MAIMON WORKING PAPERS No. 18 SEPTEMBER 2021

MORE NONSENSE ON STILTS: IMAGINARY CLAIMS BY THE INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW AND THE UNIVERSITY OF ILLINOIS AT CHICAGO, COLLEGE OF PHARMACY EXPERT IMAGINARY SIMULATION GROUP FOR THE PRICING OF ECULIZUMAB AND EFGARTIGIMOD IN MYASTHENIA GRAVIS

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

Abstract

The recently released Institute for Clinical an Economic Review's final evidence report for Eculizumab and Efgartigimod in Myasthenia Gravis is one more example of a commitment to the invention of evidence to support pricing and formulary decisions. This is the ICER business mode; without clinging to a belief system that rejects the standards of normal science and, more spectacularly, a fervid belief in the mystical ratio properties of ordinal scales that rejects the axioms of fundamental evidence, ICER would wither and die. Unfortunately there are many who share ICER's beliefs, including academic research centers and health decision makers who should know better. Whether this shared belief is by accident, ignorance or design is an open question; the fact is that it is widely held to the detriment of patients where insurers and health system latch upon ICER pronouncements as support for denial of care. This amounts to a neo-eugenic application of a bankrupt methodology.

The purpose of this brief commentary is to focus on two aspects of the ICER invented evidence modeling: the downsides of assumption driven simulation modeling, the potential creation of a multitude of competing models, and the application of EQ-5D-5L multiattribute preferences where in the misguided attempt to create quality adjusted life years (QALYs) the model builders overlook the fact that the preferences are ordinal and the QALY is a mathematical impossibility.

The result is that the ICER pricing claims for Eculizumab and Efgartigimod in Myasthenia Gravis are nonsense. Not only should the respective manufacturers ignore them, but this message needs to be transmitted to health system decision makers before care is denied and patients harmed.

INTRODUCTION

The manifest deficiencies of the belief system in health technology assessment shared by ICER and others has been extensively documented. It has been described as a meme, with high transmission fidelity; one that has persisted for some 40 years because of a decision to invent evidence to support claims for cost effectiveness at product launch. Given the option of establishing provisional pricing and access criteria subject to a targeted real world evidence research program to fill in gaps and provide an ongoing framework for disease area and therapeutic class reviews, the leaders in technology assessment (typically academic) rejected hypothesis testing in favor of modeling and the invention of evidence to support cost-effectiveness claims at product launch. One interpretation is that this approach, where claims could never be evaluated empirically, led to a win-win consulting environment; fees were earned for claims that stretched over future decades and which were impossible to justify, apart from a nuanced belief in the realism of assumptions.

It should be made clear that the ICER assumption driven simulation modelling shortchanges both patients and manufacturers. Hiding behind a smokescreen that what is done is 'an established standard and everyone does it' (which is false), ICER can continue to dominate pricing and access decisions. In part ICER's 'success' rests on the limited awareness of many of its audience members of the standards for normal science. Just as many Americans (not quite a majority) believe in creationism (rebranded as intelligent design) we cannot, quite reasonably, ask what demarcates science from pseudoscience, where ICER joints intelligent design in the Dover courtroom. But we must make the effort, not only to emphasize that science is concerned with the discovery of new yet provisional facts through a modified process of conjecture and refutation where we evaluate credible, empirically evaluable and replicable competing therapeutic claims.

Central to this process of discovery is measurement; unless our instruments for evaluating response to therapy meet the required standards we can make no claim for the superiority of one therapy over another. Again, of course, we cannot expect (?) our audience to understand the importance of distinguishing ordinal from ratio scales; or even the distinction between ratio and interval scales. But if we don't, we end up believing that the EQ-5D-5L preference scores are actually ratio scales in disguise (a strongly held ICER belief).

