ICER, ISPOR AND QALYs: Tales of Imaginary Worlds

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

Abstract

Previous commentaries in the Formulary Evaluation section of INNOVATIONS in Pharmacy have pointed to the lack of credibility in modeled claims for cost-effectiveness and associated recommendations for pricing by the Institute for Clinical and Economic Review (ICER). The principal objection to ICER reports has been that their modeled claims fail the standards of normal science: they are best seen as pseudoscience. The purpose of this latest commentary is twofold: first, to review the latest report by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) on standards and value claims for modeling imaginary cost-per-quality life year (QALY) worlds in health technology assessments and (ii) given ICER's commitment to ISPOR standards, to consider the credibility of ICER's QALYs. The concern is that ICER gives the impression that there is a common and agreed utility metric to support QALY constructs and that there is an agreed standard for creating QALYs within the imaginary reference case meme. The purpose of this commentary is to emphasize that there is no 'gold standard' QALY, let alone a 'gold standard' utility metric. A QALY is what you assume the QALY should be; it is an artificial construct which allows any number of competing QALYs to be generated for the same target population under different imaginary constructs. Similar objections apply to the application of willingness-to-pay thresholds to support ICER's business case as the sole arbiter, in pole position, of health technology assessments to support pricing recommendations. In the US. A threshold has to be defined for the QALYs and costs captured, by assumption, within the model. Even for the same target patient population a $50,000 cost per QALY threshold will yield different pricing recommendations depending upon the assumptions driving the QALY estimate, such as choice of utility metric, time spent by stage of disease, frequency and severity of adverse events and costs. While this is an obvious point, it appears not to occur to those who, with a limited technical understanding of how the ICER model is constructed (and ICER is not given to extended technical explanations) take at face value the ICER imaginary construct and the QALY value judgements. Of course, it is also beside the point that the ICER reference case imaginary world was never intended to support empirical assessments of the claims made; we have no idea of whether it is right, if it is wrong and, over the lifetime perspective, we will never know and were never intended to know.

Keywords: ICER, ISPOR, utility metric options, QALY options, imaginary worlds, pseudoscience, imaginary assumptions

Introduction

The business model of the Institute for Clinical and Economic Review (ICER) rests on the construction of imaginary incremental cost-per-QALY worlds. Previous commentaries in *INNOVATIONS in Pharmacy* have pointed to the fact that these constructions, which are widely accepted by mainstream health technology assessment groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), fail to meet the standards or normal science; they are best described as pseudoscience [1 2 3 4 5 6 7 8]. They fail the demarcation test between science and non-science (cf. natural selection vs. intelligent design). While their supporters acknowledge that they fall outside of the normal activities of hypothesis testing in problem solving, the traditional way that science is seen to add to our knowledge base through an ongoing process of conjecture and refutation, the imaginary worlds of ICER may be conjectural but are impossible to refute. It is an analytical dead end. Unfortunately, while ICER's adoption of the reference case construction kit for imaginary worlds fails to meet the standards of normal science, there are still those who are prepared to take this endeavor seriously.

The purpose of this commentary, following from a previous commentary that reviewed the status of imaginary QALY claims, is to consider ICER's embrace of the cost-per-QALY paradigm [9]. In this arcane world of constructing imaginary worlds, ICER identifies and applies utility scores to estimate QALY gains over a timeframe that may extend for decades into the future. Claims are made for competing therapies from constructed incremental cost-per-QALY comparisons set against cost-per-QALY threshold values. What is often overlooked is that creating QALYs, the choice of utility metric and time spent in a modeled disease stage is entirely discretionary, as are the costs that the model builder may decide to include. This means that any arbitrary cost-per-QALY threshold (e.g., $50,000 per QALY) will yield different recommendations for price discounting depending on the model assumptions. To achieve comparable recommendations for discounting a product price, would require a threshold calibrated for the specific model construct, utility metric and costs.

The critique presented in this commentary applies to reference case lifetime models irrespective of the type of modeling framework that is employed. Economic models can employ either a patient level simulation or a cohort level framework to

Corresponding author: Paul C Langley, PhD
Adjunct Professor College of Pharmacy
University of Minnesota
Minneapolis MN
Director, Maimon Research LLC; Tucson, AZ
Email: langley@maimonresearch.com

DOI: https://doi.org/10.24926/iip.v10i4.2266

generate lifetime estimates of average direct medical costs and outcomes, typically expressed as comparative incremental cost-per-QALY projections [10]. If patient level simulations or hybrid versions are seen as the preferred lifetime framework this does not mean that they can be absolved from application of the demarcation test.

Unfortunately, as detailed here, falling back on the reference case as sufficient to drive comparable value judgements across disease areas and hypothetical target patient populations is a non-starter. Certainly, the authors of reference case guidelines may have sought to create a barrier against untoward competition in building lifetime imaginary worlds. Unfortunately, the reference case as a defense overlooks some significant methodological deficiencies, apart from its characterization as pseudoscience. It is certainly no defense against the proliferation of imaginary cost-effectiveness models that have dominated the literature over the past 30 years.

