

MAIMON WORKING PAPERS No. 7 MAY 2021

SUPPING WITH THE DEVIL: BELIEF AND THE IMAGINARY WORLD OF MULTIPLE MYELOMA THERAPIES INVENTED BY THE INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW

Paul C. Langley, Ph.D., Adjunct Professor, College of Pharmacy, University of Minnesota

Abstract

*Previous commentaries in **Innovations in Pharmacy** and other peer reviewed journals have made the case that the analytical framework, if that is not too strong a term, to support pricing and access recommendations endorsed by the Institute for Clinical and Economic Review (ICER) fails to meet the standards of normal science. By any criteria the ICER analysis is best described as pseudoscience; it fails the demarcation test between biological evolution and intelligent design. Like intelligent design it has its believers; a meme for all seasons. ICER is fully aware of the fact that it fails these standards, yet perseveres. It justifies its cost-per-QALY framework by denying, through unsubstantiated assertions, that it meets standards for scientific credibility; it denies the possibility of negative values and utilities which undercut completely the construction of QALYs. This is nonsense: not only does the ICER framework fail those standards, to include axioms of fundamental measurement, but also a simple rule of logic in basing its models on assumptions. ICER dogmatic adherence to simulation modeling is evidence in its latest report on multiple myeloma. The report is a charade; but unfortunately not one that is rejected by Bristol-Myer Squibb, GlaxoSmithKline, Sanofi and Amgen. Their responses to the draft evidence report points to their acceptance of imaginary simulation constructs to drive pricing decisions. Whether this reflects their unqualified acceptance of the imaginary simulation modelling to create evidence or a failure to appreciate the standards of normal science is unclear. Certainly, in this case they fail to recognize the devastating impact of believing in the use of the EQ-5D-5L preferences to create imaginary or I-QALYs. The question raised in this commentary is whether the willingness to accept the ICER analytical framework reflects a belief in the role of creating evidence, ICER style, or a willingness to accept ICER imaginary conclusions as the easy way out in negotiating prices with insurers and other payers. Accepting ICER imaginary constructs is an analytical dead end that will stifle the discovery of new facts. The question is: so what?*

INTRODUCTION

One of the more intriguing features of health technology assessment in the US is the position occupied by the Institute for Clinical and Economic Review (ICER) in clinical appraisals and, for our present purposes, modelling claims to support price and access recommendations. Previous commentaries in *Innovations in Pharmacy* have made the case that ICER's recommendations should be rejected as they

are based on assumption driven simulated evidence that fail the demarcation test of normal science^{1 2}. The recently released ICER report on *Anti B-Cell Maturation Antigen CAR T-cell and Antibody Drug Conjugate Therapy for Heavily Pre-treated Relapsed and Refractory Multiple Myeloma* is in this tradition³. The imaginary modelling, as briefly detailed here, clearly meets criteria for pseudoscience as do previous ICER evidence reports⁴. More importantly, however, is the position taken by manufacturers in their response to the draft evidence report. The manufacturers seem uninterested regarding the merits of the ICER analytical framework. All four of the manufacturers responding with a public comment – Bristol-Myer Squibb, GlaxoSmithKline, Sanofi and Amgen – seem perfectly content with ICER creating imaginary claims to support pricing and access recommendations. The manifest failings of the ICER methodology, indeed the wider belief system in approximate information, are either not appreciated or not recognized. The response of each manufacturer is to suggest assumption modification; to move the deckchairs when the iceberg approaches. This raises a critical issue: to what extent do manufacturers, or their technology appraisal teams share, with ICER, a deep seated belief in the approximate information modelling meme? If so, what are the implications? This is the focus of this commentary.

