INNOVATIONS in pharmacy

Volume 8 | Number 4 Article 7

12-1-2017

Resolving Lingering Problems or Continued Support for Pseudoscience? The ICER Value Assessment Update

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

Resolving Lingering Problems or Continued Support for Pseudoscience? The ICER Value Assessment Update

Paul C Langley, PhD
Adjunct Professor, College of Pharmacy, University of Minnesota

Abstract
The Institute for Clinical and Economic Review (ICER) released its updated value assessment framework in mid-2017. This included refinements to its conceptual structure and modifications to methods of collecting and assessing evidence. Consequent to this release, a number of authors have commented on the updated value framework, addressing the question of whether the latest framework represents a major revision or merely attempts to resolve lingering problems. The purpose of this commentary is twofold: (i) to revisit what are considered to be fundamental flaws in the ICER value assessment framework and (ii) to question whether or not post-release critiques of the value framework address the fundamental weaknesses in the ICER approach: the absence of credible, evaluable and replicable claims for the benefits and harms of a therapy intervention. The commentary argues that while ICER sees the purpose of its value assessment framework as forming 'the backbone of rigorous, transparent evidence reports' in placing 'scientific methods of evidence analysis at the heart of a clearer and more transparent process' it falls far short of these ideals. Rather, in attempting to replicate in the US health care environment the evaluation framework mandated by the National Institute for Health and Care Excellence (NICE) in the UK, the ICER falls into the trap of generating value claims for product impact that fail to meet the standards of normal science.

Keywords: ICER, pseudoscience, economic evaluations, imaginary worlds, pseudoscience, simulations

Conflicts of Interest: I declare no conflict of interest or financial interest that I or members of my immediate family have in any product or service discussed in the manuscript, including grants (pending or received), employment, gifts, stock holdings or options, honoraria, consultancies, expert testimony, patents and royalties.

Introduction

Over the past 18 months a number of commentaries have been published in *INNOVATIONS in pharmacy* pointing to the absence of standards that meet the those of normal science for cost-effectiveness and budget impact claims [1] [2]. This criticism applies both for pharmaceutical products and devices. Modeled claims for product outcomes are presented that: (i) lack credibility, (ii) are not evaluable and (iii) are not replicable. Professional associations such as the Academy of Managed Care Pharmacy (AMCP) and the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), a large number of single payer health system agencies and editors of leading health technology assessment journals, actively encourage and set standards for the construction of imaginary worlds to support claims for cost-effectiveness [3] [4]. This health technology assessment 'meme' is well entrenched. Over the past 30 years, we have seen the publication of literally thousands of cost-effectiveness modeled claims and health care decisions for formulary access and pricing based on these standards, which have little if any right to be taken seriously. Constructed as imaginary worlds, focusing on lifetime cost-per quality adjusted (QALY) simulations, we have no idea whether the claims are right, whether they are wrong and, in the case of the ever popular lifetime cost-per-QALY model, we will never know. As detailed in these commentaries, judged against the standards of normal science, lifetime cost-per-QALY models that follow ISPOR recommendations for 'good research practice', that embrace the National Institute for Health and Care Excellence (NICE) reference case, are best seen as pseudoscience.

The recently released Institute for Clinical and Economic Review (ICER) updated value assessment framework falls squarely within this technology assessment tradition [5]. Claims for cost-effectiveness and pricing are generated by modeled frameworks that fail to meet the standards of normal science. They fail the standard for claims to be credible, evaluable and replicable. Unfortunately, while a number of authors have presented critiques of the updated ICER value framework, the criticisms presented fail to address the limitations implicit in ICER's adoption of a flawed modeling methodology.

This is not the first time that the commentaries in *INNOVATIONS in pharmacy* have addressed the limitations of the ICER value assessment framework. Four commentaries

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

have been published on ICER reports: heart failure (Entresto), multiple sclerosis, rheumatoid arthritis and cholesterol reduction (PCSK9 inhibitors) [6] [7] [8] [9]. Judged by the number of PDF downloads recorded by the University of Minnesota library system, these commentaries have been popular. To date, out of 2,481 PDF downloads for the 28 commentaries published to date (18 November 2017), some 619 (25%) have been for the four ICER reports. The ICER heart failure review has been the most popular with 329 PDF downloads.

The purpose of the present commentary is to revisit the ICER value assessment framework, to include the recent Neumann and Cohen critique of the updated value assessment framework and its claim for legitimacy as the leading standard for drug value evaluation. [10]. Although the ICER might be seen as a force to be reckoned with in health technology assessment, triggering alarm bells in the C-suites of pharmaceutical companies, as Neumann and Cohen put it, this commentary puts the case that this alarm is misplaced. Certainly, manufacturers should be held accountable for claims made for their products. However, seen from the perspective of the standards of normal science rather than from the standards of health technology assessment models and simulations, the ICER assessment framework should be put to one side. It may continue to interest those committed to the construction of non-evaluable claims generated by lifetime cost-per-quality adjusted life year (QALY) models. Unfortunately, we have no idea, and will never have any idea, as to whether the claims made are right or if they are wrong. Even if these modeled claims are justified as being 'for information only', formulary committees would be well advised to reject model based recommendations, relying instead of a technology assessment methodology that generates credible, evaluable and replicable claims for product acceptance, placement and pricing.