ICER in its embrace and advocacy of the construction of imaginary cost-per-QALY health technology assessments is following the standards for constructing imaginary worlds put in place by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). These standards have been recently revisited and endorsed by ISPOR in as Special Task Force report. This commentary briefly reviews this report and its recommendations for enhanced value measures to complement the core cost-per-QALY imaginary framework.

Central to this QALY critique of ICER imaginary model building is the audience for ICER value judgements, in particular the blanket use of the term QALY. Perhaps inadvertently, the impression given is of a 'unique' metric that drives ICER's modeled claims. If this is a perception, then it needs to be challenged. There is no unique, agreed QALY metric standard. Rather, there are a range of utility constructs, application of which would potentially yield competing QALY estimates.

This is not to imply, however, that utility measures do not have a place in health technology assessment. There are abundant studies where patient reported outcomes (PRO) instruments have been utilized to evaluate competing claims. The key difference, however, is that these are evaluated in the context of randomized clinical trials (RCTs) and observational studies with claims to be evaluated that are credible, evaluable and replicable.

ISPOR: State of the Art

Claims by ICER that the construction of reference case, imaginary lifetime cost-per-QALY worlds is the 'state of the art' in health technology assessment is quite accurate. The construction by assumption of imaginary worlds is not only endorsed by ISPOR and the Academy of Managed Care Pharmacy (AMCP) in the US but also by the overwhelming majority of assessment agencies for single payer health systems.

Most recently, ISPOR has reaffirmed its commitment to the construction of imaginary worlds in a Special Task Force Report issued in 2018 [11]. This task force was convened to *review relevant perspectives and appropriate approaches and methods to support the definition and use of high-quality value frameworks;* presumably 'high value' imaginary constructs, although no criteria are suggested that may distinguish 'high value' from 'low value' imaginary worlds or how a metric might be applied to inform those willing to utilize imaginary worlds in formulary decisions.

Expressing a concern that competing value frameworks in attempting to simplify the problem of value could end up making *ad hoc assumptions and simplifications not supported by theory or evidence,* the Report takes the position that frameworks that do not capture the full costs and benefits of treatment may distort decisions, although 'full' is not defined outside of broad reference case guidelines. The approach taken by ISPOR is *'to define value on the basis of microeconomic principles, recognizing that value is best defined as what individuals (or others acting on their behalf) would be willing to pay to acquire more health care or other goods and services.* This is best achieved, maintains the Report, by informing resource allocation decisions by 'approximating' the value of interventions in incremental cost per quality adjusted life year (QALY) term, capturing length and quality of life in an imaginary reference case construct.

The Report points out that thousands of these cost-per-QALY analyses have been carried out and catalogued. One such catalogue is the Tufts University, Center for the Evaluation of Value and Risk in Health's Cost Effectiveness analysis (CEA) Registry where the discerning builder of modeled imaginary worlds can scroll through hundreds of utility scores catalogued over the past 30 years and linked to diseases and disease stages to populate their model and generate QALYs [12]. It is not clear, however, from the summary provided by the Center, the basis for the listed utility score. There appear to be no codes categorizing the utility metric (e.g., EQ-5D-5L vs. SF-6D) and its construction (e.g., choice of mapping function). Of course, reader can always refer back to the source study. The risk is that prospective users, in particular those constructing imaginary worlds, will take the scores at face value and populate their models accordingly.

The ISPOR Report cautions, however, that while they believe cost-per-QALY claims are the key driver there are limitations to the metric. It may not fully capture: (i) the health or wellbeing of patients (and, presumably, caregivers); (ii) individual or community preferences for health state attributes; (iii) disease severity; (iv) equity of access; (v) unmet medical need and (vi) the perspective of the audience to whom the value claim is addressed: the patient, payer or society at large. If the patient, including the caregiver in pediatric or younger adolescent populations, is the preferred perspective then there will exist

DOI: https://doi.org/10.24926/iip.v10i4.2266

tensions as they do not necessarily meet the full costs of health interventions and may put to one side concerns regarding the opportunity costs of their choice.

It is also recognized that within any disease area and for any target pharmaceutical product or device there is a range of options, not only for deciding which imaginary value framework is appropriate but also the perspective taken within that framework. Is the perspective that of the health plan, the patient, the health plan manager, the provider, the manufacturer, a special interest group, government regulators or society as a whole? The Report cautions, therefor, against attempts to create a single value-based metric that may put to one side variations in individual preferences, together with ability and willingness to pay.

The Special Task Force Report also focuses on the range of value elements that might be captured in a technology assessment [13]. In addition to the conventional elements that are included or considered for inclusion in modeled claims (QALYs, net costs, productivity and adherence improving factors) , the report also considers: reduction in uncertainty, fear of contagion, insurance value, severity of disease, value of hope, real option value, equity and scientific spillovers. While, inevitably, saying further research is warranted, the Report considers how these elements might be included in a cost-utility framework. Options canvassed are increasing the scope of the QALY measure, broadening health state descriptions or even utilizing multi-criteria decision analysis (MCDA) [see below].

