MAIMON WORKING PAPERS

Working Paper No. 5 March 2020

Evading the Question: The EQ-5D and ICER's Rejection of Fundamental Measurement in Fabricating Imaginary Worlds

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

Abstract
A number of commentaries have been published over the past 4 years by the present author on the manifest flaws in the reference case value assessment framework of the Institute for Economic and Clinical Reviews. The recent release of a draft evidence report on sickle cell disease gave an opportunity, as part of the public comment process, to attempt to ascertain ICER's views on the criticisms of their commitment to constructing imaginary worlds and the application of EQ-5D utilities to construct QALYs. A series of detailed questions to ICER were submitted. ICER's response indicated quite clearly that they were not interested in any critique of the merits of their reference case methodology. Their argument was quite simple: it's what everyone else does. They apparently had no concept of the role of the scientific method, hypothesis testing, or of the dubious role of health technology assessment in focusing on the fabrication of 'approximate information'. The purpose of this brief working paper is to review ICER's responses and to emphasize, once again, the absurdity of their value assessments.

Keywords: imaginary worlds, sickle cell, ICER, pseudoscience, nonsense claims, nonsense recommendations

---

Introduction

Evade: to take refuge in escape or avoidance (Merriam Webster)

Over the past few years the Institute for Clinical and Economic Review (ICER) has attempted to insinuate itself as the principal arbiter for value assessments in the US. The ICER business model is built around the construction of lifetime imaginary simulations which claim to provide a framework relevant to health system decision makers concerned with pricing and access for pharmaceutical products and devices. As detailed in previous commentaries the ICER modeling approach fails to meet the standards of normal science; the discovery of new facts proceeds [3]. It is best characterized as pseudoscience (i.e., bunk). Constructing imaginary worlds to support pricing and access recommendations has certainly characterized health technology assessment of the past 30 plus years. Indeed, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), makes clear that it is not interested in hypothesis testing or the discovery of new facts in treatment impact. ISPOR sees its principal role in generating 'approximate information'. Imaginary world evidence, created by its focus on lifetime incremental cost-per-QALY models and willingness to pay thresholds, in contrast to real world evidence where meaningful claims for therapy impact and quality of life in disease areas can be evaluated from patient-centric evidence platforms.

The purpose of this commentary is to present a rebuttal to ICER's response to questions raised under the public comment umbrella in sickle cell disease [1] . The questions relate to the ICER draft evidence report on sickle cell disease released on 23 January 2020 [2] with ICER's public comment responses released on 13 March 2020. While it is unusual for responses to responses to be pursued, in this case the responses by ICER could be interpreted as self-serving as they miss the intent of the questions raised. A clarification is in order.

Playing in the Sandbox

Sandbox: a shallow box or hollow in the ground partly filled with sand for children to play in (Oxford Dictionaries)

It is not the intent here to dissuade ICER from continuing to fabricate imaginary value assessment frameworks for products and devices. Clearly, if the past 30 years is any guide, many academics and consultants have gained gain considerable pleasure (promotions and income flows) from such constructs. After all, there is no way imaginary claims for benefits and costs over lifetimes

of patients extending 30 years or more, can be challenged, apart from proposing alternative imaginary constructs; to this extent creating imaginary client specific claims can support significant marketing activities by manufacturers. ICER, of course, can never be challenged.

ICER can continue to produce evidence reports from modeled imaginary world claims. There may even be a committed ongoing audience for these constructs. After all, if the technology assessment (or pharmacoeconomic) meme has been enthusiastically embraced for 30 years, there will no doubt be a residual audience of science fiction fans. The health technology sandbox has provided a milieu that has widespread support. All ICER has to do to support its continued use of the EQ-5D-3L to construct imaginary cost-per-QALY claims is to defend their position on the grounds that 'everyone else does it' and 'it is a recognized standard'; irrespective presumably of its well-known and documented flaws. This, as noted, in previous commentaries reflects the widespread and uncritical acceptance of the health technology approximate modeled information meme.

Fundamental Measurement

Fundamental: forming a necessary base or core; of central importance (Oxford Dictionaries)

It's a puzzle why ICER persists in the quaint notion that the EQ-5D=3L has interval properties; one answer is, of course, that if ICER acknowledged that it is an ordinal manifest scale the concept of an EQ-5D-3L QALY collapses. As ICER appears oblivious to the required properties of fundamental measurement, even if they are assumed to apply without proof in the construction of the ICER imaginary worlds, it is worth restating them. The four types of measurement scale are summarized in Table 1.

Briefly, there are four measurement scales (putting to one side conjoint simultaneous measurement which underpins Rasch measurement theory). These are nominal, ordinal, interval and ratio. The argument presented here is that the EQ-5D-3L generates ordinal manifest scores. It does not have interval properties (i.e., invariance of comparisons) and it certainly does not have ratio properties as the EQ-5D-3L 'score' lacks a true zero. The result is that to construct QALYs by assuming the EQ-5D-3L has ratio properties is a mathematical nonsense.

It would not be unreasonable to make the point that neither ICER nor the contracted university based groups of model builders appear have the first clue about the issues of fundamental measurement that may impact the 'believability' of their constructed imaginary worlds. This is all too apparent in the sickle cell evidence report where they not only continue to embrace the construction of 'for approximate information' imaginary cost-per-QALY worlds, but steadfastly take the view that the EQ-5D-3L (or other generic utility measures') have both interval and ratio properties. The absurdity of this approach has been established in this and previous commentaries. Yet they persevere.