ICER Critiques

Critiques of the ICER value assessment process can be usefully categorized as (i) those that accept the underlying lifetime simulation methodology as the gold-standard for health technology assessment and formulary decisions as opposed to, as in the present case (ii) those who reject lifetime simulations and the construction of imaginary worlds in accepting only those claims for competing products which meet the standards of normal science for hypothesis testing and falsification.

It is important to keep in mind that Neumann and Cohen are committed along with ICER to a core health technology assessment methodology. This puts center stage the construction of lifetime modeled simulations as the basis for comparative product evaluation; a shared commitment to the construction of imaginary worlds. This is see in their endorsement of refinements proposed by ICER: (i) presenting budget impact claims based on a range of product uptake assumptions; (ii) using net drug prices rather than list prices; (iii) increased scrutiny of heterogeneity of treatment effects; (iv) additional measures of clinical benefit; (v) applying lower health utilities for individuals with chronic, severe conditions; and (vi) paying more attention to patient context (e.g., severity of a condition, care giver burden).

Even so, there are lingering concerns. The most prominent of these for Neumann and Cohen is the retention of the threshold budget impact calculation that allows a ceiling expenditure of $915 million per annum going forward. If this is exceeded in the ICER model calculation then the 'short-term affordability' price should be reduced to compensate. Their objections are not to a budget cap per se, but to the methodology employed. ICER should not be assessing how much payers should spend on a drug; that is the payer's prerogative. Rather, ICER should focus on estimating and disseminating net budget impact forecasts of the impact of new drugs under a range of price, uptake and horizon assumptions.

The second lingering problem is the narrow healthcare perspective taken in the ICER base-case cost-effectiveness case. While this may appeal to payer audiences, ICER should put alongside this a comprehensive societal impact model. This, presumably, would take a lifetime horizon and, once again, generate non-evaluable claims, this time for the societal impact.

The final lingering concern is with the use of cost-effectiveness thresholds. However, it is not the value of the threshold that is the problem, it is the introduction of thresholds into the ICER value assessment in the first place. ICER seems to see itself (mistakenly to the reviewer's minds) as the self-appointed decision maker and not, more appropriately, as an evaluation agency or consultant. It is not ICER's role, unlike the role of NICE in the single payer UK health system, to determine what tradeoffs payers and patients should be willing to make.

Whether these 'lingering concerns' will have any substantive impact on how ICER presents its results and recommendations is a moot point. After all, the terminology itself suggests that these issues are not that important and probably can be resolved by a more extended presentation of modeled scenarios. A more substantive 'lingering concern' might be whether the claims ICER value assessment framework is of any interest to health care decision makers? After all, as Neumann and Cohen conclude:

> *An inherent flaw is that ICER is striving to provide a public good in a fractured US healthcare market, where its private payer audiences have short time horizons and practical constraints that often do not align with societal value.*

If this is the case then the more pertinent argument, which goes to the acceptance of what is seen here as a flawed technology assessment methodology, is whether heath care decision makers are even interested in the ICER value assessment framework. In the wider societal context, whether the construction of imaginary worlds, which the authors accept uncritically, is the gold standard for cost-effectiveness claims?

Note should also be taken of the comments offered by the National Pharmaceutical Council (NPC) on the updated value assessment framework. Again, the NPC accepts the underlying technology assessment methodology[11]. Continuing concerns are: (i) the need to provide transparency through full access to the model so that results can be reproduced and tested; (ii) the continued use of the $50,000 cost-per-QALTY threshold in voting on ICER recommendations; and (iii) the need to eliminate affordability assessments. At the same time the ICER is complimented on its intention to include a 'cost per consequence' measure in its reports.

NPC concerns are also voiced in their recent comments on a proposed (announced) collaboration between the ICER and the Veterans Administration (VA) [12]. Issues raised include: (i) the need to broaden the evidence base for value assessments beyond RCTs in post-product approval analysis and research in the 'unique' VA population; (ii) accommodating the perspective of veterans and their families in complex conditions such as post-traumatic stress disorder (PTSD), loss of limbs, chronic pain and return to work; (iii) capturing a wider set of value inputs and not link benefits to as single value assessment; and (iv) avoid imposing arbitrary spending limits on drugs.

Credibility, Evaluation and Replication

Unless health care decision makers and, it must be added, pharmaceutical and device manufacturers are prepared to put to one side the currently accepted heath technology assessment commitment to base cost-effectiveness claims on imaginary lifetime constructs, then groups such as ICER will continue to flourish. ICER's access to resources, its effective management, its acceptance by key players and its independent status, will ensure its continued dominance of the health technology assessment space. This should not be taken as a criticism of ICER in its efforts to influence formulary recommendations, placement and pricing. Rather, it is a criticism directed at ICER's acceptance of a flawed technology assessment 'meme' that puts to one side the standards of normal science. If ICER was prepared to focus on claims that were credible, evaluable and replicable then its efforts could be supported. The question is, of course, whether ICER (and groups such as ISPOR and the AMCP) would be prepared to abandon their present position.