The Special Task Force Report also considers how multiple elements of value might be combined into a single value metric for individuals [14]. These include extended measures which attempt to describe the value of medical interventions along three dimensions: health gains, financial risk protection and social for population subgroups. While it is unclear how these might be aggregated to a single metric, the health gains are still assessed by a utility metric to generate QALYS or a similar measure. Going beyond the three additional dimensions captured in the extended concept of value, the Report introduces the term augmented value to capture further dimensions of value. The possible list is extensive, including scientific spillovers, reducing diagnostic uncertainty and financial insurance value. This extended and augmented value concepts are still speculative and the Report acknowledges that they cannot be fully aggregated (e.g., be monetized) across the various dimensions of value they describe. Finally, the Report suggests that if a single metric is the objective, we should consider MCDA, but without any firm direction of what the MCDA elements might be or the MCDA model most appropriate to creating a 'weighted' metric as well as addressing the critical issue of aggregating across individuals [15] [16]

The last section of the Report considers more recently proposed value frameworks [17]. The value frameworks reviewed include those of the American College of Cardiology/ American Heart Association, the Institute for Clinical and Economic Research, the American Society of Clinical Oncology, the National Comprehensive Cancer Network, and the Memorial Sloan Kettering Cancer Center. The assessment considers the diversity of frameworks presented in terms of their face and content validity, reliability and conceptual underpinning. In other words, do these value tools conform to mandated ISPOR standards. Not surprisingly, the ICER imaginary reference case incremental cost-per-QALY or cost-effectiveness analysis is seen as the standard for base health plan coverage as recommended in the Special Task Force Report. Value thresholds are a key input to coverage and reimbursement decisions with payers encouraged to introduce willingness-to-pay or cost-per-QAL metrics or similar threshold decision tools given their opportunity costs and budget constraints. These may vary by patient as well as being modified by equity and severity considerations, although budget impact should not be an integral part of value assessment. The Report recommends structured deliberative processes for health together with a plea for exploring and testing novel elements of benefit.

Mapping Health State Utilities

The degree of acceptance of the imaginary world meme in health technology assessment is seen in the time and effort over the past 20 or more years to mapping health state utilities from non-preference based outcome measures together with attempts to crosswalk between generic preference based utility measures, notably the ongoing reviews of the impact of moving from the EQ-5D-3L to the EQ-5D-5L on modeled cost-effectiveness claims. In the former case ISPOR has issued good practice guidelines for estimating the relationship between outcomes measured in clinical studies (e.g., RCTs) and health utility [18]. This mapping allows clinical markers as inputs to equations to estimate utility score approximations which can then populate modeled incremental cost-per-QALY claims. Whether this effort has been rewarded if the objective is to populate a model and generate non-evaluable lifetime incremental values is, again, a moot point. In the latter case the introduction of the EQ-5D-5L in 2009 gave clinical researchers and health technology assessors the opportunity to use a potentially more sensitive instrument as the EQ-5D-3L was considered to have unacceptable floor and ceiling effects and to lack sensitivity as it only captured three responses to the five health dimensions: no problem, some problem and extreme problem. Unfortunately, the EQ-5D-5L with its five response levels created its own set of problems as attempts to rework previous modeled claims for cost-effectiveness using the 3L version pointed to significant differences with the re-modeled claims [19].

Go Forth and Multiply

While the Special Task Force Report asks (inevitably) for more research, the base case is a commitment to the application of a reference case to support imaginary incremental cost-per-QALY modeling supported by threshold decision criteria [20]. Attempts to widen the scope of the metric employed to capture other dimensions, such as equity and disease severity as well as consideration of other perspectives and value frameworks that

DOI: https://doi.org/10.24926/iip.v10i4.2266

may utilize MCDA, are still secondary to the imaginary incremental cost-per-QALY meme.

Whether the research agenda proposed by the Special Task Force Report will ever have a practical import in health decision making outside of the activities of ISPOR and its academic following is a further moot point. At best, the existing value frameworks identified by the Report may continue to mature to include additional dimensions of interest, with special interest groups (e.g., in rare diseases) proposing value frameworks specific to the target patient populations. Even so, the question remains: will these be assessed with a framework for credible and evaluable claims or will health technology assessment continue to be dominated by the imaginary world meme.

The Special Task Force Report does not address the question of the scientific status of the proposed value frameworks, notably those that promote the lifetime incremental cost-per-QALY as central to the decision problem. This is not surprising. After all, for the last 30 years cost-effectiveness, specifically in cost-benefit or cost-utility analysis, the construction of imaginary worlds has been the mainstay of health technology assessment. To admit to the ISPOR global membership that claims from a lifetime cost-per-QALY model are not only impossible to evaluate but that they were never intended to support hypothesis testing would be embarrassing. While they are touted as 'providing approximate information' (whatever that means), the fact is that we will never, as already emphasized, intended to know if they were right or if they were wrong. We could create any number of competing cost-per-QALY models for the products in a disease area for a target population but would have no basis other than a comparison of assumptions to judge whether the information created by one model was preferable to another.

