence of the commitment to experimentation, falsification and replication, with Grayling, to give a wider European perspective, pointing to the motto of the Accademia del Cimento in Florence in 1657 and their motto *Provando e Riprovando* [27 28]. Indeed, even earlier, as Wootton points out, in the 16th century Leonardo da Vinci (1452 – 1519) in notes prepared posthumously in 1540 for his *Treatise on Painting* (published in 1641) clearly anticipated the standards for the scientific method in rejecting 'sciences which begin and end in the mind' and which fail the 'test of experience'[27].

In the early 20th century standards for empirical assessment were put on a sound methodological basis by Popper in his advocacy of a process of 'conjecture and refutation'[21 22] . Hypotheses or claims must be capable of falsification; indeed they should be framed in such a way that makes falsification likely. Life becomes more interesting if claims are falsified because this forces us to reconsider our models and the assumptions built into those models. To the extent that the proponents of the pharmacoeconomic modeled claim believe that it can be defended on the grounds that it 'reflects reality' or is a 'reasonable representation' of what is 'out there' (whatever those terms actually mean), it is worth reflecting on Popper's statement: 'never in science are inferences drawn from mere observational experience to the prediction of future events' [23]. The fundamental issue is one of demarcation: to distinguish science (e.g., natural selection) from pseudoscience (e.g., intelligent design).

Standards in Pharmacoeconomics

It is curious that in the standards proposed by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) there is no mention of the need to meet the standards of normal, experimental science in respect of (i) falsification and (ii) replication. Certainly predictive validation (rather than falsification) is mentioned, but it is considered as more of an optional (and even preferred) extra to other recommended forms of model validation. The possibility of committing to an ongoing process of replication and assessing the presence of false positives is entirely absent. A reasonable question to ask is why this has this occurred?

Looking back over the past 20 years there are literally hundreds (if not thousands) of published simulated and modeled claims that fail to meet the standards of normal science. A reviewer might then reasonably ask how did this situation arise? Is it an accident? Do we assume that those accepting these standards (and the number of ISPOR standards that support such models) were simply unaware of the standards of normal science? Or is there an implicit (if not explicit) acceptance that in pharmacoeconomics different standards should apply? Is there a research program, as it would be understood in the physical sciences, that underpins pharmacoeconomic studies? Or is pharmacoeconomics willing to continue to support the creation of imaginary worlds with studies that are immune to falsification?

A curious feature of the pharmacoeconomics literature is the apparent absence, after some 30 years, of successor studies that have attempted to evaluate claims put forward in published modeled claims. Unlike disciplines like physics where there is an ongoing appraisal of claims and an accumulation of theoretical and empirical knowledge generated by practitioner's ability to develop practical and useful models, pharmacoeconomics is silent.

From the late 1980s when attention began to be given to the applications of decision modeling to clinical decision making, the role of decision models was seen as 'informing' health care decision makers. It was not clear, however, as to what form this provision of information should take and the standards by which this evidence should be judged. Weinstein and Fineberg in their seminal textbook on decision analysis argue that decision analysis 'offers a prescriptive model for clinical decision making ….an aid, not a substitute for clinical reasoning' [24]. More recently Drummond et al in the 3rd edition of their text on economic evaluation see the 'ultimate purpose of economic evaluation is …to inform different types of decision makers about the efficient allocation of health care resources' with the 'greatest use of these methods is to inform particular decisions in specific jurisdictions providing the means of bringing evidence from a range of sources together… and providing a framework for decision-making under uncertainty' [25]. Unfortunately, if 'information' as an aid to decision making is presented either in the form of untestable hypotheses or where it is immune to falsification (and replication) it is difficult to see whether a decision maker should take the evidence presented seriously.

Truth is Consensus

But we do not have to subscribe to the standards of normal science; pharmacoeconomics is perfectly within its rights to adopt a relativist position where all perspectives are equally valid. Indeed, support for a relativist position is widespread, at least outside of the natural sciences. Following from the publication of *Philosophical Investigation* by Wittgenstein, and the view that truth is merely consensus, the emergence of the so-called Science and Technology Studies in the 1960s in the UK, as described by Wootton, in their advocacy of the equivalence principle saw a widespread acceptance of the relativist position [26 24]. Applying the principle of symmetry, for relativists the content of science is to be understood sociologically. Relativists reject the notion that one body of evidence is superior to another. We cannot adopt the viewpoint of one community and reject another. As Wootton summarizes the relativist position: *The success of a scientific research programme thus depends not on its ability to generate new knowledge but on its ability to mobilize the support of a community* [24]. In this community research program evidence is always constructed; it is never discovered. Rather than attempting to come to grips with reality, science is about *rhetoric, persuasion and authority*.