ICER REPORT FINDINGS

The imaginary ICER model was designed to produce empirically non-evaluable invented estimates of the lifetime cost-effectiveness of three CAR-T products: Idecabtagene vicleucel (Abecma, Bristol Myer Squibb; Bluebird Bio Inc) [Ide-Cel], Ciltacabtagene (Janssen, Legend Biotech)[Cilia-Cel] and Belantamab mafodotin-blmf (Blenrep, Glaxo Smith Klein). The model was ‘informed’ by key clinical trials, previous economic imaginary models and assumptions from the literature, and guesses from key opinion leaders. Within a decision analytic framework the hypothetical patient cohort for the model included all patients who were eligible for CAR-T therapy and who had undergone leukapheresis. Three health states were modelled: alive or progression free or responding to therapy; alive and not responding to therapy/subsequent relapse; and dead from multiple myeloma related complications or other causes. This is known as an absorbing state but one that is contradicted by the ability to create negative utilities or states worse than death. There is, perhaps, a Lazarus-like back door. Patients remained in the model until they died. Health state occupancy was derived from the application of partitioned survival techniques providing estimates of progression free survival and overall survival. Modelled imaginary outcomes included total life years gained, I-QALYs gained, equal value life years gained, time progression free and total imaginary costs over a lifetime horizon discounted by 3%. None of these outcomes is empirically evaluable.

Base case results compared each of the three therapies to a market basket. The modelled imaginary intervention costs, QALYs and life years are presented in Table 1. The apparent plausibility of these imaginary modelled estimates, even for the short time frames involved, disguise the fact that they fail the demarcation test between science and pseudoscience: none of the claims are empirically evaluable, let alone replicable. Indeed, with a judicious choice of assumptions they can be overturned. A possibility that is not put to rest by claims for sensitivity analyses to protect the ‘core’ imaginary model.

TABLE 1
ICER BASE CASE IMAGINARY NON-EVALUABLE MODELLED OUTCOMES FOR
MULTIPLE MYELOMA CAR-T THERAPIES

Treatment	Total Cost	I-QALYs	Life Years	Cost per I-QALY gained	Cost per life year gained
Ide-Cel	\$646,000	2.24	2.97	\$329,000	\$250,000
Comparator basket	\$276,000	1.08	1.50		
Cilta-Cel	\$617,000	3.40	4.52	\$147,000	\$113,000
Comparator Basket	\$276,000	1.08	1.50		
Belantamab	\$254,000	1.15	1.60	\$98,000	\$70,000
Comparator Basket	\$218,000	0.78	1.08		

Source: ICER Report Tables 4.4, 4.5, 4.6 and 4.7

The final step in the modelling is to apply cost-per-I-QALY threshold analysis and required price adjustment to meet the various cost-per-I-QALY thresholds. For an intermediate threshold of \$150,000 per I-QALY the required unit prices are:

- Ide-Cel (WAC \$419,500) required unit price \$245,000
- Cilta-Cel (WAC placeholder \$419,500) required unit price \$427,000
- Belantamab (WAC \$8,277) required unit price \$9,300

The result is no price discounts against WAC for Cita-Cel and Belantamab but discounts from WAC in range 42% – 54% for Ide-Cel. These thresholds fail to meet minimum requirements for fundamental measurement; as such they are meaningless.

While the I-QALY may be considered by some to be the ‘coin of the realm’ in cost-effectiveness analysis, the fact is the coin is counterfeit; a dud. It would have failed the re-coining of Sir Isaac Newton in the 1690s. This failure to meet required measurement standards means that any calculation involving the appropriately designated impossible or I-QALY is meaningless. If we add to this the he belief that logic can be put to one side in the modeling of assumptions, then the entire analysis is a waste of time.

FOUNDATION BELIEF IN PSEUDOSCIENCE

The strength and resistance of a belief system should not be underestimated. In health technology assessment the focus is not upon hypothesis testing and the discovery of new yet provisional facts regarding competing therapy interventions, but on the invention of approximate information ⁵. This sets technology assessment apart from normal science (including drug development) and the mainstream social sciences (e.g., education, psychology, economics). On all criteria, ICER technology assessment

shares the Dover courtroom with intelligent design. Failing the demarcation test between normal science and bunk, the discovery of new yet provisional facts, any focus on empirical evaluation is entirely foreign. Rather, the focus is on inventing evidence through lifetime imaginary simulation models. There is, unfortunately, a willing, if uninformed audience who accept the ISPOR relativist position that truth is consensus, even if it is invented ⁶.