There is no simple answer to the question of which preference based multi-attribute health status system to use, or whether to opt out. As Drummond et al point out: (i) the decision does matter as the systems are far from identical, they differ in the health dimensions and levels assigned to each dimension, in the description of those levels and in the severity of the most severe level; (ii) they differ in the population surveyed in the construction of the system and the instruments used to determine the preference based scoring; and (iii) they differ in the theoretical approach taken to modeling the preference data into a scoring formula [21].

The diversity of generic preference-based multi-attribute systems is not, of course, to endorse them in constructing imaginary cost-per-QALY worlds. While this meme has had widespread support as shown by, for example, the two US Panels on cost effectiveness of 1996 and 2016, together with the various do-it-yourself imaginary world construction textbooks, on the standards of normal science this approach to health technology assessment is best seen as pseudoscience (c.f., intelligent design) [22] [23] [24]. Science is a process of discovery, not the creation of imaginary worlds. As Newton (Isaac Newton 1642-1727), with Descartes as his target (René Descartes 1596-1650) stated '*hypotheses non fingo*' (I do not feign hypotheses) Descartes, in Newton's view had 'produced fantastic and untestable ideas, then assumed them to be true and used them as building blocks of his philosophy [25].

Imaginary ICER Worlds

ICER's embrace of the imaginary worlds meme is consolidated in its so-called reference case. This sets out the standards for constructing imaginary worlds. Unlike NICE in the UK or other technology assessment agencies such as the PBAC in Australia, these reference case standards are not directed to manufacturers who have been asked to make modeled submission, but are directed to ICER itself (or, more accurately, the contracted groups at the universities of Washington, Illinois and Colorado who construct the imaginary models). ICER is always the final arbiter on the model although manufacturers and other stakeholders may offer suggestions if they decide to engage with ICER in what many see as a non-productive activity. Few suggestions are taken up.

Guidelines set out in the ICER reference case make it clear that wherever possible (with a few noted exceptions) ICER should strive in building its in-house model to generate its QALY estimates from either generic health preferences captured directly from the a representative sample of the US or mapped from clinical markers. As ICER states:

- Health preferences should reflect those of the general US population (preferably), providing a rationale if patients with the condition, individuals at heightened disease risk, or a different population is used
- Health preferences should be from an indirect method of measurement based on a generic measurement system (e.g. EQ-5D-5L)
- When there are challenges to outcome measures used in clinical trials and available patient-reported data translating into QALYs, mapping studies that allow translation of surrogate outcomes into quality of life measures should be considered
- If using ultra rare disease framework, acknowledge and highlight additional uncertainty in translating additional outcomes into QALY measures [26]

Although it appears implicit in the ICER reports that the EQ-5D is the preferred generic preference metric, it is not clear whether it is the EQ-5D-3L or the EQ-5D-5L that is preferred. As it stands these two versions of the EQ-5D should be seen as separate instruments. Certainly it is possible to crosswalk and the EUROQoL group has produced value set tables for the EQ-5D-5L cross walked with the EQ-5D-3L [27]. In the case of NICE the present position (October 2019) is that the preferred measure of health-related quality of life in adults is the EQ-5D-3L value set for reference case purposes [28]. If data are collected using the EQ-5D-5L system, utility values in reference case analyses should be calculated by mapping the descriptive system data to the 3L value set. The selected mapping function

DOI: https://doi.org/10.24926/iip.v10i4.2266

is that developed by Van Hout [29]. In prospective clinical studies the EQ-5D-5L descriptive system is supported.

If a cost-effective model framework has been developed which provides for the assessment of credible, evaluable and replicable claims for target populations across disease areas then the analyst is open to utilizing a preference-based multi-attribute system for health related quality of life (HRQoL) outcmes. If issues are raised as to the relevance of the established generic instruments, then options are open to either augment the health dimensions or substitute a disease specific instrument. This may not be a generic instrument but a disease specific one which may be independent of generic measures or an addendum to those measures. Indeed, there is a substantive literature questioning the relevance of generic instruments to both HRQoL and QoL considerations in disease states. This does not apply just to rare diseases that impose a substantial burden on caregivers and the wider family (e.g., Duchenne muscular dystrophy, sickle cell disease), but to more prevent disease states such as schizophrenia and, more generally, the range of mental health conditions [30]. Perhaps ICER should be advised that a one-size reference case construction kit for imaginary worlds (with presumably the EQ-5D-3L as the 'preferred' metric) does not fit all and that persevering with arguments that, by ad hoc additions, where other considerations might be set alongside the core imaginary value judgements, is an analytical dead-end.

But this is not all. Even if a specific preference metric were mandated, any modeled claim could be challenged on the grounds that: (i) the modeled assumptions and particular mathematical form of the model (e.g., Markov process) will generate time spent on average for a hypothetical target population in a disease stage that is only one of many other potentially modeled stages and (ii) in assessing the numerator of the incremental claims, the assumptions made regarding which medical inputs and their projected unknown lifetime costs are modeled are again open to a range of constructs. Rather than seeking to add to our knowledge of cost-effectiveness claims and the most effective way of presenting these claims to formulary committees (e.g., as modeled extensions of RCTs) the technology assessment meme directs its followers to the construction of imaginary world where each imaginary world for a target population is one of a potential multiverse of models and non-evaluable value judgements for competing products in that therapy area.