The embrace of a relativist position in pharmacoeconomics sets the stage for the acceptance of simulations to construct new evidence. Truth is constructed. Hypothesis testing and replication are redundant. Even, presumably, if the predictions relate to the lifetime discounted cost-per-QALY outcomes of a target chronic disease population. We put to one side the possibility that simulations can fail and accept that any simulation is 'doomed to succeed' [27].

If decision makers are prepared to accept the relativist position then, presumably, this is the end of the story. Journals will continue to publish claims that are non-testable and immune to falsification, claims that are factored into formulary and pricing decisions. It is difficult to believe that this position is acceptable. It is difficult to take seriously a consensus belief that models and simulations are sufficiently representative of a target reality.

An imaginary world yields imaginary claims. A consideration that needs to be kept in mind when a simulated or modeled claim is published that is sponsored by and which favors a manufacturer's product or if recommendations are made by independent assessors. If a claim is immune to falsification then decisions for formulary listing, discounted or premium

pricing are open to challenge. A formulary committee cannot point to evidence, other than the constructed evidence of a simulation, to support their decision. Indeed, it is all too easy, given the inherent flexibility allowed in constructing imaginary worlds, for competitors to present contradictory claims. How could this existential threat be handled within a relativist paradigm? With two imaginary worlds competing for attention, both with untestable competing claims, there is no basis for a resolution. The protagonists could argue over modeling techniques, over assumptions and over thresholds – a debate that is unlikely to lead anywhere. Even if ISPOR or some other organization took it upon itself the task of 'simulation adjudication' it would seem a pointless exercise when testable short-term modeled competing claims could be readily adjudicated by an appeal to the facts: to falsification and replication.

Attempts in the US attempts to assess comparative QALY product performance has, more by accident than an appeal to the standards of normal science, met resistance with the Patient Protection and Affordable Care Act (2010). The act forbids the Patient-Centered Outcomes Research Institute (PCORI) to use cost-per-QALY "or similar measure that discounts the value of a life because of an individual's disability as a threshold to establish what type of health care is cost effective or recommended". The fact that there appears to be little interest, in any case, by health care systems in the US in claims expressed in terms of QALYs, is a further positive feature. Indeed, from a global perspective, the US is not alone in failing to put procedures in place to collect QALYs on a regular basis – even if a QALY standard could be agreed as a common metric across health care systems. Even in the UK, where the reference case is the assessment standard, the fact is that no health authority collects QALYs. Given the absence of any acceptance and implementation of a QALY metric, the emphasis on QALY modeled claims seems even more difficult to defend.

Conclusions

Unfortunately, practitioners, journal editors and groups such as ISPOR do not have the luxury of putting the question of scientific credibility to one side. It may be a difficult decision to make particularly when it is pointed out that there are (and have been) substantial commitments made by manufacturers and others to supporting cost-effectiveness studies which in retrospect generate claims that are immune to failure. If we consider the resources manufacturers have devoted internally in staffing health technology assessment groups and to supporting consultant activities, the question may reasonably be asked as to what benefits have been derived from these activities and, for the future, what benefits might be expected if there are concerns as to the robustness and credibility of these activities? Should they continue to accept that decisions in medicine can be based on modeled or simulated claims that are untestable and which fail to meet the standards of normal science?

What are the options open to ISPOR, *Value in Health* and to practitioners in pharmacoeconomics if they are to avoid the charge that all too often modeled claims lack credibility and fail to provide an input to health decision makers that is practical and useful? An important step would be to acknowledge that, for the past 25 years, those advocating standards and good practice for modeled claims have failed to consider the need to meet the standards of normal science. This may be a difficult decision to take; after all, to admit that literally dozens if not hundreds of published studies lack credibility and should be put to one side is not an easy decision. The next important step would be to reconsider the standards for 'good' modeling or simulation and ensure that these recognize the essential place of falsification and replication. The Editor of *Value in Health* could request that when authors submit simulations or models that lack testable claims that they state explicitly that the study does not meet the standards of normal science. Authors should state (i) that the claims are not open to (a) falsification and (b) replication and (ii) that in the absence of experimentation the claims may be right but also that they may be wrong.