It is into this charade of false belief that manufacturers are cast once ICER decides to make an example of them. In all fairness, manufacturers are all too often on the back foot as, with few adequately trained internal resources, there is nowhere to turn to challenge ICER from day one [1]. This is not helped by the possibility that the ICER invented evidence meme is shared by

The very staff employed by the manufacturer who should be defending the employer's interests.

ICER's failure to come to grips with normal science is seen in the Health Benefit Price Benchmark (HBPB). This relies entirely on assumption and false belief in measurement theory. As will be noted here, there can be a multitude of HBPB's each driven by different models and assumptions in the same disease area and for the same therapies.

This commentary is in five parts. First, a brief overview of the ICER simulated imaginary model; second, a short introduction to the required scientific standards with particular reference to measurement; third, a deconstruction of the centerpiece of the ICER model, the EQ-5D-5L preference score and the imaginary QALY; fourth, the imaginary multiverse of modeled claims; and finally a brief statement of the required standards for evaluating therapeutic claims.

THE SIMULATED ICER IMAGINATION

The apparent purpose of the ICER modelling, undertaken by the expert modelling group at the College of Pharmacy is to create the imaginary cost-effectiveness of eculizumab and efgartogimod with each added to conventional therapy by assumption versus conventional therapy alone. The base case analysis to invent evidence for cost-effectiveness claims used a two-year four state Markov model (Markov was shot by Stalin) where a simulated cohort of hypothetical patients entered the simulated imaginary model, assumed to be unimproved on initial treatment, receiving eculizumab plus conventional therapy or conventional therapy alone. The primary outcome measure, assumed not observed, was quantitative myasthenia gravis score (QMG). The QMG score it should be noted is a multiattribute score than fails the standards of fundamental measurement; it is a ordinal score; any comparisons of one ordinal score with another are invalid as are claims for transition matrices based on ordinal OMG scores are also invalid. Unfortunately, due to lack of familiarity with measurement standards, the transition matrices in the expert model are useless as they rely on impossible to measure 'three point improvements' as the scale is ordinal.

Unsurprisingly, there is no apparent awareness by this expert group of the standards of normal science and the role of measurement; a position shared in the leading textbook [2]. Apparently, in agreement with the dominant ICER meme they reject the standards of normal science, rejecting any notion of the discovery of new facts, preferring instead to recycle previous assumptions ('facts') to create claims for the future. This violates a simple logical point that was made by the Scottish philosopher David Hume (1711-76), the problem of

induction [3], that *It cannot be established by observation, since we cannot observe future events. And it cannot be established by logical argument, since from the fact that all past futures have resembled past pasts it does not follow that all future futures will resemble future pasts [4].*

Perhaps we should, for the benefit of the College of Pharmacy expert modelers, make clear that truth, as proposed by ICER is not, as Wittgenstein asserts, just consensus. If truth for the experts and the other academic consulting centers supported by the coffers of ICER, is consensus then, as Wootton points out it is 'a notion incompatible with an understanding of one of the fundamental things science does, which is to show that a consensus view must be abandoned when it is at odds with the evidence'[5] . Unfortunately, the approximate imaginary modeling for claims endorsed by the College of Pharmacy effectively precludes the first steps in the discovery of new facts. There can be no appeal to superior evidence; we are in a Taliban world of dubious (if not patently false) beliefs and assertions.

The result of this assumption driven Markov simulation is the imaginary claim that on the basis of a series of assumptions the experts at the College of Pharmacy ask for a discount of $97% to 98% from the Federal Supply Schedule to achieve a HBPB price range for both products. The expert group appears not to understand the required standards for normal science in making claims but, more perniciously, the limitations imposed by the axioms of fundamental measurement.

In its rush to invent evidence to support the mythical HBPB ICER and the expert group put a major obstacle in the way of meaningful pricing options and access to care. These pricing claims are actually taken seriously and then, by the time a coherent evidence based argument is in place, the damage has been done. As will be conclusively demonstrated here these conclusions and the analysis undertaken by the experts (under contract to ICER) is arrant nonsense. The querulous reader might raise the question, if there is only a two-year time horizon, why not produce empirically evaluable claims instead of making them deliberately non-viable. It seems apparent that the expert imaginary simulation group, for whatever strange reason, fail to appreciate the standards required in science as opposed to pseudoscience, notably in respect of an ignorance (or lack of awareness) of the axioms of fundamental measurement. Their agreement with the ICER meme is undeniable; it is, unfortunately misplaced.