If the focus of criticisms of ICER is the QALY, then the criticism should be more focused on the utility construct, the estimated average time spent in a disease stage and the assumptions underpinning the selection and valuing of direct medical costs. ICER needs to inform patient advocacy groups who may raise issues regarding the relevance of the utility construct to HRQoL or QoL that either (i) ICER is wedded to a specific generic measures (e.g., EQ-5D-3L) and will not countenance any deviation or (ii) that it recognizes that as threshold values are unique to individual models that it will accept recommendations for other utility metrics in creating an imaginary world.

The Universal QALY Threshold

Adoption of willingness-to-pay thresholds implies that a common generic metric should be mandated in model construction. In the US, with the requirement for US preferences, the choice is between, in practical terms, the SF-6D and the EQ-5D-3L/5L [31] [32] [33] [34]. If one instrument is mandated, then utility scores embedded in the model should only be those that are generated by this instrument or mapped to that instrument. Combining scores from different instruments, let alone scores which fail to meet US preference standards, should be put to one side. As ICER controls the construction of imaginary worlds this should not be an obstacle. There is, of course, no reason ICER has to follow its own standards. In fact, ICER has cobbled together scores from different instrument on grounds that they will give similar scores - an example is the ICER evidence report for oral semaglutide in Type 2 diabetes [35] .

In the case, as noted above of the EQ-5D-3L and the EQ-5D-5L it should be made clear that the two instruments will yield different utility scores and hence QALY claims, *ceteris paribus*, for the same stage of disease in the same target population. If the object is to generate identical recommendations for price discounting irrespective of the generic measure then the ICER willingness-to-pay threshold(s) must accommodate this with different generic utility measures mapped to specific thresholds. A $50,000 threshold for one metric will not yield the same price discounting claims as the same threshold for a competing metric in the same disease state. Indeed, it would be possible to take the ICER model final version and substitute utility scores from other generic instruments, to include the possible specific form of mapping algorithm. This would result in decision makers having the option of choosing between modeled imaginary worlds and associated value judgments.

ICER's much published value judgements for threshold based price discounting are a mirage. Unless the threshold is specific to a QALY measures (which involves not just the selected metric system but the consistent use of direct medical cost estimates and the underlying model that constructs time spent in different health states over the hypothetical average lifetime of the hypothetical target patient population) then the application of a fixed threshold across products in the various disease areas has no validity. A threshold only has relevance in its relation to the model driving the specific QALY estimates and the estimated direct medical costs. If different models are constructed within a disease area, a common recommendation for price discounting will be driven by thresholds specific to the various imaginary world constructs.

Of course, given the imaginary nature of the modeled QALY claim, ICER could put aside its commitment to generic US

DOI: https://doi.org/10.24926/iip.v10i4.2266

preference based utilities by stage of disease drawn from the literature, to simply assuming arbitrarily that a utility score is 'approximate' and reflects the modeler's judgment that it is a 'reasonable proxy'. As it stands, ICER's position that it is consistent in applying across the board the same cost-per-QALY thresholds ($50,000, $100,000 and $150,000) to support recommendations for price discounting is misleading as it is virtually impossible to apply the reference case to support identical model structures and assumptions.

In practice, unfortunately, the ICER contracted model builders have not adopted a standard utility measure. Rather, as suggested by the various evidence reports, the model builders seek out utility metrics where they can find them from a disease specific literature review. Whether the metric is the EQ-5D-3L, the EQ-5D-5L or the HUI Mk3, ICER modelers rely on being able to extract scores from the literature. The selected scores depend upon what the model builders can find rather than an *a priori* justification for the relevance (or otherwise) of the metric in that disease state and the appropriateness of a preference based instrument in that disease state. A further potential complication that abstracting utility measures mapped or reported from clinical trials or observational studies reported in the literature for a disease state, may not match the characteristics of the ICER assumed target population or modeled disease stage. The added complication is that even if ICER explicitly mandated the EQ-5D-3L in the reference case, it has not mandated the mapping algorithm that is appropriate (or criteria to be applied) when it is modeling different disease states. As the ISPOR Task Force Report on mapping makes clear, if a mapping algorithm is required (or selected from the literature) it has to be justified with an audit trail detailing its choice.

If a generic instrument is judged by patient groups to only encompass a minimum set of disease experience attributes, then it may reasonably be argued that there is little difference between an arbitrary choice of generic instrument and the adoption of a disease specific metric from an on-line catalogue such as the Tufts utility emporium. This opens, to apply a cliché, Pandora's Metric QALY box to address the more pertinent question of why, in constructing cost-effectiveness claims we focus on QALYs when the outcomes, in a lifetime framework, not only lack credibility but are, by definition, impossible to evaluate empirically and replicate across target patient populations. The ICER imaginary world is but one more addition to a shelf of existing imaginary worlds, none of which can be judged to be necessarily more 'informative' or 'realistic' than any other. It would be just as appropriate to select one at random.