For those who have followed the commentaries published in *INNOVATIONS in pharmacy* over the past 18 months, the limitations of the current health technology assessment 'meme' should be well known. For present purposes, as the focus is again on the ICER value assessment framework it is worth restating the key arguments. These can be considered under the following heads:

- Standards of normal science
- Replication and clinical claims
- Claim timeframe
- Willingness-to-pay-thresholds
- Constructing QALYS
- Adherence and persistence
- Pricing and costs
- Potential confounding factors
- Access to the ICER model
- Immunity to feedback

Standards of Normal Science

The standards of normal science, as they should apply in health technology assessment, are aptly summarized by the motto of the Royal Society (founded 1660; Royal Charter 1662): *Nullius in verba* (take no man's word for it). If a modeled cost-effectiveness claim is to meet these standards then it has to: (i) involve the construction of an empirically evaluable, coherent theory and (ii) facilitate the testing of hypotheses through experimentation of observation. A theory is not to be judged by the realism of its assumptions. To argue that a lifetime cost-per-QALY based claim is to be accepted as an input to formulary decision making on the grounds that the simulation is 'realistic' is unacceptable. Not only is it impossible to claim that a simulation projecting forward 20 or 30 years is realistic but, as detailed below, health technology assessment standards in respect of pricing and compliance, ensure that it is not intended to be 'realistic'. Should the claim be that a model is realistic, but not too realistic? Perhaps ICER could collaborate with ISPOR to construct a 'lifetime model realism' scale to rank models where a multiattribute weighting yields a score on a range from 0 = not realistic to 1 = completely realistic?

The standards for hypothesis testing through experimentation and observation have been in place since the 17th century [13]. They demarcate science from pseudoscience or, as more strongly stated by Pigliucci, they demarcate natural selection from intelligent design [14]. Failure to accept and meet the standards for experimentation and observation, the core feature of health technology assessment models such as those central to the ICER value assessment framework, effectively put claims made outside the realm of normal science. This criticism applies equally to the ICER as well as single payer agencies such as NICE in the UK, the PBAC in Australia and

PHARMAC in New Zealand. Of course, the fact that NICE embraces the reference standard of constructing imaginary worlds does not mean that others should follow.

It is worth noting that this rejection of the standards of normal science in health technology assessment is recognized and endorsed by practitioners. The position taken in the Canadian health technology guidelines, for example, is quite explicit: *Economic evaluations are designed to inform decisions. As such they are distinct from conventional research activities, which are designed to test hypotheses* [15]. Taken a face value, therefore, health technology assessments in 'informing' decisions on products and devices are not concerned with the benefits and harms that might be evaluated in treatment practice, or on the belief or otherwise in the claims from randomized clinical trials (RCTs) and whether they can be replicated, but in the construction of simulations (imaginary worlds) which are designed to generate non-evaluable claims. In respect of this last point, it should come as no surprise that the Canadian Agency for Drugs and Technologies in Health (CADTH) embraces lifetime cost-per-quality adjusted life year (QALY) models as the gold standard [16].

Replication and Clinical Claims

A feature that the ICER has in common with single payer assessment agencies and, in the US, formulary committees, is the willingness to take clinical claims from Phase 2/3 randomized clinical trials (RCTs) at face value. Rather than taking a critical perspective on the validity of these claims and the extent to which the claims have been replicated, the analysis typically proceeds to an indirect modeling comparison to rank competing products in disease and therapy areas from a limited evidence base, driven by modeled indirect comparisons. It is always important to remember that indirect comparisons and the ranking of interventions are only 'as good' as the technique employed; the question of claims replication is the critical issue.

Clinical claims from pivotal RCTs are typically the primary input to indirect product comparisons. Given that these claims are then extrapolated over the lifetime of the patient in the value assessment simulation, it seems odd that ICER puts to one side any attempt to validate these clinical claims or to suggest how these claims might be validated either through head-to-head RCT comparisons or, more to the point, an assessment of these claims in clinical practice. This neglect stands in contrast to the standards proposed in the Minnesota formulary submission guidelines [17] [18]. In the Minnesota guidelines manufacturers are asked to submit, as part of their product dossier, a protocol detailing how the claims made are to be evaluated in treatment practice with the results reported back to the assessment formulary committee. Feedback, as detailed below, is seen as a critical element in formulary acceptance, placement and pricing. Manufacturers are not asked to put forward lifetime simulations to support non-evaluable incremental cost-outcomes claims. Rather, the focus is on clinical and cost-outcomes claims that are credible, evaluable and replicable.

It is worth noting that the requirement for credible claims and a protocol to support those claims is not new. This approach was proposed over 12 years ago in the development of draft guidelines to support formulary committee evaluations of new products as well as disease area and therapeutic class reviews in the WellPoint (now Anthem) health system outcomes based formulary [19] [20].

Claim Timeframe

The position taken by ICER, and one, as noted above, that is entirely consistent with standards advocated by the AMCP, ISPOR and agencies such as NICE, is that 'the grounding of any evaluation of value should recognize the long-term perspective on both outcomes for patients and costs'. This is achieved by simulations that 'estimate outcomes and costs at the longest feasible time horizon, usually the full lifetime of patients'. In practical terms this means, presumably, that claims for incremental cost-effectiveness are appropriately based on simulations, driven by selected assumptions, which can extend over 20, 30 or even more years. Simulated benefits which might only be realized over many years are thus a 'core' element in the value framework.