REQUIRED SCIENTIFIC STANDARDS: ICER's IMAGINARY OUTCOMES

Formulary decisions must rest upon claims for therapy impact, notably comparative claims that are consistent with the standards of normal science and the axioms of fundamental measurement. Instead, after 30 plus years of health technology assessment we face exactly the opposite commitment: therapy claims that are contrived in their focus on inventing evidence, on pseudoscience, and the implicit rejection of any concern with meeting the standards of fundamental evidence [6]. This was a deliberate choice; in the early 1990s, with the focus on a single, comprehensive claim for cost-effectiveness the multiattribute preference score entered center stage with the QALY acclaimed as the only valid construct to support modelling with pricing and access recommendations[7] . We now realize that this is, as noted, arrant nonsense.

We now realize with the benefit of hindsight and a better appreciation of the measurement standards that apply in the physical sciences, education and psychology that the decision in favor of approximate invented information in lieu of hypothesis testing to generate new evidence, was a major error, setting back for decades a commitment to the standards of normal science. Unfortunately, a number of key organizations and academic modelling centers have yet to receive, let alone understand, the memo.

The ICER assumption driven simulation model produces six imaginary outcomes; claims that are not credible, empirically evaluable or replicable. This is inevitable, given ICER's commitment to imaginary models. These ersatz outcomes are:

- **Total drug costs**
- **Total costs**
- **Quality Adjusted Life Years (QALYs)**
- **Life years**
- **Expected value of life years (evLY)**
- **Expected value of life years gained (evLYG)**

If a simulation model is designed to project claims decade into the future then it fails at the first hurdle to meet the standards of normal science. The ICER drug and total costs are a total fabrication not only because they are a patchwork of assumptions built on present and prior costs but because the notion of cost in terms of the claimed 'cost-effectiveness' is not anchored to any unit cost classification (e.g., CPT codes). Costs, for ICER, are what you assume them to be; there is no intention that they should ever have to face a credibility check.

ICER's commitment to the mathematically impossible QALY dooms both aggregate estimates of lifetime QALYs and cost-per-QALY thresholds but also the equal value of life year metrics (evLY and evLYG) as QALY estimates are integral to both measures. If one assumption (or belief) dooms the QALY based claims it is ICERs dogmatic insistence that the ordinal preference scales generated by direct and indirect or multiattribute instruments are ratio scales in disguise. This mystical transformation (where ICER cannot provide a proof or justification) means, by the axioms of fundamental measurement, that these preference scales, which lack construct validity, can only support nonparametric statistics. Which means no multiplication of time spent in disease states to create QALYs. At the same time, as the preference instruments create negative scores or health states worse than death, the scales have no true zero. Unless this condition is met there can be no multiplication or division; if in doubt think number lines and the rules for positive and negative numbers.

It is worth noting that the breakthrough ICER construct of perfect health (currently 0.851 on a scale of 0 to 1) is a classic example of misunderstood fundamental measurement. You cannot 'adjust' preference scores by each other if you are trying to standardize for age and gender differences between health states. It is disallowed as the preference score is ordinal. Also, the preference score has negative values for health states. ICER gives no indication of how these are accommodated in the 'aggregate' measure. This completely invalidates the evLY and evLYG outcomes.

Life years are the product of Markov states and the transition probabilities association with those states (including the one way transition to death). Any redesign of the Markov states and transition probabilities will lead to changes in life year claims. The life years are entirely imaginary, created by the assumption driven simulation model.