It is difficult to see how ICER might escape from this predicament. Mandating a generic instrument raises two problems: (i) there may not be evidence on generic utility metrics from the literature to support 'assumptions' to populate the imaginary world of a specific target population in a disease area (unless any assumption for scores 'will do') and (ii) there may not be an algorithm, specific to a disease area, that maps the utility score from the specific study to the mandated ICER 'gold standard' utility score. In the case, for example, of the EQ-5D-5L metric with US preferences, there are unlikely to be more than a handful (if that) of published studies that would allow utility scores to be plucked from the literature for inclusion in the imaginary world model. Falling back on the EQ-5D-3L then raises the issue, as detailed above, that the two are effectively separate systems. Although ICER takes approximately 8 months to produce a final evidence report for public review and then voting by a selected group of experts, it is unlikely that ICER would allocate time and resources to meet evidence gaps. Rather, ICER attempts to populate its reference case; falling back reluctantly is it appears on review that the evidence base is too weak to support the lifetime incremental imaginary world (as witnessed in the recent withdrawal and reissue of the JAK inhibitors in rheumatoid arthritis evidence report) [36] .

Conclusions

If ICER value judgements are built on nothing more than a series of assumptions, where each model relies on a different set of assumptions, a point that might be of interest to those wishing to examine the entrails of ICER's imaginary worlds, is to consider the choice of assumption. After all, if the object is to convince, possibly by now a somewhat skeptical audience, of the importance of providing policy makers with constructed information from imaginary worlds, what are the inclusion and exclusion criteria for the model structure and the 'verified' assumptions: Gormenghast or Narnia? Both require a suspension of belief yet both embody assumptions and characterizations that have 'real world' analogs (although Aslan is a stretch).

Unfortunately, for those such as ISPOR and ICER who argue for the 'state of the art' imaginary world meme, the reference case is not a robust enough framework to fall back on in justifying an incremental cost-per-QALY, threshold driven approach to cost-effectiveness. It is not just the argument for a 'master' metric. In the case of NICE, the EQ-5D-3L is the mandated metric; or at least until there is an agreement on a value scoring for the EQ-5D-5L. Apart from the issue of the choice of mapping algorithm where, as occurs in most cases, the metric is created from clinical markers, there is the wider issue of bringing together other elements to create QALYs for a cost-per-QALY estimate. This brings into question the structure of the model and the assumptions populating the model. Not least is the bizarre requirement that the model should track, in chronic disease states, the natural course of the disease (the lifetime perspective). Believers presumably accept at face value that it is possible to make 'reasonable' or 'realistic' assumptions about the future, although the ICER reference case puts to one side modeling pharmaceutical price increases. Is the model meant to be 'realistic' or not? Is it just an artificial construct that meets certain speculative standards and which is just one imaginary

DOI: https://doi.org/10.24926/iip.v10i4.2266

future world among other equally valid, by assumption, imaginary future worlds? This is not an issue for NICE as the submission and review process ensures that there is a 'NICE approved' imaginary world. All parties accept this and the resulting threshold implications for pricing. The playing field is level and all parties know the rules of the 'game'. There are even imaginary world referees, typically in academic institutions, who will adjudicate the manufacturer's imaginary submission. They can pronounce whether it is acceptable, modifiable or should be replaced by the referees own proposal for an imaginary world. NICE, as senior referee, is the final judge.

ICER is in a quite different position: it has no legislative or regulatory mandate for its assumed role in health technology assessment. It faces a further hurdle that under the Affordable Care and Patient Protection Act (2010) it is made clear that the Patient Centered Outcomes Research Group (PCORI) exclude discounted cost-per-QALY or similar measures for threshold values for priority setting by the Centers for Medicare & Medicaid Services [37]. Even putting this hurdle to one side, the objections to the construction of imaginary worlds should alert health decision makers to the inadvisability of the ICER reference case to drive formulary decisions. While this does not mean an objection in principle to the application of cost-per-QALY models that generate credible and evaluable claims, supporting scientific discovery, but the rejection of the anachronistic and counter-productive role that ICER has taken upon itself in health decision making.

Unfortunately, few recipients of the ICER evidence reports seem to be interested in or willing to question the lack of scientific status in the ICER reference case. Decision makers 'take ICER's word for it'; a belief that, in some sense, the ICER imaginary construct, as one of a possible multiverse of competing imaginary worlds and value judgements, is 'useful' information. ICER clinical assessments, pricing and affordability are reported through the media and taken up by decisions makers who are typically not in a position to undertake a post-mortem of the ICER model. As noted in previous commentaries, the scientific revolution of the 17th century was built on the construction of empirically verifiable theories and hypotheses[38]. As evidence for this consider the motto of the Royal Society (founded 1660; Royal Charter 1662) *nullius in verba* (take no man's word for it)[39].