Bizarre as it may seem, the meme for inventing evidence has survived for over 30 years. Textbooks support it ⁷; academic centers support it, together with global professional groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). ISPOR, in its commitment to technology assessment, has published numerous good practice guidelines for constructing imaginary worlds; but at no time has ISPOR published guidelines detailing the standards of normal science or the axioms of fundamental measurement. The creed must not be challenged. The young modeler applying ISPOR practice standards must subscribe to the status quo. After all, his or her future professional advancement depends on this acceptance of the ISPOR meme and the accolades of key ISPOR believers.

This widespread belief in the technology assessment meme to invent evidence, which has been described as a relativist rather than a realist position, creates barriers to any criticisms of the ICER reference case methodology. If advisers on technology assessment within companies share the belief in this approximate information meme with ICER, then it is not unreasonable that any criticisms of ICER will not see the wood for the trees: criticisms will focus on assumptions within the invented ICER simulation. The fact that an ICER imaginary determination can cost their employer hundreds of millions of dollars in lost sales will be of only passing interest; the commitment must be to the approximate information meme. This acceptance puts manufacturers in difficult situations where employees in health technology assessment, in subscribing to the approximate invented information meme effectively put their belief in the ICER imaginary modelling system ahead of the interests of the employer and the standards of normal science. The result is that criticisms of an ICER evidence report may be muted; rather than, as deserved, an outright rejection of the ICER claims. We find instead a series of minor challenges to the model assumptions. Even then, the fundamental issues of the scientific method are never addressed; notably the reliance on value or utility claims from direct and indirect preference instruments that overlook the presences in all instruments of negative values or health states worse than death.

ICER PSEUDOSCIENCE

The hallmarks of ICER pseudoscience may be summarized as follows:

- **Normal Science:** a rejection of the standards of normal science and acceptance of pseudoscience in rejecting the role of empirically evaluable claims in favor of inventing non-evaluable claims
- **Fundamental Measurement:** a rejection of the axioms of fundamental evidence in refusing to accept that generic multiattribute instruments produce only ordinal scores ⁸

- **Dominating Assertions:** defending its position by assertion rather than proof that preference scores from multiattribute instruments have ratio properties
- **Assumption Driven:** creating evidence from simulation models driven by assumptions despite the elementary logical point that what has happened in the past cannot lay claim to what might happen in the future⁹
- **States Worse than Death:** ignoring the fact that direct (standard gamble, time trade off) and indirect (EQ-5D-3L/5L, HUI Mk2/3) preference instruments produce states worse than death (negative utilities). This invalidates their application as other than ordinal scores. They cannot have ratio properties as the 0=death mark is contrived and meaningless. A ratio scale is defined by a true zero where no measurement is possible below this point (e.g., height, weight)^{10 11}.
- **Ordinal Utilities:** the failure to acknowledge that the preference scores are ordinal (e.g. TTO¹²); that is, you can rank the scores but have no idea of the distance between them. ICER asserts that they are either interval or ratio; to support the imaginary I-QALY analysis they have to be ratio.
- **Impossible QALY (I-QALY):** the failure to acknowledge (for obvious business reasons) that the I-QALY is an impossible mathematical construct as you cannot multiply time by an ordinal score to create a QALY¹³.
- **Single Attributes:** Attempting to combine symptoms or attributes of disease states into a single score (e.g., EQ-5D-3L/5L) violates the axioms of fundamental measurement as the so-called aggregate score lacks dimensional homogeneity, unidimensionality and construct validity (e.g., multiple criteria decision analysis scores)¹⁴. The standards in the physical science and the mature social sciences are to construct instruments to measure single attributes with required measurement properties. If required ratio measures can then be combined (e.g., body mass index combines two single attribute ratio measures: height and weight).
- **Multiple Models:** the ICER model is only one of a potential multitude of models that can provide competing claims, and even be engineered to produce such claims (this is called marketing)¹⁵.
- **Validation:** this is an old concept with the belief that it is possible to prove a claim; for the almost 100 years this has been rejected in favor of falsification. ICER is apparently unaware of the demise of logical positivism¹⁶. ICER validates its model and claims by (i) reviewing the structure of the model and (ii) comparing its model to similar models. At no stage is any consideration given to empirical evaluation of claims.
- **Time Horizon:** the ICER reference case asks us to believe in a invented future reality tracking hypothetical patient cohorts through the natural course of the disease with hypothetical responses based on limited pivotal trial data with 'realistic' assumptions driving non-evaluable claims for pricing and patient access for results for up to 30 or more years in the future
- **Induction:** a quaint lingering belief in logical positivism and a failure to appreciate Hume's problem of induction where, in logic, what has held in the past cannot be expected to hold in the future
- **Direct Medical Costs:** ICER believes it is possible to project selected direct medical costs (but not drug costs) years and decades into the future (appropriately discounted) based on just assumptions and previous cost estimates
- **Cost-per-QALY Thresholds:** as the I-QALY is an impossible construct then the notion of cost-per-QALY thresholds is also nonsensical as the basis for pricing recommendations