Response to therapy is also an impossible claim. Therapy response requires either a ratio scale or, as a minimum, an interval scale where there is an invariance of comparisons between points on the scale. In the ICER simulation framework the absence of an interval scale means that any claims for response to therapy are impossible. The ratio scale has interval properties as it is the addition of a true zero that allows the full range of arithmetic operations.

ICER cannot fall back on non-parametric statistics (to review changes in rank orders) because both direct and indirect preference scales lack construct validity. They represent health state descriptions that cobble together a mix of clinical attributes and ordinal response levels (e. g., Likert scales). They are not unidimensional (i.e., the lack dimensional homogeneity). We might believe that comparing health states is valid, but with a mix of attributes typically

measured on ordinal scales the exercise is a waste of time. Trying to decipher the attribute components of change is just nonsense. Response to therapy should be in terms of a single attribute or construct (latent or otherwise) with, if attributes are to be combined, each having ratio properties (e.g., body mass index, a ratio measure, combines two ratio measures, weight and height).

Given the manifest failures and falsification that characterizes the ICER reference case model, it is difficult to understand how it has survived (and indeed embraced) by those who should know better. But they have set their sights in a different direction. To create assumption driven simulated claims that have no chance of empirical evaluation

PREFERENCES AND MEASUREMENT

It is always a puzzle as to the decisions of journal editors and peer reviewers when they come of judge the applications of preference scores in cost-effectiveness claims. It appears, in the present case, that In respect of measurement theory, that ICER and the expert College of Pharmacy group have little if any clue as the constraints imposed by the levels of evidence required in statistical analysis. This lack of awareness is shared by the Editor of the *Journal of Neurology* (and its peer reviewers), notable because the journal appears to accommodate a number of papers focused on ordinal measure scores in disease area, including the EQ-5D-5L; an entirely wasted effort.

Notable among these is a paper (with supplementary information) that provides the expert modelers wit ordinal EQ-5D-5L scores to populate the simulation. None of the authors appear, apparently, to have a sufficient grasp of measurement theory to recognize that the EQ-5D-5L is a failed instrument producing ordinal and not the required ratio measure with a scale from 0 = death to 1 = perfect health; defined in turn by a limited symptom list and response levels determined as ordinal scales within the model [8].

The EQ-5D-5L preference scores care central to the simulated claims; it is important to deconstruct and abandon this preference measure. The first point to note is that it is, by intent, a multiattribute instrument with 5 symptoms or attributes and 5 response levels for each symptom. These responses are attached to a community preference weight which, through a translation to a weighted algorithm create an ordinal score. The intriguing feature of the EQ-5D-5L is that of the 3,125 health states (combination 5^5) 625 or 20% are valued as negative score or states worse than death; at point not alluded to in the evidence report. Second, it is invalid to add together scores that are not on a ratio single attribute scale.

The fatal fly in the ointment is the ability, at least in the US valuation, to create negative scores or states worse than death. While not raised in the ICER report by the expert group, this is the final objection: a ratio preference score (presumably capped at unity) cannot have negative health states. If it actually is a ratio scale then it must have a true zero with no possibility whatsoever of negative preferences.

If the expert group comes to (reluctantly) recognize this requirement, then the modeling edifice collapses as the QALY is an impossible mathematical construct. This invalidates incremental cost per QALY (imaginary) claims as well as the application of cost-per-QALY thresholds to create the ICER invented HBPB. The HBPB is mathematically impossible; planes fly, magic carpets and broomsticks don't.

THE IMAGINARY MULTIVERSE

If modeled simulated assumption driven claims can be invented under one scenario, then they can be contrasted to a potential multitude of other modeled scenario claims given a change in assumptions. There only one caveat: unless ICER and the College of Pharmacy expert modelling simulation group can claim that their assumption choice sets them aside from any other potential model, either now or in the future, they are just one among many. Irrespective of the claims that can be challenged and the imaginary claims disputed (but not in terms of real word outcomes), we are reduced to a playground debate over assumptions. Sensitivity analyses, let alone that technical masterpiece, probabilistic sensitivity analysis, will not save the day; any one of the potential for thousands of other modeled claims can be defended in precisely the same terms.