If we are concerned with the growth of knowledge in health technology assessment, then a discussion of the use of utility metrics and QALYs in the imaginary worlds of ICER seems pointless. After all, the construction of imaginary worlds, even if defended on the grounds that they represent the 'state of the art' in the consensus on health technology assessment, represent an analytical dead end and should be abandoned. Against this there are certainly many analysts that would accept the information role of imaginary worlds or at least have a vested interest, after 30 years and thousands of published imaginary models, in affirming that the emperor is well dressed despite claims to the contrary. ICER has an undeniable impact with many prepared to defend ICER's recommendations, including retail groups such as CVS [40]. That being said, one way of illustrating the inevitable shortcomings of imaginary constructs is to point out that any finite number of observations can be proved to be accommodated within an indefinitely large number of different explanations (Gottfried Liebniz 1646-1716). The ICER imaginary world is only one of many. These alternative explanations can claim to be 'realistic' yet rely on plucking a competing set of utility scores from the literature. Formulary committees could then select, from the range of possible imaginary worlds, which imaginary world's information they feel most comfortable with and its attendant non-evaluable incremental cost-per-QALY claims. Or, they could reject the ICER approach and ask for credible and evaluable claims.

ICER, of course, will continue to construct imaginary worlds, produce evidence reports and after their 8 month process of review, issue imaginary value judgements for pricing and access based on willingness to pay thresholds. There is a responsibility, therefore, to point out, even if we object to the relevance of constructing imaginary worlds, that there is no unique QALY measure. Competing measures will lead to different imaginary claims. Even if ICER mandated, for example, the EQ-5D-3L as the relevant utility measure to generate imaginary QALYs, any conclusion could be challenged on the grounds that the as a generic measure it fails to capture relevant aspects of either QoL or HRQoL within that disease state. ICER is in a no win position. To which might be added the fact that, unless directly administered to a target patient population, the claimed EQ-5D-3L is a product of the analysts choice of mapping algorithm. Different algorithms will generate different utility scores with the inevitable debate over which mapping algorithm is econometrically robust and yield the 'best fit'.

Even if ICER, under pressure, abandoned a mandated EQ-5D-3L and recognized a role for disease specific instruments, the situation would not change. There would be a continuing debate over the relevance of disease specific instruments with the added complication that there are many disease states, notably in rare diseases, where there is no utility metric.

Similar arguments could be made against any of the assumptions made to support ICER's imaginary claims. The model structure can be challenged, assumptions regarding costs can be challenged and attempts by ICER to hide behind alternative scenarios and the application of probabilistic sensitivity analysis, can all be challenged. This brings us back to the fundamental point: creating imaginary worlds by assumption to influence pricing and access decisions is to acknowledge the appropriateness of Bentham's (Jeremy Bentham 1748-1832) memorable phrase for the technology assessment meme: 'nonsense on stilts'. Or, to consider earlier memorable remarks, Kepler's (Johannes Kepler 1571-1630)

DOI: https://doi.org/10.24926/iip.v10i4.2266

comment on Fludd (Robert Fludd 1574 – 1637) the English alchemist and Rosicrucian with whom he had a long standing and acrimonious relationship: *It is obvious he derives his main pleasure from unintelligible charades about the real world, whereas my purpose is, on the contrary, to draw the obscure facts of nature into the bright light of knowledge* [25].

Conflicts of Interest PCL is an Advisory Board member and Consultant to the Institute for Patient Access and Affordability, a program of Patients Rising.

References

[1] 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/index.php/innovations/article/view/449

[2] 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/index.php/innovations/article/view/492

[3] 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/index.php/innovations/article/view/515

[4] 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/index.php/innovations/article/view/516

[5] Langley PC. Another Imaginary World: The ICER Claims for the Long-Term Cost-Effectiveness and Pricing of Vesicular Monoamine Transporter 2 (VMAT2) Inhibitors in Tardive Dyskinesia. *Inov Pharm*. 2017;8(4): Article 12 https://pubs.lib.umn.edu/index.php/innovations/article/view/927

[6] Langley PC. Resolving Lingering Problems or Continued Support for Pseudoscience? The ICER Value Assessment Update. *Inov Pharm*. 2017;8(4): Article 7 https://pubs.lib.umn.edu/index.php/innovations/article/view/933

[7] Langley PC. Transparency, Imaginary Worlds and ICER Value Assessments. *Inov Pharm*. 2017;8(4): Article 11 https://pubs.lib.umn.edu/index.php/innovations/article/view/926

[8] Langley PC. Alternative Facts and the ICER Proposed Policy on Access to Imaginary Pharmacoeconomic Worlds. *Inov Pharm.* 2018;9(2): Article 10 https://pubs.lib.umn.edu/index.php/innovations/article/view/1300

[9] Langley PC. Great Expectations: Cost-utility models as decision criteria. *Inov Pharm*. 2016;7(2): Article 14. https://pubs.lib.umn.edu/index.php/innovations/article/view/437

[10] Davis S, Stevenson M, Tappenden P et al. NICE DSU Technical Support Document 15: Cost-effectiveness Modeling Using Patient Level Simulation. Decision Support Unit, ScHARR. University of Sheffield. April 2014

[11] Neumann P, Willke R, Garrison L. A health economics approach to US value assessment frameworks – Introduction: An ISPOR Special Task Force Report [1] *Value Health*. 2018;21:119-123