This is a remarkable litany of error and sheer nonsense. ICER was doomed from the outset in the invention of evidence to support imaginary pricing recommendations. Whether by accident or design, the ICER reference case shows a disregard of the standards of normal science, in particular the axioms of fundamental measurement and simple logic¹⁷. We have known for over 30 years, to give one example, that the various direct and indirect preference instruments yield negative values and utilities. Presumably ICER never received the memo.

SAVAGING A DEAD SHEEP

To characterize the ICER methodology as a metaphorical 'dead sheep' is no exaggeration. As detailed above, the various evidence report models should not be taken seriously. But the departed sheep is amazingly resilient. ICER is aware of the criticisms, but cannot reply other than through a series of unsupported assertions^{18 19}. This raises a key question: does ICER recognize the manifest deficiencies of its methodology or does it truly believe in the importance of inventing non-evaluable evidence? Does its continued application of this simulation framework across disease states rest on a deeply held belief or a belief that, from a more cynical view, it is useful just for business purposes to an audience of faux believers?

A recent commentary has provide a deconstruction of the belief system held by ICER as evidenced in ICER responses to critiques of their methodology and, most recently, a series of assertions by ICER that apparently are all that hold the ICER methodology together². The weakest link in the ICER house of cards is the denial of the axioms of fundamental evidence as they apply to the invention of QALY based claims. ICER asserts, with no supporting evidence whatsoever, that multiattribute utility or preference scores have ratio measurement properties. This act of faith is fundamental to their belief in imaginary modelling. In this they are joined by the majority of those in health technology assessment. Yet there is no evidence that utility scores such as the EQ-5D-3L have anything other than ordinal properties. That is, scores can be ranked but the distance between scores is unknown. The simplest demonstration is to point out that the EQ-5D-3L and other multiattribute scores, not to mention standard gamble (SG) and time trade-off (TTO) claims, can support negative values or health states worse than death. This is a common characteristic of all direct and indirect preference instruments. Theoretically, if an algorithm produces negative values then the lower limit is not the lowest negative score but minus infinity. Indeed the denominator for the scores that can range from unity to minus infinity is infinity. This has an interesting outcome: attempting to rescale preference scores to a zero to unity scale fails because dividing any ordinal number by infinity always yields (in the limit) zero.

As Bond and Cox point out in their own deconstruction of the difference between models where the observed data have primacy and those models where the requirements of the model have primacy, as evidence by Rasch Measurement Theory (RMT). If you want your model to conform to the axioms of fundamental measurement, typically interval scores, then items for that measure have to be selected to meet measurement requirements (p. 303)²⁰. If the results of an analysis, such as the creation of preference scores, are descriptive of those data then they are merely exploratory; they will only produce

ordinal scores. They have to account for all of the data, which means a process of successive fitting (tweaking) of a model to the data such as the rules determining the EQ-5D-3L preference algorithm or the HUI Mk 2 preference algorithm. The result is a 'forced fit' with unavoidable 'left overs' in the form of negative utilities. One solution, as in the case of the HUI Mk2/3 is to hope no one is looking and rejig all negative scores to zero. This illustrates the fact that there is no true zero in the scale but a contrived zero we call death (or something worse). For those who believe in conforming to the axioms of fundamental measurement, to develop confirmatory and predictive models, the required task is to ensure item fit to the model. In Rasch modelling the focus is on the size and structure of residuals to ensure that the underlying principles for conjoint simultaneous measurement, to capture latent attributes, can be justified to create a measurement scale with invariant, interval level properties. There exist a large number of disease specific patient centric need fulfillment quality of life instruments with Rasch properties. More recently, it has been demonstrated how the output from these instruments can be transformed to bounded ratio scales to support needs fulfillment quality of life claims and create N-QOLS with the required measurement properties ²¹.