Claims based on these simulations, together with the discounting of the simulated stream of benefits and costs, are clearly not intended to be evaluable. Indeed, given the inevitable entry of new products within this timeframe, long-term benefits and harms would presumably accrue to the succession of new products, given patient switching behavior, and not to the present product configuration which is, by assumption, to remain unchanged with patient's adherent to this product over their lifetimes. At the same time, within this modeling framework considerable attention is given to justifying and the construction of health states through which the modeled target population are assumed to progress. This is supported by ISPOR with a range of publications detailing what they see as professional standards for constructing imaginary worlds, together with imaginary world workshops presented regularly at their various global conferences [21] [22] [23]. This level of commitment and reinforcement has guaranteed that the health technology assessment framework is well entrenched. Or, to put it in the language of memes and memeplexes, the role of ISPOR and other technology assessment agencies as high fidelity replicators [24]

Willingness-to-Pay Thresholds

Although it would be unfair to characterize published models that claim cost-effectiveness for their sponsor's product as nothing more than marketing exercises, it is intriguing that all too often the model generates claims that fall within a

willingness-to-pay threshold, and as such the product is claimed to be cost-effective. This has been well documented in commentaries in the *INNOVATIONS in pharmacy* that have reviewed cost-effectiveness studies published in *Value in Health*, *PharmacoEconomics* and the *Journal of Medical Economics* [25] [26] [27].

It is also instructive to review for individual products the NICE assessment process in the UK where a manufacturer, given the mandatory reference case guidelines for constructing imaginary worlds, submits a modeled claim that is then passed on to a review group who proceed to dissect, restructure and present their version of the modeled imaginary world. This is further evaluated by NICE and, eventually, an agreed future world is presented. Willingness-to-pay thresholds (usually £20,000 per quality adjusted life year) are applied and, with some exceptions, NICE approval of the product. The ICER follows a similar process (NICE-lite) with the exception being that while the model may be constructed by an academic group under contract to ICER, the value assessment model is not put out for independent review. There is certainly a public comment phase following the ICER draft report, but given the time frame involved there is little scope for those comments from stakeholders to be taken on board. The key issue is the absence of a well-documented independent review of the ICER value assessment reporting in the public domain. This, of course, assumes that the participants are in agreement on the contribution of constructing imaginary worlds.

Assuming a specific QALY measure is mandated (in the case of NICE the EQ-5D) with the single payer's agency also mandating willingness-to-pay thresholds, then the there is no debate over whether or not the product, by definition, is cost-effective. This is not, as Neumann and Cohen point out, the case with ICER as it is not the agent of the various US health payers who have, let us assume, agreed on modeled willingness-to-pay thresholds. Assuming, further, that willingness-to-pay is relevant to decision making in the US, they cannot claim a product is cost-effective (with recommendations for price discounting) unless the decision maker concurs with (i) the application of the specific ICER lifetime cost-per-QALY model as one among many modeled options; (ii) the application of an agreed QALY measure; and (iii) the relevance of the ICER willingness-to-pay threshold (or thresholds). At best, as Neumann and Cohen point out, ICER is an evaluator not a self-appointed decision maker.

Unfortunately, in the case of NICE, moving from the early EQ-5D-3L to the later EQ-5D-5L version of the instrument a slight problem has arisen which rather undercuts the claim for a robust and invariant QALY measures as a reference case. A recent report by Wailoo et al of the Decision Support Unit at the University of Sheffield compared response distributions between the 3 level and 5 level versions of the instrument in an assessment of the ability to map between the two measures [28]. The report found significant statistical differences in the covariates and latent factors between most dimensions of the two instruments and, as a result, moving from three levels to five levels 'is just not just a uniform realignment'. As a result, the two versions of the instrument 'produce substantially different estimates of cost-effectiveness'. While it is possible to map between the two instruments, this may be at the expense of claims for the instrument as a generic measure, with Wailoo et al pointing out that mapping may be only appropriate within disease areas. At the same time, as noted in previous commentaries, there are recent claims that the preferences expressed over hypothetical health states are inconsistent with the assumptions of the multiattribute utility theory that underpins measures such as the EQ-5D [29].

The fact that the US is a diverse ('fractured') health care system with multiple decision makers, makes any broad claim for cost-effectiveness in the US impossible. A situation that is made more complex (if not absurd) by the fact that individual decision makers would have had no input whatsoever into the ICER model framework, even to the extent of denying the relevance of a lifetime cost-per-QALY model and thresholds as cutoffs. A situation that remains unchanged with proposed supplementary cost-per-outcome measures. What is the threshold value for cost-per-event avoided? Can diverse health systems agree on a model framework that supports lifetime cost-per-outcome claims? Finally, it should be noted that under the Patient Care and Affordable Care Act (ACA), the Patient Centered Outcomes Research Institute (PCORI) that cost-per-QALY or similar measures cannot be used 'to determine coverage, reimbursement, or incentive programs under Title XVIII' [30].

The US is not, however, alone in putting cost-per-QALY claims to one side. While the ICER defends its position by pointing out that this standard has been accepted by many academics, manufacturers and patient groups, it should be pointed out that the acceptance is not universal. Germany, for example, is not interested in QALYs and they have been rejected by the Spanish Ministry of Health [31] [32]. In New Zealand, while cost-per-QALY claims are accepted, the Pharmaceutical Management Agency (PHARMAC) has rejected the application of thresholds. The argument is that given the range of criteria that contribute to formulary decisions and the requirement in New Zealand that pharmaceuticals are kept within a fixed budget, thresholds are inappropriate[33]. As budgets vary from year-to-year, there is no threshold below which a pharmaceutical is considered cost-effective.