Table 1: Imaginary Worlds: Economic Evaluation Studies *Value in Health* January 2015 to November 2015

Paper (author)	Target Population and Intervention	Sponsor (if any)	Modeling Technique and Claims Status	Claims Assessment and Credibility
Danese et al.[28]	Breast cancer life years saved by trastuzumab plus chemotherapy alone as first line therapy vs. trastuzumab plus pertuzumab plus chemotherapy	Genentech	Simulated life years saved from using trastuzumab as first line therapy in HER2+ metastatic breast cancer compared to counterfactual scenario without trastuzumab (which was approved in 1998) for period 1999-2013 and projected life years to be saved from using trastuzumab and pertuzumab as first line therapy with chemotherapy in period 2013 to 2027 compared to trastuzumab plus chemotherapy. From estimates of life years saved an example calculation of the value of trastuzumab and the combination of trastuzumab and pertuzumab at a population level at a value of $150,000 per life-year-saved. Cumulative life years saved from 1999-2013 from first line trastuzumab use was 156,413 (if 100% utilization life years saved would have been 281,948). Life years saved projected for 2013-2027 with trastuzumab projected to be 328,200. With addition of pertuzumab (100% utilization) life years saved increased to 466,159. Incremental economic value of life years saved (for 2013) was estimated to be $1.66 billion for trastuzumab with chemotherapy and an additional $1.37 billion if pertuzumab had been added (in 2015 dollars). In terms of mean times to progression the gain from adding pertuzumab is $0.06 billion. Results supported sponsor's product pertuzumab.	Claims made impossible to verify both for life years saved and for incremental economic value to support addition of pertuzumab to therapy. No indication given by authors as to how a health care system might evaluate survival or economic value benefits from addition of pertuzumab and increasing utilization of both products in HER2+.
Vertuani et al.[29]	Comparison of minimally invasive surgery (MIS) compared to open surgery (OS) for the treatment of degenerative lumbar spinal conditions using lumbar spinal fusion in UK and Italy	Medtronic International	A 2-year procedure-based cost-per-QALY model. with sensitivity analysis to identify key cost drivers in surgical procedures. Incremental cost effectiveness values calculated with post-operative health state measured as QALYs (EQ-5D) from Swedish source. MIS was estimated to be the dominant therapy with the more expensive MIS procedure offset by reduced costs of perioperative hospitalization and less complications. HRQoL significantly improved 2 years post-operation for both techniques with 0.72 QALYs for MIS and 0.68 QALYs for OS (a gain of 0.04 QALYs for MIS over OS). Calculated ICERs showed MIS to be the dominant therapy in both UK and Italy. ICERs were below recognized threshold values.	Although an acute intervention the authors give no direction as to how a surgical unit or health care system might evaluate these claims. The claims (e.g., resource utilization, reduction in length of stay, reduced blood loss, fewer complications) are potentially testable but no guidance is given for possible evaluation, and

				provision of feedback. Impossible to assess whether these claims would be supported in a treating environment.
Legrand et al.[30]	Comparison of three minimally invasive hemodynamic monitoring and fluid therapy strategies in intermediate and high risk abdominal surgery applications: APPWA (arterial pulse pressure waveform analysis); CCA (conventional clinical assessment; and ED (esophageal Doppler).	None	Decision model with three possible outcomes: discharge without death or complications, major complications and death. Outcomes based on meta-analysis of RCTs. ICERs calculated for costs per major complication avoided and cost per death avoided. One way sensitivity analysis and probabilistic sensitivity analysis performed with results presented as cost-effectiveness acceptability curves. In one-way sensitivity analysis APPWA was always dominant compared to CCA and ED dominant in most scenarios. ED was dominant over APPWA in three situations, but was dominated in all others. In the probabilistic sensitivity analysis APPWA wand AD were dominant compared to CCA in 97.3% and 76.1% of cases and were in 0% and 13.3% of cases respectively. ED compared to APPWA was dominant in only 23.8% of cases and dominated in 76.1% of cases. In deaths avoided, the probabilistic sensitivity analysis showed that APPWA and ED dominated CCA in 92.9% and 69.5% of cases respectively. ED compared to APPWA was dominant in 20.8% cases and dominated by APPWA in 27.6% of cases.	Although an acute intervention the authors give no indication as to how a surgical unit or health system might evaluate the short term claims for cost and clinical outcomes (hospital mortality, major complications) to choose between APPWA and ED. As the threshold values for the probabilistic sensitivity analysis are unlikely to mean anything to decision makers, they add nothing to the analysis. There is no guidance for possible evaluation, and provision of feedback.
Nguyen & Gordon[31]	Comparison of repetitive transcranial magnetic simulation (rTMS) versus antidepressant pharmacotherapy in treatment resistant depression	Griffith University assessment group contracted to the Australian Government	Hypothetical 3-year health state transition (Markov) model with key outcome incremental cost per QALY, the additional cost of rTMS over additional QALY compared to antidepressants; both discounted at 5%. QALYs gained were greater with rTMS while costs were slightly lower. At a threshold of A$50,000 per QALY the probability that rTMS was dominant was 32% and likelihood it was cost-effective was 41%. Resources and costs were standardized to meet constraints of Markov cycle.	There is no guidance for possible evaluation, and provision of feedback. Unclear, for example, how QALY claims would be assessed. Impossible to assess whether these claims would be supported in a treating environment. No support given for A$50,000 threshold and how treatment centers should or would interpret the