Given the argument that the utility ordinal scores are the weak link, it is surprising that more attention is not given by manufacturers, should they choose to do so, to the utility scores utilized in the ICER modelling. In the case of this Report the evidence base for the choice of utilities is limited to those from a single trial the *Efficacy and Safety Phase 2 Study of bb2121 in subjects with relapsed and refractory multiple myeloma (KarMMA)* [NCT03361748]. This study (n=149) used the EQ-5D-5L instrument but as a secondary (underpowered) endpoint. These results were presented as a paper at the American Society for Hematology Meeting in 2020; there is no evidence for any peer reviewed publication since then (PubMed search 15 May 2021) ²².

Belief in a ratio preference score without a true zero is patent nonsense. Yet none of the responses by these manufacturers recognize this; they still accept the QALY as a meaningful, if impossible, construct. In the case of the EQ-5D-5L we have known since its launch in 2009 that it can generate negative preferences; a situation no different from the EQ-5D-3L. Following some ten years of disputes over its application, culminating in the decision by NICE in 2019 to reject its application in imaginary simulation modelling (with the requirement to map to the EQ-5D-3L equivalent scores). None of the manufacturer responses appeared to appreciate the limitations of this scale. This may be, of course, just a failure to keep up with the literature, but there is a key analysis in the ISPOR house journal *Value in Health* published in 2019 which details the extent to which negative preferences appear ²³. From an assessment of EQ-5D-5L value sets covering five countries (Singapore, the Netherlands, China, Thailand, Canada) the study found that between 9% and 33% of health states were given negative values.

SUPPING WITH THE DEVIL

It is surprising how many manufacturers support ICER financially, engaging with ICER in the creation of its imaginary evidence modelling. It is not as though there is an expectation of 'being gentle' in a review; ICER has effectively killed this expectation through its modelling of claims and willingness to demand

substantial price discounts and patient access. This is entirely appropriate given its business model. What is not appropriate, from the position of third parties expecting to benefit from approvals of new products, is the insistence by ICER to develop modelled claims from limited evidence with the excuse that as more evidence emerges the models assumptions may be varied to produce alternative recommendations for pricing and access. By which time the damage will have been done if payers take recommendations at face value. The best ICER can offer is to utilize *ICERAnalytics*, a new cloud based platform where manufacturers can modify ICER modeled assumptions and come to different imaginary outcomes. An exercise in futility as it points to the possibility of any model being re-engineered to come to a required set of claims. Manufacturers could demand ICER release the model platform for multiple myeloma through *ICERAnalytics*, but this would probably be refused. Any release could be years in the making for what will be a wasted exercise of challenging assumptions. After all, why give the manufacturer any more ammunition that what is unavoidably necessary?

So why sup with the Devil? If the spoon is not long enough, why not ignore ICER altogether? Given the manifest deficiencies of the ICER approach, this would seem a sensible decision. Yet engagement persists. One possibility is that if your advisers and ICER share a common imaginary information belief system, you will focus on the trees and not the wood: criticism is restricted to challenging assumptions not the invented evidence methodology. It is difficult not to underestimate the strength of a belief system which rejects as an article of faith normal science for pseudoscience, inventing approximate evidence to support formulary decisions, and its impact on value claims. This has been the dominant meme for over 30 years endorsed by the leaders in health technology assessment, professional groups such as ISPOR, health departments in single payer systems and academic centers. The transmission fidelity of the meme to generations of students and the acceptance of mysteries such as the I-QALY have been remarkable²⁴. In the ICER world truth is consensus.