Interestingly, ICER has shot itself in the foot with the release of the cloud ICERAnalytics software system. Ostensibly defended in terms of transparency and the ability of decision makers in a health care system to assess (tweak) the impact of changing model parameters (i.e. assumptions) in the ICER model that supported claims for denial of care and restricted access criteria. The possibility, which will no doubt be exploited, is to reverse ICER claims by a judicious choice of assumption.

If there are an infinite, or at least, a potential multiverse of modeled claims then we will never be able to create and compare therapy options. In the case of eculizumab and efgartogimod any ICER claim cannot be considered even provisional; just one of a multiverse of competing model claims which can each produce an infinite series of progeny, each of which can produce a series of progeny through changing assumptions, each with its impossible HBPB claim. ICER (and the expert modelers are in a bind: they cannot claim a

unique status for their model because there are no criteria for uniqueness that will apply now and for all future modifications this model.

There is a pervasive failure to recognize the manifest limitations of health technology assessment modeling and the required standards for coherent value claims [9]. These include a mistaken belief in generic quality adjusted life year claims (QALYs) as well as disease specific PRO measures. At the same time assumption driven lifetime simulation models also fail these standards, notably those claims for pricing and access produced by the Institute of Clinical and Economic Review (ICER) [10]. We are in the unfortunate situation; after 30 years of QALYs and imaginary simulation modeling we now realize we cannot invent evidence to support formulary submissions.

RASCH MEASUREMENT AND STATISTICAL MODELLING

In the social sciences statistical modelling is the dominant analytical techniques to describe a data set. The object is to fit the model to the data, if necessary by the rejection of potential explanatory variables. This stands in contrast to the physical sciences where the measurement task is to obtain data that fit the model. The requirements of the model, construct or trait that is to be measured drive data collection and item selection. The distinction is between exploratory/descriptive models, fitted to the data (e.g., econometric modelling), and confirmatory/predictive models utilizing probabilistic conjoint measurement, where the requirement is for the data to fit the model. It is this latter approach that drives the Rasch model [11]. This leads to the consideration of the measurement properties of the required instrument. In human subject research where the objective is to measure latent traits we must start with a substantive theory about what it is we are trying to measure. Item development and selection must be driven by our knowledge of the latent trait. It may turn out that the latent trait is not actually quantitative. At the moment the Rasch model is the only one available to test the hypothesis that we are measuring a quantitative latent trait. Two questions are central to this: (i) how well do the empirical data fit the measurement model requirements and (ii) Does the instrument yield invariant interval-level measures for the intended purposes?

The attractive features of measurement in Rasch modeling – unidimensionality with linear, additive, invariant values on an interval-level measurement scale – exist only to the extent that the data fit the Rasch model requirements; guided, of course, by an understanding of how the latent trait will be captured in practice. To this should be added the recent development that allows us, under certain conditions, to transform the interval-level measurements to a bounded ratio scale [12] . No other patient reported outcome instrument,

whether generic or disease specific can meet these requirements for fundamental measurement. They are locked into a paradigm that dismisses (or is unaware of) the required axioms of fundamental measurement, relying on their belief in the primacy of data over substantive theory; the notion of quality control in the selected data does not arise. We have to use all the data regardless of quality and measurement properties.

Rasch measurement supports strong inferences that measured behaviors reflect the underlying latent trait or construct. The first step in Rasch modeling must be to measure the construct and evaluate construct validity as an umbrella concept to encompass content validity, face validity and concurrent validity. This can only occur at the disease or target patient group level if the resulting value claims are to have any merit in meeting the standards in normal science for credibility, empirical evaluation and replication. We must reject in their entirety value claims based on direct or indirect generic preference measures; claims from time trade off (TTO) and standard gamble (SG) direct measures as well as the EQ-5D-3L/5L, the HUI Mk2/3, the SF-36/6D and the AQoL among others; add to this the majority of patient reported outcome measures (PROs) which similarly fail the required measurement standards [13].