There is the real possibility that technology assessment decisions may have to abandon cost-per-QALY assessments as the universal gold standard for resource allocation and formulary assessment in favor of more disaggregated, disease specific measures. Even if ICER continues to argue for a cost-per-QALY value assessment, there are concerns that in

constructing incremental cost-per-QALY simulations with QALY measures that are taken from the literature, the measures captured in the simulation may not be compatible across the products assessed or for comparisons between ICER reports for different products in a range of disease areas. Threshold cost-per-QALY claims are only relevant, assuming that the QALY threshold is accepted, if resources are allocated across disease areas utilizing a common QALY instrument. ICER should commit, presumably, to a common QALY measure used across all its value assessments.

It is worth noting that the ICER updated value assessment rejects multicriteria decision analysis (MCDA) because it does not believe 'that the methods for weighting individual elements are robust enough' to ensure the reliability of value judgements . It has found the technique 'too complicated for reliable use'. A criticism that might equally well apply to its modeling for value assessment thresholds and budget impacts.

Constructing QALYS

It is possibly surprising that, given the lack of interest in the US and the outright bar on the application of cost-per-QALY based claims and thresholds under the ACA that the ICER perseveres with a value assessment framework that has, as its centerpiece for pricing recommendations, cost-per-QALY thresholds generated by imaginary worlds. Given this, it is important also to note that the QALY measures utilized in lifetime cost-per-QALY claims are often cobbled together from the literature. Literature sources are culled for utility measures, covering both generic and disease specific measures in an attempt to generate 'acceptable' measures to capture the health states of the constructed target population as it progresses over its lifetime health state experiences. Heath states which, in turn, are literature based constructs. This practice is defended and any criticisms are deflected by the model builder arguing for input parameter distributions and various sensitivity analyses to support claims expressed in probabilistic terms.

Adopting the QALY as the outcome gold standard, of course, effectively precludes any empirical evaluation of clams made for competing products. In the US (and in other jurisdictions) QALYs are not regularly generated as part of patient electronic medical records (EMRs) or as part of administrative data sets whether these are from physician practices or hospital records. Indeed, even if there was an effort to collect QALY measures, there would be the further question of which QALY measure, generic or disease specific, to collect. A situation that is made even more absurd by the difficulty of crosswalking one QALY measure to another. As noted, it seems as if the intent of those advocating standards for health technology assessment have gone out of their way to ensure that claims made, with the lifetime cost-per-QALY (NICE reference case) paradigm are impossible to evaluate [34].

Adherence and Persistence

A question that ICER needs to address in each of its modeled lifetime value assessments is compliance with therapy. There is now abundant evidence that few patients are either adherent to or persistent with therapy in the long-term; that is, beyond two or three years. If this is the case, as pointed out some time ago in the commentary on the ICER heart failure analysis, if patients are not persistent with therapy why do simulations continue to assume compliance over the lifetime of the patient? This neglect is also puzzling given the attention to product uptake assumptions proposed by the ICER is this revised value assessment. This is not always the case. ICER in its recent value assessment of the VMAT2 inhibitors includes discontinuation of therapy in modeling tardive dyskinesia [35] . However, this appears to be the exception rather than the rule. The obvious question is why this is not introduced across the board.

It is certainly possible to introduce assumptions regarding adherence and persistence with therapy into lifetime simulation models. Whether it is worth the while is debatable. A more useful approach, if the focus is on a commitment to generating evaluable claims, would be to focus on the short term. With access to an ICER model it would be entirely possible to reformulate the model to accommodate compliance behavior. This, from ICER's perspective would be risky as it could change completely claims for incremental cost-effectiveness to those based on short term QALY outcomes (assuming that QALYs measured over a short time horizon are meaningful). If, for example, the claim made is that the benefits to patients only accrue in the long term *to the minority who are compliant with therapy over this timeframe* then any long-term incremental cost-effectiveness claim would be qualified by outcomes for those patients who dropped out from therapy. This situation would be made more complex if assumptions were made as to differential compliance behavior across products in the therapeutic area. Perhaps ICER could consider as a standard element of their value assessment evaluable modeled claims for each of the first 3 years following index prescription.

Pricing and Costs

The ICER modeling of cost-effectiveness 'will not routinely make estimates of price changes across comparator treatments linked to patent and exclusivity time horizons'. As pointed out in the ICER commentaries in *INNOVATIONS in Pharmacy*, to assume the absence of pricing changes in pharmaceuticals (and other medical resource inputs) is rather odd given the evidence for manufacturer's pricing policies to regularly increase unit price, often far in excess of any medical price inflation index. ICER's defense is that 'assumptions about price changes are not currently the standard in health technology assessment agency cost-effectiveness analyses ' as it is 'very difficult to predict the pricing landscape many years into the future …'. There is no disagreement with that point although it might be added that (i) if price increases are built

into the imaginary world(s) scenario, there is no reason to assume there will be a consistent pricing policy across products in a disease or therapy area and (ii) pricing decisions may respond to ICER recommendations for discounting. A more believable response might be: if this is the case, why not admit the virtual impossibility of long-term simulations that might conceivable bear any relation to a possible future reality and focus, as recommended here, on short-term credible claims with feedback to decision makers? After all, if you claim that your lifetime simulation should be realistic why put to one side pricing assumptions which are, on the evidence for past pricing behavior, an integral part of any cost-effectiveness simulation?