MANUFACTURER RESPONSES

If we consider the tone and content of manufacturer responses to, initially, the draft evidence report²⁵ and then the final evidence report, it is clear that for those responding there is an acceptance of the approximate information meme and the role of simulated imaginary evidence. There are no suggestions that the standards of normal science, including fundamental evidence and rules of elementary logic, are of relevance. This is puzzling because in drug development, manufacturers are scrupulous in following the required evaluation standards of normal science. Apparently this transmutes to a belief in inventing evidence once ICER enters center stage, like the demon king rising through a stage trapdoor surrounded by a fog of dry ice.

- **Bristol-Myers Squibb**

Perhaps the most pertinent comment by BMS concerns the limited data at product launch. If data are limited then this may lead to an inappropriate assessment of value which then factors into the pricing of medicines, with long-term data continuing to evolve with both new trials and the ongoing release of

data from existing trials. One concern being an understatement of benefits from ide-cel in terms of overall survival and progression-free survival. ICER gives its stock response: we need to release our recommendations even if the data are limited and we need to guess input assumptions to support modelling decisions. New data can always be introduced later. Apparently where there is uncertainty, sensitivity analysis can always allay our fears.

In addition to a number of technical points regarding ICER's modelling assumptions (which were apparently addressed) other points concerned transparency and consistency in modelling. While BMS clearly spent a long time reviewing the assumptions and techniques within the modelling framework, at no time was there a concern expressed as to the scientific status of the ICER methodology. There were no concerns expressed regarding, as noted above, the status of quality of life claims and the mathematical impossibility of constructing QALYs.

- **GlaxoSmithKline**

The GSK submission focuses, once again on the interpretation of the clinical data and assumptions. Worth noting is the debate over HRQoL results. A major concern is that the data are limited giving an incomplete picture of possible long term improvements. The balance of comments related to the need for text clarification in respect of citations, vision effects and inconsistent descriptions of the treated population. Once again, there is no evidence that GSK are aware of the lack of scientific status of the ICER modelling.

- **Sanofi**

A primary focus of the Sanofi review is on the needs for consistency between protocol designs, population characteristics, and the reporting of trial results. ICER's response to a series of detailed questions on methodological transitivity are ignored in favor of a blanket response defending the use of a limited hypothetical population in modeling. The net result is a minimum change involving deletion of references. The focus is primarily on the methodology of reporting clinical data; not on the model. Again, there is no evidence that Sanofi were aware of the lack of scientific status of the ICER modelling.

- **Amgen**

Amgen's principal concern is with assumptions regarding overall survival estimates in the ICER model. ICER agrees that there are a wide range of potential outcomes, although modifications 'appear' to provide more consistent imaginary estimates. A further issue is consistency in patient populations across clinical trials. A further issue raised concerned utilities. Amgen is not apparently concerned with the measurement status of these instruments but that the possibility of adjustments to accommodate patients possibly refractory to therapy options. ICER demurs; pointing out that data on health state utilities are limited and we have to do with the limited data available. Presumably this is the ICER 'any port in a storm' defense. The comments then proceed to the role of 'as treated' as opposed to an 'intent

to treat' population in the modeled claims; ICER again demurs arguing that an 'as treated' approach would result in biases within the modelled claims. Further concerns were expressed regarding cost assumptions and the basis for cost estimates (e.g., assumed time in hospital and readmission). ICER's response is to agree that previous costs from real world evidence to support modeled assumptions is to be preferred to unsupported assumptions, although this once again fails to see the logical error.

Amgen, in common with other manufacturers, does not challenge the ICER approximate information imaginary simulation framework. The concern is with assumptions. This seems an odd preoccupation given the well documented lack of scientific merit in the ICER methodology and the logical error in basing future events on past assumptions.

- **Overview**

The common theme in all four critiques is not to challenge the overall merits of inventing evidence to support cost-effectiveness claims but to suggest how changing assumptions might 'improve' the invented imaginary comparative cost-effectiveness and pricing claims. The focus is on modifying assumptions. It is not clear, given the status of invented evidence, what this is intended to achieve; possibly a comparison of modelled assumptions without any evaluable claims that might support choice of assumption?