				claims for dominance and cost-effectiveness.
Collins & Schwemm[32]	Comparison of vancomycin versus linezolid	None	Cost per QALY decision model, building on previous ZERHyR cost-effectiveness models, in a hypothetical patient cohort with a lifetime horizon. Primary outcome measure incremental cost per QALY gained in a probabilistic sensitivity analysis driven by a Monte Carlo simulation. Base case analysis showed linezolid patients experienced a 6% cost increase while gaining a 0.15 QALY increase. Results of the sensitivity analysis showed that linezolid 78% cost effective at $100,000 US threshold and in range 72.1% to 79.7% if threshold varied between $50,000 and $150,000. Vancomycin dominated in population with documented MRSA.	The lifetime framework for this model excludes any possibility of establishing testable claims. There is no guidance for possible evaluation, in particular given the acknowledged variability in the modeled results, and no provision for feedback for health centers evaluating the relative coast-effectiveness claims for the treatment options.
Ting et al.[33]	Comparative cost-effectiveness analysis of erlotnib, afatinib and cisplatin-pemetrexed for first line treatment of epithelial growth factor receptor mutation-positive non-small-cell lung cancer in the US	None	Lifetime Markov model with cycle length of one month using Monte Carlo simulation. Cohort simulated until all patients were dead. Primary outcomes cost and QALYs. UICER ratios ranked treatments. A cost-effective treatment was defined as an ICER < $100,000 per QALY. Uncertainty captured through cost-effectiveness acceptability curve. Base case found erlotnib cost effective compared to options at threshold of $100.000 QALY in 75% to 80% of model iterations. Expected value of information (EVPI) techniques applied to identify cost of research to reduce parameter uncertainty for effective lifetime of treatments.	There is a potential for testable predictions given the short survivorship among patients. The lifetime framework for this model excludes any possibility of establishing testable claims. There is no guidance for possible evaluation, and provision of feedback for health centers evaluating the relative cost-effectiveness claims for the treatment options.
Nazir et al.[34]	Overactive bladder Mirabegron vs antimuscarine agents	Astellas Pharma	Cost per QALY Markov model with a one-month cycle over a 5 year time horizon. Patients were distributed across 25 symptom severity profiles (from the SCORPIO trial) and assigned at entry to either therapy. Transitional probabilities estimated from multinomial logic regression model. In absence of data, assumed 90% of patients	Authors claimed that from UK perspective the sponsor's product was cost-effective compared to antimuscarine agents. Given the