Potential Confounding Factors

Again, there is now ample evidence that the response to therapy within a target patient population depends on a range of potential confounding factors. These may include comorbidities such as diabetes and cardiovascular disease in older populations, the presence of depression, anxiety, sleep disorders and pain, as well as access to care and employment status. While ICER may consider these as 'contextual considerations' which may be put to one side consequent to a vote by 'stakeholders' as to their potential benefits and disadvantages, the concern is that the modeling framework is ever further distanced from any claims for realism in its future scenario-driven reality.

Creating lifetime simulation models that deliberate exclude factors such as compliance, pricing and costs and a range of other confounding factors must raise doubts again as to whether or not the commitment to a lifetime technology assessment 'meme' is a worthwhile endeavor. The defense that 'this is what everyone does' is not acceptable. It puts to one side any notion of progress in health care decision making in favor of adhering to and attempting to further refine what might be considered a bankrupt methodology. One is reminded of the question of why a driver who has lost his car keys at night only looks in the area around the lamppost: the response being – " that is where the light is".

Access to the ICER Model

The question of whether or not access to the ICER cost-effectiveness model might resolve issues depends on the underlying structure and assumptions of the model itself and the belief reviewers have in the merits of the lifetime model exercise.. If the model is designed to generate non-evaluable claims then, unless by dint of major restructuring, it is doubtful if one can translate the model into one that is capable of generating credible and evaluable claims. If so, then access to the model is clearly a waste of time. All access allows is to simply rearrange deckchairs. Competing claims may emerge, but these will still be variations on a theme: the claims are still non-evaluable. Presenting external reviews with access to a lifetime cost-per-QALY model which, within presumably limits set by ICER on the ability to challenge or manipulate the underlying structure of the model means we are still left with a model that fails to meet the standards of normal science.

Understandably, ICER appears reluctant to release their model to independent assessors or to health system decision makers. The problem is, of course, that there can be, as demonstrated in the literature, a plethora of models, each sponsored by competing interests. Releasing the ICER model to public scrutiny runs the risk of encouraging competing models with the result that, from the payers' perspective in a fragmented health system, decisions would have to be made across competing imaginary futures. It is entirely possible that with model competitions, one class of models might incorporate pricing projection and assumptions regarding adherence and persistence behavior while another class, including the ICER model, excludes these. A situation, it might be noted that has been addressed in the *INNOVATIONS in pharmacy* commentaries for lifetime simulation models in diabetes mellitus. Taking its cue from the Mt Hood challenge meeting in diabetes modeling, ICER might consider sponsoring model competitions in specific disease areas as a prelude to releasing a value assessment report [36].

Immunity to Feedback

Of course, if ICER was prepared to concede that there is a need to develop modeling frameworks that were capable of generating short-term evaluable claims that could be driven by a protocol that generated feedback to a formulary committee, then the issue of competing imaginary worlds might be usefully resolved. Unfortunately, this is most unlikely to occur given ICER's commitment to a NICE-lite lifetime cost-per-QALY methodology. This is made quite clear in the revised value assessment framework which is, to make the point once again, only one of a potential range of competing value assessment frameworks that could be sponsored by other independent agencies.

If manufacturers, pharmacy benefit managers and health care systems are to take ICER claims for cost-effectiveness, pricing and budget impact at face value, then it is presumably incumbent upon ICER to provide a value framework that generates feedback to decision makers. Unfortunately, this is not the case. Indeed, the ICER value framework would appear to be designed to exclude effectively any feedback to decision makers. Decision makers are forced to either accept or reject the ICER lifetime modeled recommendations and their claims for comparative benefits and harms. There is no mechanism which, as a new product or device is introduced to treatment practice, allows decision makers to evaluate the merits of the ICER claims; let alone the validity of the assumptions which are built into the ICER models. While a model, as noted above should not be judged on the validity/reality of its assumptions, there are critical assumptions, notably in respect of adherence and persistence that if excluded cast doubt on the overall merit of the exercise.

The current ready access to internet-based reporting platforms to support RCT and observational studies at low cost and in real time provide a vehicle for feedback on evaluable claims in a timeframe relevant to health care decision makers, physicians and patients. Protocols can be implemented in target populations where, as detailed below, patients can report on-line in real time with these responses supported by reports from the electronic medical record (EMR). The ready access to these platforms gives further support to the development of short-term modeled claims that meet the standards of normal science. Hypotheses can be tested through experimentation and observation. There is also feedback to model builders who can assess the strengths and weaknesses of their models, setting the stage for more robust model frameworks.

Conclusions

Once a commitment is made in incremental cost-outcomes analysis to a lifetime perspective any claims made for clinical outcomes, cost-consequences and quality of life are, by definition, non-evaluable. Even if the model builder builds in credible claims that may be evaluated in the short-term, we are still in the position of not knowing whether the ICER lifetime recommendations are right or whether they are wrong; and we will never know. While this commitment to the construction of what are best described as imaginary worlds may be defended in terms that they are not meant to generate testable hypotheses but rather to provide 'information' for decision makers, the decision maker has no basis for either accepting or rejecting the claims. As far as the decision maker is concerned the simulation is best seen as a 'black box' that could equally well be re-engineered to generate competing claims. Arguments that the simulation is 'realistic' are simply not acceptable when the simulation fails to accommodate, for example, assumptions as to anticipated compliance and pricing behavior. Unfortunately, even if such assumptions were built into the model, the lifetime perspective adopted guarantees that any claims made fail to meet the standards of normal science.