More concerning is the possibility that ICER is being given a free pass. In none of these critiques is there any attempt to challenge what we may call fundamental assumptions. As noted above, the weakest link in the ICER house of cards is the failure to recognize that both direct and indirect preference measures can generate negative scores. The implication is clear: there is no true zero for any of these instruments which mean none can support a ratio scale or the construction of QALYs. There is no way around this as the term 'true zero' in measurement means that a scale with a true zero can never have negative values. It only requires, for these preference instruments for one respondent to present with a negative value for any health state description for the scales to be, by default, an ordinal scale. Certainly, interval scales can have negative values (as a zero point is arbitrary) but then it has to be shown to have interval properties by construction.

What are we to conclude? The simplest explanation is that the respective manufacturers lack trained resources to challenge ICER. Less appealing is the argument that the manufacturer's staff buy into the ICER approximate invented evidence meme and see no need to raise questions that they consider irrelevant. If this is the case, then manufacturers will have to reflect on the implications of this for staffing, pricing and lost revenue. The fact is that the criticisms raised here about the lack of appreciation of the standards of normal science are not new; we have been aware of them for the past 30 or more years. Of course, a further explanation for these selective criticisms of assumptions is that manufacturers view ICER as mistaken yet necessary; an easily surmountable obstacle to pricing and access. The ICER model and criticisms of ICER are just a *rite de passage* (similar to being interrogated by

a Senate committee); the critical issue is to negotiate and contract. ICER can easily be put to one side; who remembers (or understands) an ICER final evidence report model?

MEDIA RESPONSE

It is clear that the media have only a superficial understanding of the ICER modelling framework. The ICER assumptions are taken at face value. They parrot the ICER media releases. FiercePharma, for example, just provides a summary of press releases. It does not, apparently, have time for a more informed analysis. At no stage is there any substantive criticisms of the ICER analytical framework. The fundamentals of the ICER analysis in the focus on the limited evidence from clinical trials, the focus on generic multiattribute preference scores and the incremental cost-per-QALY model remain unchanged. The criticisms of the I-QALY are ignored; the fact that the I-QALY is an impossible mathematical construct fails to resonate.

CONCLUSIONS

It is unfortunate, but must be a major consideration, that holding to a common belief in the approximate imaginary simulated assumption driven modelling belief system, while not discouraging debate and engagement between ICER and manufacturers, may either by accident or design put the standards of normal science to one side, accepting mysteries such as the impossible or I-QALY as articles of faith and sustained belief. Supping with the devil is more palatable if you share a belief; a long spoon may not be necessary, particularly with rank mutton on the menu, as long as you share the mint sauce.

It is difficult to judge the likelihood of rejecting approximate imaginary information as the common belief system in favor of an acceptance of the standards of normal science in evaluating ICER modelled claims for pricing and patient access. If this final evidence report is any guide, the commonly held belief system is well entrenched. Perhaps salvation lies with third parties who, in rejecting the QALY for reasons unrelated to negative preferences, restrict significantly the application in formulary decision making of incremental imaginary cost-per-QALY models. This may lead to consideration of the role of the formulary committee as the agent for patients and caregivers; a potentially beneficial role if the formulary committee is to set standards for submissions for formulary negotiations and approval. ICER has taken on this role as arbitrator, but through the application of a belief system that is clearly nonsensical.

REFERENCES

¹ Langley P. Nonsense on Stilts – Part 1: The ICER 2020-2023 value assessment framework for constructing imaginary worlds. *InovPharm*. 2020;11(1):No. 12