NEXT GENERATION QUALITY OF LIFE

Rejecting invented evidence also means rejecting ordinal multiattribute preference scores and the QALY. Both are well past their use by date; indeed, if they ever had one in the first place [14]. Fortunately, in respect of quality of life defined from the patient perspective, with the latent construct of need fulfillment, we have a measure that meets the required measurement standards with instruments meeting Rasch measurement standards in some 30 disease states. These are not preference measures, community valuations of multiattribute health states are not involved. Indeed, as noted, attempts to value health states are absurd. This is typically overlooked with the result we have a smorgasbord of unfit preference scores which lack dimensional homogeneity and construct validity. After some 80 years of agreement on the levels of measurement this is a ridiculous situation.

Avoiding community preferences for health states defined in terms of a bundle of symptoms and functions, does not mean that the next generation measures ignore clinical symptoms and functional status. The potential contribution of these attributes is seen through the lens of the patient (or caregiver) as elements in a broader holistic framework. As the patient (or caregiver) is the ultimate beneficiary of a therapy intervention the value claim focuses on the need of the patient and the extent to which that need is fulfilled. It is the benefit a patient derives from an intervention specific to a disease state defined in the patient's own terms.

Patient focused QoL measures are not new; they have just been ignored in favor of ordinal multiattribute preference measures. Developed over the past 25 years for specific chronic disease states there are now some 30 disease states covered (including: atopic dermatitis, psoriasis, growth hormone deficiency, Crohn's disease, depression, asthma, COPD, sickle cell disease, herpes, ulcerative colitis). These various measures are based on a coherent outcome model. They determine the extent to which respondents can meet their fundamental human needs. Items or statements are presented (with a binary True/Not true response) derived directly from relevant patients (and caregivers) and provide data on the value these groups derive from alternative interventions. This ability is clearly related to the symptoms and functional limitations they experience. However, in contrast to clinician determined HRQoL quality of life measures such as the EQ-5D, these new measures generate a basis for a single index of patient value or QoL rather than adding together (inappropriately) a basket of clinical outcomes defined as ordinal scales [15].

As disease specific measures they identify the overall impact of living with a particular disease from the patient's perspective. This provides the framework for evaluating the extent to which patient (or caregiver) need is met with competing therapy interventions. The items selected for each instrument are subject to an extended process of item selection through the application of Rasch Measurement Theory. Items finally selected are ranked in terms of the difficulty of a need being met and the ability of the respondent to meet that need expressed in probabilistic terms. The number of items selected is relatively small, typically in the range 25-30. The instrument can be completed in 4 or 5 minutes.

This single index of patient value is transformed to a bounded ratio scale that is unique to each instrument. This creates the Need-QOL (or N-QOL) measure, which is robust and accurate, meeting all the required standards detailed above. As the N-QOL is on a bounded ratio scale in the range 0 =no needs are met to 1 = all needs are met It can be used to create need-based quality of life claims by multiplying time in a disease stage by the N-QOL score to create the N-QAL. By design, negative values are impossible; scores for different instruments across disease states can be compared.

MINIMUM STANDARDS FOR VALUE CLAIMS

There are six standards that must be met for credible and evaluable value claims, including clinical endpoints, patient reported outcomes (PROs), QoL and, and resource allocation. Failing to meet any one of these standards means the value claim must be rejected. In many cases claims will have already ratio properties based on agreed clinical measurement

together with measurable (e.g., CPT code) claims for resource allocation impact; costs are not an acceptable claim. The focus of Rash measurement, as the only acceptable analytical framework is, of course, focused on latent construct of which QoL is the most relevant.

1. MEETING THE STANDARDS OF NORMAL SCIENCE

The single most important standard is to meet the requirements of normal science: *All value claims must be credible, evaluable and replicable.* If not, like the QALY, the claim is nothing more than pseudoscience and must be rejected. Invented value claims have been the mainstay of health technology assessment for 30 years; to overcome this will be difficult.