Can the ICER position be challenged? Quite clearly it can be as this commentary has outlined. The challenge, however, is not to ICER itself as it can continue to further refine its value assessment framework and the continued modeling of imaginary worlds. The challenge is to the mainstream health technology assessment belief in the information value of lifetime, cost-per-QALY imaginary worlds irrespective of whether these are narrowly payer focused or whether they attempt a wider, yet still imaginary construct, which takes a societal perspective.

The essential point is that those taking up the challenge to this commitment to constructing imaginary worlds accept the standards of normal science: health technology assessment claims must be credible, evaluable and replicable in a timeframe that is relevant to decision makers. But who will mount the challenge? ISPOR and associated professional groups are not candidates: they have too much sunk capital in the 'information vs. hypothesis testing' technology assessment meme. The obvious candidates are, on the one hand, health care systems and their various formulary committees while, on the other hand, pharmaceutical and device manufacturers.

From the formulary committee perspective standards have already been proposed in the Minnesota guidelines. Adoption of these guidelines by a major health system would clearly act as a catalyst in bringing manufacturers in line. A complementary scenario would be where a major manufacturer, possibly acting in tandem with a health system, announced that it was focusing its health technology assessment resources on developing models that, once again, were designed to generate credible, evaluable and replicable claims in a timeframe that is relevant to decision makers. Submissions to healthcare systems, following an unsolicited request for a product dossier would detail and, at the same time, propose how the claims could be evaluated in target patient populations.

From a strategic perspective, the commitment by a manufacturer to abandoning the current, mainstream technology assessment 'meme' in favor of one that met the standards of normal science has few downside risks. Apart from objections from professional and other technology assessment groups, the principal beneficiaries would be decision makers, physicians and patients in health care systems. For the first time a manufacturer would commit to generating claims and assessment protocols that would be evaluated in real time, providing ongoing feedback. This engagement would not, of course, have to be repeated across clients. Given the typical rate of uptake of new products, 'first adopter' health systems could be the vehicle for claims assessment and feedback. These 'first adopters' could reflect the disparate nature of the US healthcare system with evaluations targeted to specific patient groups within systems such as the Veterans Administration, capturing patient characteristics specific to that system.

The first step for a manufacturer would be to announce that it was investing in a technology assessment process that recognized the importance of generating credible and evaluable claims. This would be seen as a natural extension of its commitment to high quality RCTs that, by definition, met the standards of normal science. The shortcomings of the present standards would be detailed, pointing to the real time information vacuum that exists in focusing technology assessment activities on the construction of any number of competing imaginary worlds, where claims for cost-effectiveness and pricing adjustments are not to be taken seriously.

Second, the manufacturer would have to align internal activities to drawing up a set of standards for short term modeling to support clinical, cost-effective and budget impact claims. Protocols for claims assessment, specific to disease and therapeutic area, would be agreed. These elements would be combined in a dossier structure to meet unsolicited requests from health systems.

Third, standards would have to be agreed for protocol implementation and reporting. Given time and resource constraints, observational tracking studies would be the preferred option. The design of the tracking study, which could combine elements from the electronic medical record (EMR) as well as study specific patient questionnaires supported by internet-based platforms, would be detailed and agreed with the health system and physician practices. Finally, a publication strategy would be detailed, to include prior media release and presentations announcing the adoption of this evaluable claims strategy.

From the perspective of meeting the standards of normal science in health technology assessment, it is a moot point as to how sustainable is the present commitment to the construction of lifetime cost-per-QALY models. Certainly, the alternative presented here is achievable. At the same time, it is of interest to speculate on how adoption of this commitment to evaluable claims would resonate outside of the US. As detailed in the *INNOVATION in Pharmacy* commentaries, many single payer systems have mandated lifetime cost-per-QALY models in product and device submissions, and where willingness-to-pay thresholds are in place, their impact for pricing and formulary acceptance. However, the ability to actually track evaluable modeled claims through low cost internet-based reporting platforms offers agencies the option of redrafting their guidelines to focus on evaluation, replication and feedback. Whether this option is taken up remains an open question.

References

[1]*INNOVATIONS in pharmacy*. Formulary Evaluations. https://pubs.lib.umn.edu/innovations/formulary-evaluations/

[2] 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): Article 22. https://pubs.lib.umn.edu/innovations/vol8/iss1/22

[3] Langley PC. Modeling Imaginary Worlds: Version 4 of the AMCP Format for Formulary Submissions. *Inov Pharm.* 2016;7(2): Article 11. https://pubs.lib.umn.edu/innovations/vol7/iss2/11

[4] 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. https://pubs.lib.umn.edu/innovations/vol8/iss2/14

[5] Institute for Clinical and Economic Review (ICER). Overview of the ICER value assessment framework and update for 2017-2019. https://icer-review.org/wp-content/uploads/2017/06/ICER-value-assessment-framework-update-FINAL-062217.pdf

[6] Langley PC. Cost-Effectiveness and Formulary Evaluation: Imaginary Worlds and Entresto Claims in Heart Failure. *Inov Pharm.* 2016;7(3): Article 6. https://pubs.lib.umn.edu/innovations/vol7/iss3/6

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

[8] 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): Article 10. https://pubs.lib.umn.edu/innovations/vol8/iss2/10

[9] Langley PC. Rush to Judgement: Imaginary Worlds and Cost-Outcomes Claims for PCSK9 Inhibitors. *Inov Pharm.* 2017;8(2): Article 11. https://pubs.lib.umn.edu/innovations/vol8/iss2/11