-
- ² Langley P. Peter Rabbit is a Badger in Disguise: Deconstructing the Belief System of the Institute for Clinical and Economic Review. *InovPharm*. 2021; 12(2): No 22
<https://pubs.lib.umn.edu/index.php/innovations/article/view/3992/2855>
- ³ Lee SJ, McQueen RB, Beinfeld M, Fluetsch N, Whittington MD, Pearson SD, Ollendorf DA. Anti B-Cell Maturation Antigen CAR T-cell and Antibody Drug Conjugate Therapy for Heavily Pre-Treated Relapsed and Refractory Multiple Myeloma; Final Evidence Report. Institute for Clinical and Economic Review, May 11, 2021.
<https://icer.org/assessment/multiplemyeloma-2021/#timeline>
- ⁴ Pigliucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010
- ⁵ Neumann PJ, Willke R, Garrison LP. A Health Economics Approach to US Value Assessment Frameworks – Introduction: An ISPOR Special Task Force Report. *Value Health*. 2018;21:119-123
- ⁶ Wootton D. The Invention of Science: A new history of the scientific revolution. New York: Harper Collins, 2015
- ⁷ Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes. 4th ED. New York: Oxford University press, 2015
- ⁸ Langley P, McKenna S. Fundamental Measurement and Quality Adjusted Life Years. *Value Health*. 2021;24(3):461[letter]
- ⁹ Magee B. Popper. London: Fontana, 1974
- ¹⁰ Patrick D, Starks H, Cain K et al. Measuring preferences for health states worse than death. *Med Decis Making*. 1994;14:9-18
- ¹¹ Gandhi M, Rand K, Luo N. Valuation of health states considered to be worse than death – an analysis of composite Time Trade-Off data from 5 EQ-5D-5L valuation studies. *Value Health*. 2019;22:370-6
- ¹² Lugnér AK, Krabbe PF. An overview of the time trade-off method: concept, foundation, and the evaluation of distorting factors in putting a value on health. *Exp Rev Pharmacoecon Outcomes Res*, 2020;20(4): 331-42
- ¹³ Langley P. The Great I-QALY Disaster. *InovPharm*. 2020; 11(3): No 7
<https://pubs.lib.umn.edu/index.php/innovations/article/view/3359/2517>
- ¹⁴ McKenna S, Heaney A. Composite outcome measurement in clinical research: the triumph of illusion over reality. *J Med Econ*. 2020; 23(10):1196-1204
- ¹⁵ Langley P. Let a Thousand Models Bloom: ICER Analytics Opens the Floodgates to Cloud Pseudoscience. *InovPharm*. 2021;12(1): No. 5 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3606/2668>
- ¹⁶ Lakatos I. Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of Knowledge*. I Lakatos, A Musgrave. Eds. London: Cambridge University Press, 1970
- ¹⁷ Langley PC and McKenna SP. Measurement, modeling and QALYs. *F1000Research*. 2020; 9: 1048 <https://doi.org/10.12688/f1000research.25039.1>
- ¹⁸ Langley P. The Impossible QALY and the Denial of Fundamental Measurement: Rejecting the University of Washington Value Assessment of Targeted Immune Modulators (TIMS) in Ulcerative Colitis for the Institute for

Clinical and Economic Review (ICER). *InovPharm*. 2020;11(3): No. 17 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3330/2533>

¹⁹ Langley P. To Dream the Impossible Dream: The Commitment by the Institute for Clinical and Economic Review to Rewrite the Axioms of Fundamental Measurement for Hemophilia A and Bladder Cancer Value Claims. *InovPharm*. 2020;11(4):No. 22 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3585/2642>

²⁰ Bond T, Fox C. Applying the Rasch Model: Fundamental measurement in the human sciences. 3rd Ed. New York: Routledge, 2015

²¹ Langley P. McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*. 2021;12(2):No. 6 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3798>

²² Delforge M, Shah N, Rodriguez-Otero P, et al. 3465 Health State Utility Valuation in Patients with Triple-Class-Exposed Relapsed and Refractory Multiple Myeloma Treated with the BCMADirected CAR T Cell Therapy, Idecabtagene Vicleucel (ide-cel, bb2121): Results from the KarMMa Trial. Paper presented at American Society of Hematology; 2020.

²³ Gandhi M, Rand K, Luo N. Valuation of health states considered to be worse than death – An analysis of composite Time Trade-Off Data from EQ-5D-5L valuation studies. *Value Health*. 2019;22:370-75

²⁴ Dawkins R. A Devil's Chaplain. New York: Houghton Mifflin: 2004

²⁵ ICER. Response to Public Comments on Draft Evidence Report. April 5, 2021. https://icer.org/wp-content/uploads/2020/10/ICER_Multiple-Myeloma_Response-to-Comments_040521.pdf