2. SUBMITTING VALUE CLAIM PROTOCOLS

Manufacturers and others submitting value claims must demonstrate how the claim can be evaluated: *All value claims must be accompanied by an evaluation protocol.* Failure to provide a claims evaluation protocol must lead to a rejection of the claim.

3. RECOGNIZING THE AXIOMS OF FUNDAMENTAL MEASUREMENT

All value claims must conform to fundamental measurement standards; this means that *the claim submitted must have ratio measurement properties* with a true zero and invariance of comparisons.

SUBMITTING SINGLE ATTRIBUTE CLAIMS

Following the standards of measurement of the physical science, all value claims should be for a single attribute whether this is for clinical, outcomes, PROs, QoL or resource utilization: *value claims must be for single attributes defined by a ratio scale meeting requirements for construct validity, content validity and unidimensionality.*

4. SUBMITTING DISEASE SPECIFIC CLAIMS

As the patient (or caregiver) is the presumed beneficiary of therapy intervention, value claims to support that intervention must be *specific to a target patient population within a disease area.*

5. REPORTING VALUE CLAIM EVALUATIONS

Value claims must, in the case of formulary submissions, be *evaluated and reported to the formulary committee* or other health system decision makers in a reasonable or meaningful time frame.

These standards are not new; indeed they are consistent with the standards of normal science. Perhaps the College of Pharmacy experts might take these on board in future attempts to undertake health technology assessment and ignore ICER.

OVERVIEW: A PARADIGM SHIFT?

The ICER analytical framework, the reference case, it a complete and unmitigated disaster. To accept the imaginary simulations as critical inputs to formulary decision making and social prices requires a major suspension of belief in simple logic, the standards of normal science and the axioms of fundamental measurement. Unfortunately, this suspension of belief or, possibly more appropriately a relativistic belief that no one system of 'truth' is superior to another and that no one source of knowledge is superior to another, is an article of faith. To a relativist, we cannot make claim to superior evidence; alternative belief systems as an analytical framework are equally valid; even those such as the ICER reference case that are clearly nonsensical. Science is not necessary to come to grips with reality; any belief system will suffice to make a decision. Evidence for ICER and its acolytes is never discovered but constructed within a social community populated by, among others, academic imaginary modeling groups. ICER rests its laurels on rhetoric, persuasion and authority. If you believe in the ICER belief system, you will enable its recommendations for the denial of care; including the unfortunate eugenic undertones [16]. This appeal to rhetoric and the fantasy land of ICER is the antithesis of what science does: to show that a consensus view must be abandoned when it is at odds with the evidence.

There will be pushback; a belief system is not overturned by logic and demonstration. The believers will continue to drink the cool aid. Abandoning a belief, a faith, based upon imaginary constructs is difficult. What Dawkins describes as a mind virus is tenacious in its hold on analysts. As a first bastion will be the plaintive defense: everyone does it. ICER is no different from others who have promoted the role of inventing evidence to support probabilistic claims. This misses the point: claims must be credible, evaluable and replicable not judged by some variant of probabilistic sensitivity analysis within a blinkered view of a relevant simulation. Perhaps, as noted in previous commentaries, belief is strongest when the object of that belief is clearly impossible: *Certum est quia impossibile est* [17]. If this is true then we can look forward to a protracted defense as the wagons are circled and the cool aid is distributed.

With ICER there is no progress, the belief system shared by many analysts in health technology assessment, is to abandon discovery. Rather than a commitment to a deeper understanding of therapy impact, of the contribution of new therapies as part of a structured research program to uncover new, yet provisional facts, ICER merely recycles information, from limited clinical trials, assumptions from the literature and 'educated' guesswork. It is a barren activity. ICER is not concerned with research, just to a business model that the credulous are prepared to accept for decision making and the denial of therapy based on an imaginary construct. More importantly, it is not just a question of abandoning progress but of actively blocking progress through imaginary claims. The denial of hypothesis creation and assessment is a barrier to new hypotheses; accepting imaginary claims to support pricing reductions and denial of care can discourage further activities in disease areas. It is not just that discovery is put to one side but of denying that discovery has any role.