[10]Neumann PJ, Cohen JT. ICER's revised value assessment framework for 2017-2019: A critique. *PharmacoEconomics*. 2017;35:977-980

[11] NPC evaluates ICER's revised value assessment framework. http://www.npcnow.org/blog/npc-evaluates-icers-revised-value-assessment-framework?utm_source=NPC+Contact+List&utm_campaign=1d1b048abf-EMAIL_CAMPAIGN_2017_06_22&utm_medium=email&utm_term=0_3ddd3927eb-1d1b048abf-198281001

[12] http://www.npcnow.org/blog/why-va-should-proceed-caution-icer-collaboration

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

[14] Pigliucci M. Nonsence on Stilts: How to tell science from bunk. Chicago: Chgicago University Press, 2010

[15] Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the Economic Evaluation of Health Technologies: Canada (4th Ed). Ottawa: CADTH, 2017i

[16] Langley PC. True North: Building Imaginary Worlds with the Revised Canadian (CADTH) Guidelines for Health Technology Assessment. *Inov Pharm.* 2017;8(2): Article 9. https://pubs.lib.umn.edu/innovations/vol8/iss2/9

[17] Program in Social and Administrative Pharmacy, College of Pharmacy, University of Minnesota. Guidelines for Formulary Evaluation (Proposed) version 2. December 2016. https://www.pharmacy.umn.edu/sites/pharmacy.umn.edu/files/minnesota_guidelines_december_2016.pdf

[18] Langley PC. Nullius in verba: Version 2.0 of the University of Minnesota, School of Social and Administrative Pharmacy Program, Proposed Guidelines for Formulary Evaluation. *Inov Pharm.* 2016;7(4): Article 16. https://pubs.lib.umn.edu/innovations/vol7/iss4/16

[19] Sweet B, Tadlock C, Waugh W et al. The WellPoint outcomes based formulary: Enhancing the health technology assessment process. *J Med Econ*. 2004;8:13-25

[20] Langley PC. *Recent developments in the health technology assessment process* in Fulda T and Wertheimer A. *Handbook of Pharmaceutical Public Policy*. New York: Haworth Press, 2007.

[21] Caro JJ, Briggs A, Siebert U et al. Modeling good research practices – Overview: A report of the ISPOR-SMDM Modeling good practices task force – 1. *Value Health*. 2012 (15:796-803

[22] Eddy DM, Hollingworth W, Caro JJ et al. Model transparency and validation: A report of the ISPOR-SMDM modeling good research practices Task Force – 7. *Value Health*. 2012;15:843-850

[23] Caro JJ, Eddy DM, Kan H et al. Questionnaire to assess relevance and credibility of modeled studies for informing health care decision making: An ISPOR-AMCP-NPC good practice task force report. *Value Health*. 2014;(17):174-182

[24] Blackmore S. *The Meme Machine*. Oxford: Oxford University Press, 1999.

[25] Langley PC. Imaginary Worlds: Modeled claims for cost-effectiveness published in PharmacoEconomics January 2015 to December 2015. *Inov Pharm.* 2016;7(2): Article 9. https://pubs.lib.umn.edu/innovations/vol7/iss2/9

[26] 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. https://pubs.lib.umn.edu/innovations/vol8/iss2/14

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

[28] Wailoo A, Alava M, Grimm S et al. Comparing the EQ-5D-3L and 5L versions. What are the implications for cost-effectiveness estimates? Decision Support unit, ScHARR, University of Sheffield. March 2017. http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2017/05/DSU_3L-to-5L-FINAL.pdf

[29] Beresniak A, Medina-Lara A, Auray JP et al. Validation of the underlying assumptions of the quality-adjusted life years outcome: Results from the ECHOUTCOME European Project. *PharmacoEconomics*. 2015;33:61-69

[30] Newmann P, Weinstein M. Legislating against use of cost-effectiveness legislation. *New Eng J Med*. 2010. 365:1495-7

[31] Langley PC. Imaginary Worlds and Efficiency Frontiers: Should We Abandon the IQWiG Health Technology Assessment Model?. *Inov Pharm.* 2017;8(1): Article 13. https://pubs.lib.umn.edu/innovations/vol8/iss1/13

[32] https://www.diariofarma.com/2017/01/13/sanidad-rechaza-uso-avacs-la-fijacion-precios-medicamentos?&utm_source=nwlt&utm_medium=masleidas&utm_campaign=23/01/17

[33] Langley PC. He Ao Pohewa: The PHARMAC Prescription for Pharmacoeconomic Analysis in New Zealand and the Standards of Normal Science. *Inov Pharm.* 2016;7(2): Article 13. https://pubs.lib.umn.edu/innovations/vol7/iss2/13.

[34] Langley PC. Sunlit Uplands: The Genius of the NICE Reference Case. *Inov Pharm.* 2016;7(2): Article 12. https://pubs.lib.umn.edu/innovations/vol7/iss2/12

[35] Institute for Clinical and Economic Review (ICER). Vesicular monoamine transporter 2 inhibitors for tardive dyskinesia: Effectiveness and value. Evidence Report. November 2017.

[36] Langley PC, Rhee TG. Imaginary Worlds: The Status of Simulation Modeling in Claims for Cost-Effectiveness In Diabetes Mellitus. *Inov Pharm.* 2016;7(2): Article 17. https://pubs.lib.umn.edu/innovations/vol7/iss2/17