			experiencing adverse events would discontinue. Utility values (EQ-5D) varied by symptom severity and were estimated by regression modeling from reported trial symptoms. Discount rate of 3.5% applied to costs and health benefits. Primary output ICER expressed as cost per QALY gained. Results supported sponsors product with ICERs in range £367 to £15,593 QALY gained, below threshold of £20,000	short time frame for the model there is a potential for developing testable predictions. However, there were no testable claims presented or guidance as to how the claims made might be evaluated. As the claims are presented in cost-per-QALY terms with QALYs estimated by regression model, no guidance was given as to how QALYs might be generated for treating populations. There is no guidance for possible evaluation, and provision of feedback for health centers evaluating the cost-effectiveness claim for the treatment options.
Bermingham et al.[35]	Assessment of seven alternative antiviral strategies HBeAg-positive or HBeAg-negative hepatitis B	Funding from National Institute for Care and Clinical Excellence (NICE) to National Care Guideline Centre	Probabilistic Markov cohort model to estimate lifetime costs and QALYs of competing therapies. UK NHS and personal social services perspective. Costs and QALYs discounted at 3.5%. Entry to model via either HBeAg-positive or HBeAg-negative hepatitis B. Average age at start of treatment 31 and 40 years respectively. Nineteen relevant treatment sequences were tracked. Model claimed peg-IFN α-2a most effective first line antiviral. For those with HBeAg-positive and HBeAg-negative CHB failing to achieve seroconversion or viral suppression the sequence peg-IFN α-2a →TDF → TDF + LAM most effectives with cost of £7,488 per QALY gained compared to no treatment. For those with HBeAg-negative CHB peg-IFN α-2a→ETV →TDF is most effective treatmnent with cost of £6,981 per QALY gained. Authors recognized that model limited by lack of long term evidence of efficacy, resistance and off-	As a lifetime cost-per-QALY model, the claims are impossible to evaluate. No attempts were made by the authors to generate short-term testable claims even though they acknowledged that long term data to support modeling was sparse or non-existent. There is no guidance for possible evaluation, and provision of feedback for health

			treatment durability, and only a few trials that have evaluated effectiveness of sequential treatment therapies in patients developing resistance, effectiveness and safety of antiviral combinations and on treatment rates of seroconversion.	centers evaluating the cost-effectiveness claims for the various treatment options and sequences.
Dhankar et al.[36]	To assess population level impact and cost-effectiveness of hepatitis A vaccination programs in the US	Merck Sharp & Dohme Corp	Modeled universal vs. regional childhood lifetime hepatitis A vaccination deterministic, age-structured epidemiologic model with an equilibrium age distribution and a stationary population to evaluate transmission of and vaccination against HAV infection in the US. Population for the model was divided in distinct classes (maternal antibodies, susceptible, exposed etc.) with each component further categorized into 110 age groups. Direct treatment costs and indirect costs (work loss) included, as well as public health disease control costs. QALYS were assumed with discounted weight of 0.68 for all outcomes except liver transplant which had a weight of 0.73. In all other health states weight was unity. Outcomes were incidence of hepatitis and cost and QALYS over 100 years. Probabilistic sensitivity analysis indicated that universal vaccination less costly and more effective that regional vaccination in all simulations.	As a lifetime cost-per-QALY model, the claims are impossible to evaluate. No attempts were made by the authors to generate short-term testable claims (e.g., complications from vaccine). There is no guidance for possible evaluation, and provision of feedback for policymakers who might initiate a universal vaccination program.
de Waure et al.[37]	Comparison of cost-effectiveness and budget impact of Therakos online extracorporeal photopheresis (ECP) to commonly used alternatives (mycophenolate, pentostatin, imatinib) for management of steroid-refractory/resistant chronic graft versus-host-disease (cGvHD) in Italy.	Therakos Inc	A Markov model with a 7-year time frame and applied via a Monte Carlo simulation to a hypothetical cohort of 1000 patients. The model captured 3 health states: complete response, partial response and stable disease. The model cycle was 3 months. Health state utilities were derived from the literature with estimated direct medical costs. A discount rate of 3% was applied to costs and QALYs. The economic evaluation supported the dominance of Therakos online (ECP) generating lower costs and higher QALY gains. The probability sensitivity analysis supported this conclusion. The results supported the sponsor's product.	There was the potential, given the number of outcome measures (to include QALYs) to have proposed a set of testable propositions; evaluable well within the 7 year model timeframe. No attempts were made by the authors to generate short-term testable claims. There is no guidance for possible evaluation, and provision of feedback for policymakers who might initiate the proposed treatment