We must abandon the search for a single value Holy Grail to drive formulary decisions, with acceptance or denial of care. We must base decisions on attributes specific to a disease state and established by formulary committees. Factoring in a range of attributes with required measurement properties together with input from patients themselves should be sufficient to negotiate an acceptable provisional price and conditions for access to care that can be modified over time as new data become available as part of ongoing disease area and therapeutic class reviews.

The advent of the disease specific N-QOL means the end of multiattribute ordinal preference scores and the impossible QALY. This provides an assured basis for value claims that represent he need of patients (and caregivers) and a measure of the extent to which that need is met. The key development that has made this possible is the ability, recently developed, to transform a single index of patient value from these instrument to a bounded ratio scale with all necessary properties to evaluate need and its determinants as well as robust and accurate measures of therapy response.

The next step, given the number of instruments already developed, is to initiate a research program to evaluate need in these diverse disease areas, supported by trials and observational studies to create value claims for therapy interventions. There is no longer any need to invent evidence for non-evaluable QALY claims.

REFERENCES

[1]Langley P. Suppling with the Devil: Belief and the Imaginary World of Multiple Myeloma Therapies Invented by the Institute for Clinical and Economic Review. *InovPharm*. 2021; 12(3): No 6
https://pubs.lib.umn.edu/index.php/innovations/article/view/4215/2917

[2] Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes. New York; Oxford University Press, 2015

[3] Hume D. An Enquiry Concerning Human Understanding. 1748

[4] Popper. Magee. B. New York; Fontana, 1974

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

[6] 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

[7] Neumann P, Willke R, Garrison L. A Health Economics Approach to US Value Assessment Frameworks – Introduction: An ISPOR Special Task Force Report. *Value Health*. 2018;21:119-123

[8] McKenna S, Heaney A. Composite outcome measurement in clinical research: the triumph of illusion over reality. *J Med Econ*. 2020;23(10):1196-1204

[9] McKenna S, Heaney A, Langley P. Fundamental Outcome Measurement: Selecting Patient Reported Outcome Instruments and Interpreting the Data they Produce. *InovPharm*. 2021;12(2): No. 17
https://pubs.lib.umn.edu/index.php/innovations/article/view/3911/2764

[10] Langley PC. Nonsense on Stilts –Part 1: The ICER 2020-2023 value assessment framework for constructing imaginary worlds. *InovPharm*. 2020;11(1): No. 12
https://pubs.lib.umn.edu/index.php/innovations/article/view/2444/2348

[11] Bond T, Fox C. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 3rd Ed. New York: Routledge, 2015

[12] Langley P, McKenna S. Fundamental Measurement: The Need Fulfillment Quality of Life (N-QOL). *InovPharm*. 2021;12(2):6
https://pubs.lib.umn.edu/index.php/innovations/article/view/3798/2697

[13] McKenna S, Heaney A, Langley P. Fundamental Outcome Measurement: Selecting Patient Reported Outcome Instruments and Interpreting the Data they Produce. *InovPharm*. 2021;12(2): No. 17
https://pubs.lib.umn.edu/index.php/innovations/article/view/3911/2764

[14] Langley PC and McKenna SP. Measurement, modeling and QALYs.*F1000Research.* 2020; 9: 1048 https://doi.org/10.12688/f1000research.25039.1

[15] 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

[16] Langley P. Abandoning Eugenics and the QALY. *InovPharm*. 2021;12(3): No.20

https://pubs.lib.umn.edu/index.php/innovations/article/view/4291/2939

[17] Dawkins R. A Devil's Chaplain. New York: Houghton Miflin, 2004