[12] Tufts University. Center for the Evaluation of Value and Risk in Health. Cost Effectiveness Analysis (CEA) Registry. https://cevr.tuftsmedicalcenter.org/databases/cea-registry

[13] Garrison L, Pauly M, Willke R et al. An overview of value, perspective and decision context – A health economics approach; An ISPOR Special Task Force Report [2]. *Value Health*. 2018;21:124-130

[14] Lakdawalla D, Doshi J, Garrison L et al. Defining elements of value in health care – A health economics approach: An ISPOR Special Task Force Report [3]. *Value Health.* 2018;21:131-139

[15] Danzon P, Drummond M, Towse A et al. Objectives, budgets, thresholds, and opportunity costs – A health economics approach: An ISPOR Special Task Force Report [4]. *Value Health*. 2018;21:140-145

DOI: https://doi.org/10.24926/iip.v10i4.2266

[16] Phelps C, Lakdawalla D, Basu A et al. Approaches to aggregation and decision making – A health economics approach: An ISPOR Special Task Force Report [5]. *Value Health*. 2018;146-54

[17] Review of Recent US Value Frameworks—A Health Economics Approach: An ISPOR Special Task Force Report [6]. *Value Health*. 2018;21:155-60

[18] Wailoo A, Hernandex-Alava M, Manca A et al. Mapping to estimate health-state utility from non-preference-based outcome measures: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health*. 2017;20:18-27

[19] Hernández-Alava M, Wailoo A, Grimm S et al. EQ-5D-5L versus EQ-5D-3L: The impact on cost-effectiveness in the United Kingdom. *Value Health*. 2018;21(1):49-56

[20] Garrison L, Neumann P, Willke R. A health economics approach to US value assessment frameworks – Summary and recommendations of the ISPR Special Task Force Report [7]. *Value Health*. 2018;21:161-165

[21] Drummond M, Sculpher M, Torrance G et al. Methods for the Economic Evaluation of Health Care Programmes. 3rd Ed. Oxford University Press 2005.

[22] Weinstein M, Siegel J, Gold M et al. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996;276(15):1253-8

[23] Sanders G, Neumann P, Besu A et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses. Second Panel on Cost Effectiveness in Health and Medicine. *JAMA*. 2016;316(10):1093-1103

[24] Neumann P, Sanders G, Russell LK et al. Cost-effectiveness in Health and Medicine. New York, NY: Oxford University Press, 2016

[25] Briggs R. The Scientific Revolution of the seventeenth century. Longman, 1971.

[26] ICER. Modifications to the ICER value assessment framework for treatments for ultra-rare diseases Final Version November 2017 https://icer-review.org/wp-content/uploads/2017/11/ICER-Adaptations-of-Value-Framework-for-Rare-Diseases.pdf

[27] EuroQoL. https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/crosswalk-index-value-calculator/

[28] NICE. Position statement on the use of the EQ-5D-5L value set for England (updated October 2019). https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l

[29] Van Hout B, Janssen M, Feng Y et al. Interim scoring for the EQ-5D-5L value set for England. *Value in Health*;15:708-15

[30] Hastrup H, Nordentoft M, Hjorthø C et al. Does the EQ-5D measure quality of life in schizophrenia in schizophrenia. *J Ment Health Policy Econ*. 2011;14(4):187-96

[31] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21(2):271-92

[32] Craig BM, Pickard S, Stolk E et al. US valuation of the SF-6D. *Med Decis Ma*king. 2013;33(6):793-803

[33] Craig BM, Rand K. Choice defines QALYs: A US valuation of the EQ-5D-5L. *Med Care*. 2018;56(6):529-536

[34] Mulhern B, Feng Y, Shah K et al. Comparing the UK EQ-5D-3L and English EQ-5D-5L value sets. *Pharmacoeconomics,* 2018;36960;699-713

[35] ICER. Oral Semaglutide for Type 2 Diabetes: Effectiveness and Value. Draft Evidence Report (updated) 12 September 2019. https://icer-review.org/wp-content/uploads/2019/04/ICER_Diabetes_Draft-Evidence-Report_091219-2.pdf

DOI: https://doi.org/10.24926/iip.v10i4.2266

[36] ICER. Janus Kinase Inhibitors for Rheumatoid Arthritis: Effectiveness and Value. Draft Evidence Report. October 11, 2019 https://icer-review.org/wp-content/uploads/2019/03/ICER_RA_Draft_Evidence_Report_101119.pdf

[37] Persad G. Priority setting, cost -effectiveness and The Affordable Care Act. *Am J Law Med* . 2015;41:119-166

[38] Piglucci M. Nonsense on stilts. How to tell science from bunk. Chicago: Chicago University Press, 2010.

[39] Wootton D. The Invention of Science: A new history of the scientific revolution. New York: Harper Collins, 2015

[40] Langley PC. CVS Health and the Imaginary Worlds of the Institute for Clinical and Economic Review (ICER). *Inov Pharm*. 2018;9(4): Article 4. https://pubs.lib.umn.edu/index.php/innovations/article/view/1461

DOI: https://doi.org/10.24926/iip.v10i4.2266