				sequence.
Linssen et al.[38]	To evaluate the cost-effectiveness of screening 50 to 70 year adults for hearing loss in the Netherlands comparing no screening, telephone screening, internet screening, screening with a handheld device and audiometric screening.	Partial funding by the Heinsius Houbolt Foundation	A lifetime Markov model to evaluate cost-effectiveness of screening from a health care perspective. A cost-per-QALY model that simulated no intervention against four types of nationwide screening: telephone, internet, Hearcheck and audiometric screening (both with general practitioner). A total of 76 screening strategies considered in model and compared to no intervention. Utilities assed via HUI Mark 3 instrument questionnaire to generate age-dependent utility scores. Costs and effects discounted at 4% and 1.5% respectively. Probabilistic sensitivity analysis yielded cost-per-QALY screening strategies less than €20,000 /QALY. Telephone and internet strategies dominated. Telephone strategies were either dominated or extendedly dominated by internet screening strategies.	As a lifetime cost-per-QALY model, the claims are impossible to evaluate. No attempts were made by the authors to generate short-term testable claims. There is no guidance for possible evaluation, and provision of feedback for policymakers who might initiate an internet screening program (as compared, for example, to a telephone screening program).
Visser et al.[39]	Cost-effectiveness of cognitive behavioral group training versus wait-list control for patients with unexplained physical symptoms (UPS: DSM-IV somatoform disorder).	None	A probabilistic Markov model with three health states: poor health, average health; death based on SF-36 PCS summary score. Model utilized 3 month cycles over a four year period. Assessment in cost per QALY. Data from UPS randomized trial (n = 162). After 4 years group training dominant with 0.06 QALYs gained and €828 reduction in costs. Cost-effectiveness improved with time achieving threshold €30,000 QALY at 18 months and group cost saving after 33 months.	Although a relatively short time frame no attempts were made by the authors to generate short-term testable claims. There is no guidance for possible evaluation, and provision of feedback for policymakers who might initiate the group training.
Simons et al.[40]	Assessment of patient outcomes and cost-effectiveness of redesign of care processes in patients with head and neck cancer.	None	Markov lifetime model utilized to evaluate biopsy process of suspicious lesions under local instead of general anesthesia, and combining computed tomography and positron emission tomography for diagnostics and radiotherapy planning. Patients in model stratified by disease location and stage (8 groups). Costs and QALY (EQ-5D) estimates calculated for each tumor site. Different gains in waiting time were realized for each patient group. New care process cost-effective for all studied treatment sites (using thresholds of both €80,000 and €20,000 QALY.	As a lifetime cost-per-QALY model, the claims are impossible to evaluate. No attempts were made by the authors to generate short-term testable claims. There is no guidance for possible evaluation,

				and provision of feedback for policymakers as a national policy, or health centers who might initiate a new care product in different treatment settings.
Liu et al.[41]	Cost-effectiveness of high dose hemodialysis (HD) versus conventional in center HD (ICHD) over a lifetime horizon from a UK payer's perspective.	Baxter Healthcare Corp	Markov cost per QALY model comparing HD with conventional ICHD with current and hypothetical HD reimbursement tariffs. Outcome cost-per-QALY over lifetime of patient cohort. High dose HD in-center associated with higher costs and QALYs versis conventional ICHD and thus not cost effective at UK thresholds. HD at home associated with lower costs and QALY increase compared to ICHD. High-dose HD potential to offer improved clinical and QALY outcomes over conventional ICHD, under current UK payments policy, considered cost-effective if conducted at home (61.8% at £20,000 and 83.7% at £30,000). High-dose HD is the sponsor's product.	As a lifetime cost-per-QALY model, the claims are impossible to evaluate. No attempts were made by the authors to generate short-term testable claims. There is no guidance for possible evaluation, and provision of feedback for policymakers as a national policy, for at home versus in-center HD hemodialysis under alternative reimbursement schemes.
Carlson et al.[42]	Cost-effectiveness of tocilizumab (TCZ) monotherapy versus adalimumab (ADA) in persons with rheumatoid arthritis for whom methotrexate is inappropriate.	Genentech Inc.	Cost-effectiveness of TCZ versus ADA assessed over two time horizons: The treatment initiation phase of 6 months and the patient lifetime. The latter timeframe utilized a patient level simulation model to estimate incremental cost per QALY. EQ-5D scores were mapped from the HAQ scores. One-way and probabilistic sensitivity simulation was used to capture uncertainty. In the 6 month model TCZ cost more than ADA with the ICER ranging from $6,570 per additional achievement of LDAS to $14,265 per additional ACR70 response. In the lifetime model, the incremental QALY gain of 0.04 life years and 0.23 QALYs while increasing cost by $8,532. This produced an ICER of $36,944/QALY for TCZ compared to ADA. The results of the probabilistic sensitivity analysis demonstrated that there is more than 50% probability that TCZ is cost-effective compared to ADA mono if threshold is $40,000/QALY. The probability that TCZ is	While there is a potential for generating evaluable claims from the 6 month model, this was not explored. As a lifetime cost-per-QALY model, the claims are impossible to evaluate. There is no guidance for possible evaluation, and provision of feedback for treatment centers.

			cost-effective compared to ADA is 100% at all levels of willingness to pay. The authors concluded that ADA is cost-effective in this patient group. TCZ is the sponsor's product.	
Chen et al.[43]	Cost-effectiveness of rituximab maintenance (MR) and radioimmunotherapy consolidation (RIT) versus observation in progression free survival following frontline therapy in follicular lymphoma (FL).	Partial support by Spectrum	Markov model lifetime cost and QALYs for the MR and RIT treatments compared to observation for those with advanced stage FL. Health states defined were: before first progression, first progression, second progression and death. Health utility estimates from the published literature. Primary analyses of effectiveness and costs were compared within each clinical trial. Compared to observation QALY gains for both MR and RIT were in the range 1.026 to 1.399. Incremental costs per QALY gained were in range $37,412 to $40,851.Both MR and RIT demonstrated favorable and similar cost effectiveness profiles.	While there is a potential for generating evaluable claims from the trials (e.g., progression free survival) identified in the study this was not explored. As a lifetime cost-per-QALY model, the claims are impossible to evaluate. There is no guidance for possible evaluation, and provision of feedback for treatment centers.

References

[1] Langley P. The status of modeled claims. *J Med Econ.* 2015;18(12):991-992

[2] Schommer J, Carlson A, Rhee G. Validating pharmaceutical product claims: questions a formulary committee should ask. *J Med Econ.* 2015;18(12):1000-1006

[3] Belsey J. Predictive validation of modeled health technology assessment claims: lessons from NICE. *J Med Econ*. 2015;18(12):1007-1012

[4] Wasser T, Haynes K, Barron J, Cziraky M. Using 'big data' to validate claims made in the pharmaceutical approval process. *J Med Econ.* 2015;18(12):1013-1019

[5] Langley PC. Imaginary Worlds: Modeled claims for cost-effectiveness published in Pharmacoeconomics January 2015 to December 2015. *INNOVATIONS in Pharmacy*. 2016;7(2): No.2

[6] Langley PC. Imaginary Worlds: Modeled claims for cost-effectiveness published in the *Journal of Medical Economics*. January 2015 to December 2015. *INNOVATIONS in Pharmacy*. 2016;7(2):Article 9

[7] Langley PC. Great Expectations: Cost-utility models as decision criteria. *INNOVATIONS in Pharmacy*. 2016;7(2):Article 14

[8] National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE, April 2013

[9] Langley PC. Sunlit Uplands: The genius of the NICE reference case. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 12

[10] Langley PC. Na domhain shamhlaíochta: formulary submission guidelines in Ireland and the standards of normal science. *Curr Med Res Opin*. 2016;32(5): DOI:10.1080/03007995.2016.1190699

[11] Langley PC. He ao pohewa: The PHARMAC Prescription for Pharmacoeconomic Analysis in New Zealand and the standards of normal science. *INNOVATIONS in Phar*macy. 2016;7(2):No 7

[12] Academy of Managed Care Pharmacy. Format for Formulary Submissions (Version 4). AMCP: April 2016.

[13] Langley PC. Modeling Imaginary Worlds: Version 4 of the AMCP Format for Formulary Submissions. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 11.

[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] Husereau D, Drummond M, Petrou S et al. Consolidated health economics evaluation reporting standards (CHEERS) statement. *J Med Econ*. 2013;16(6):713-19

[16] Editorial. *Nature*. 2016;529:256.

[17] Camerer CF, Dreber A, Forsell E et al. Evaluating replicability of laboratory experiments in economics. *Science*. 10.1126/science.aaf0918 (2016)

[18] Kassirer JP, Angell M. The Journal's policy on cost-effectiveness analysis (Editorial). *NEJM*. 1994;331:669-670

[19] Wootton D. The Invention of Science. New York: Harper Collins, 2015

[20] Grayling AC. The Age of Genius: The seventeenth century & the birth of the modern mind. New York: Bloomsbury, 2016.

[21] Popper KR., The logic of scientific discovery .New York: Harper, 1959.

[22] Lakatos I, Musgrave A (eds.). Criticism and the growth of knowledge. Cambridge: University Press, 1970.

[23] Popper KR. Objective knowledge (Rev. Ed.) Oxford: Clarendon Press, 1979.

[24] Weinstein MC, Fineberg HV. Clinical Decision Analysis. Philadelphia: Saunders, 1980.

[25] Drummond MF, Sculpher MJ, Torrance GW et al. Methods for the economic evaluation of health care programmes. (3rd Ed.) Oxford: University Press, 2005.

[26] Wittgenstein L. Philosophical Investigations. Oxford: Blackwell, 1953.

[27] Editorial. *Machine Design*. 13 September 2007.

[28] Danese MD, Masaquel A, Santos E et al. Estimated life-years saved in women with HER2-positive metastatic breast cancer receiving first-line trastuzumab and pertuzumab in the United States. *Value Health*. 2015;18:876-83

[29] Ventuani S, Nilsson J, Bergman B et al. A cost-effectiveness analysis of minimally invasive versus open surgery techniques for lumbar spinal fusion in Italy and the United Kingdom. *Value Health*. 20155; 18:810-16

[30] Legrand G, Ruscio L, Benhamou et al. Goal-directed fluid therapy guided by cardiac monitoring during high-risk abdominal surgery in adult patients: cost-effectiveness analysis of esophageal Doppler and arterial pulse pressure waveform analysis. *Value Health*. 2015;14:605-13

[31] Nguyen K-h, Gordon LG. Cost-effectiveness of repetitive transcranial magnetic simulation versus antidepressant therapy for treatment-resistant depression. *Value Health*. 2015;18:597-604

[32] Collins CD, Schwemm AK. Linezolid versus vancomycin in the empiricx treatment of nosocomial pneumonia: A cost-utility analysis incorporating results from the ZEPHyR trial. *Value Health*. 2015;18:614-21

[33] Ting J, Ho T, Xiang PO et al. Cost-effectiveness and value of information or erlotinib, afatnib, and cisplatin-pemetrexed for first-line treatment of advanced EGFR mutation-positive non-small-cell lung cancer in the United States. Value Health. 2015;18:774-82

[34] Nazir J, Maman K, Neine M-E et al. Cost-effectiveness of mirabegron compared with antimuscarinic agents for the treatment of adults with overactive bladder in the United Kingdom. *Value Health*. 2015. 18:781-90

[35] Bermingham SL, Hughes R, Fenu E et al. Cost-effectiveness analysis of alternative antiviral strategies for the treatment of HBeAg-positive and HBeAg-Negative chronic hepatitis B in the United Kingdom. *Value Health*. 2015;18:800-09

[36] Dhankar P, Nwankwa C, Pillsbury M et al. Public health impact and cost-effectiveness of hepatitis A vaccination in the United States: A disease transmission dynamic modelling approach. *Value Health*. 2015;18:358-67

[37] Waure C de, Capri A, Veneziano MA et al. Extracorporeal photopheresis for second-line treatment of graft-versus-host diseases: Results from a health technology assessment in Italy. *Value Health*. 2015;18:457-66

[38] Linssen AM, Anteunis LJC, Joore MA. The cost-effectiveness of different hearing screening strategies for 50- to 70- year – old adults: A Markov model. *Value Health*. 2015;18:560-69

[39] Visser MS, Zonneveld LNL, Spijker A van et al. The cost-effectiveness of cognitive-behavioral group training for patients with unexpected physical symptoms. *Value Health*. 2015;18:570-577

[40] Simons PAM, Ramaekers B, Hoebers F et al. Cost-effectiveness of reduced waiting time for head and neck cancer patients due to lean process design. *Value Health*. 2015;18:587-96

[41] Xiaoqing F, Treharne C, Arici M et al. High-dose hemodialysis versus conventional in-center hemodialysis: A cost-utility analysis from a UK payer perspective. *Value Health*. 2015;18:17-24

[42]Carlson JJ, Ogale S, Dejonckheere F et al. Economic evaluation of tocilizumab monotherapy compared to adalimumab monotherapy in the treatment of severe active rheumatoid arthritis. *Value Health*. 2015;18:173-79

[43] Chen Q, Ayer T, Nastoupil LJ et al. Comparing the cost-effectiveness of rituximab maintenance and radioimmunotherapy consolidation versus observation following first-line patients with follicular lymphoma. *Value Health*. 2015;18:189-97