ICER's continued embrace of this 'nonsensical' imaginary gold standard is readily appreciated. As long as ICER can continue to extol its 'virtues' to a receptive audience of believers, the ICER business model, questionable though it may be, stands firm. A bulwark against more prosaic interests in real world evidence; supporting the primacy of imaginary world evidence in formulary decisions from symptoms and ordinal response levels.

TABLE 1

A GUIDE TO MEASUREMENT SCALES

The measurement scales used in statistical analysis are nominal, ordinal, interval and ratio.

Each scale of measurement meets one or more of the following properties:

- Identity: Each value of a scale has a unique meaning (i.e., a descriptive category) but with no inherent numerical value in respect of magnitude, e.g., gender, race
- Magnitude: values on the measurement scale have an ordered relationship to each other (i.e., larger or smaller) but we don't know the distance between them; it has the properties of identity and magnitude
- Interval: scale units are equal to each other in an ordered relationship where the distances are known (i.e., by how much larger or smaller; but not how far from zero it is); it has the properties of identity, magnitude and equal intervals.
- Ratio Scale: the scale has a 'true zero' or a minimum value of zero (e.g., on a weight scale there can be no weight less than zero). A ratio scale has all four properties of identity, magnitude, interval and ratio.

Implications for statistical analysis:

- Scales with identity and magnitude properties can support median and modal claims but no other operations
- Scales with identity, magnitude and interval properties can only support the operations of addition and subtraction from any point on a real integer line (i.e., change the point relative to where it was before)
- Scales with all four properties (a true zero) can support the further operations of multiplication and division (i.e., they can change the point on the integer line relative to zero)

Unidimensionality and Latent Constructs

Unidimensionality: one dimension represented by a single number line

Measurement scales should have the property of unidimensionality. The focus should be on one attribute at a time. We must avoid confusing a number of attributes into a single score. Mutiattribute scales reduce confidence in predictions and the score is a less useful summary. In Rasch modeling, estimates of item difficulty and person ability are meaningful if every question contributes to the measurement of a single underlying attribute. Our analytical procedures if we are to meet the property of unidimensionality must incorporate indicators of the extent to which the persons and items fit our concept of an ideal unidimensional line. Items should contribute in a meaningful way to the construct/concept being investigated.

Apart from the lack of a single attribute (e.g., needs-based quality of life utilizing Rasch modeling) ICER does not appear to recognize that the responses to the five symptom levels will vary by disease state (e.g., no problem response for mobility vs. major problem in another disease state). If the EQ-5D-3L is used to create imaginary QALYs in that disease state then ICER has to demonstrate that the ratio property holds for that application. This has been ignored in all ICER value assessment models. Unfortunately, the EQ-5D-3L, applying the tariff for symptoms that are characterized ordinally as having extreme problems, can easily be shown to have a negative 'utility' Or a state worse than death (death scores a zero). The problem for ICER and others using the EQ-5D-3L is that it was not developed to meet the standards for fundamental measurement, in this case for constructing QALYs, a true zero for a ratio scale; let alone a ratio scale with both a true zero and an upper limit of unity (QALYs require time to be multiplied within a range of 0 to 1). It is not clear, if death is zero, but we measure states worse than death, how we would interpret a QALY if the EQ-5D-3L is negative (the lowest possible EQ-5D-3L tariff is -0.59)? Can we have negative QALYs when we calculate aggregate QALYs over a hypothetical lifetime? If there are only a handful of respondents who would score a negative EQ-5D-3L then we could assume that rather than experiencing a health state worse than death we could 'adjust' their score so that they can be assumed to be 'at deaths door' (but not dead as they are responding to the questionnaire); perhaps a utility of 0.000001? But this is just fudging a scale.

In the case of the EQ-5D-3L the notion of unidimensionality is absent. While it is claimed to capture health related quality of life (HRQoL), there is no single attribute or latent construct. It comprises 5 symptoms (mobility, self-care, usual activity, pain/discomfort, anxiety depression) with three ordinal response levels (no problem, some problems and major problems); creating a Mutiattribute scale with ordinal properties. Each of the symptoms is an attribute that could be the foundation for its own unidimensional scale. But

we just lump them together and attach community preference weights to the ordinal responses and assume the result is a ratio scale with a true zero even though we can have negative scores. While ICER maintains the EQ-5D-3L has interval properties they need to assume ratio properties to create QALYs (multiplying imaginary time spent in the modeled disease by the assumed ratio utility score).

The situation becomes even more bizarre when we move from the EQ-5D-3L to the EQ-5D-5L (introduced in 2009) where there are 5 response levels. Individuals in the same health state will have assigned preference scores to generate a score for five response levels quite different from 3 response levels (including states worse than death). If the ICER sickle cell model applied 5L weights then the count of QALYs (including presumably negative ones) would be different.

Even if ICER were willing to recognize the absence of fundamental measurement properties in the EQ-5D (and other generic utility instruments), this does not mean that this would give succor to the belief in fabricated imaginary evidence. The ICER value assessment framework would still fail the demarcation test as pseudoscience (i.e., bunk). It is also difficult to see how ICER might underwrite a 'utility' instrument that met the standards required. After all, instruments developed by application of Rasch Measurement Theory (RMT) focus on the response to interventions on a constructed interval scale rather than attempting to go the further step of creating instruments which have ratio properties (i.e., a true zero). The EQ-5D-3L horse has well and truly bolted.

Obscurantism

Obscurantism: the practice of deliberately preventing the facts or full details of something from becoming known: (Oxford Dictionaries)

ICER was asked to respond to a series of questions to probe their understanding of the standards of normal science and the role of fundamental measurement, with specific reference to the EQ-5D, in the construction of their imaginary reference case world for sickle cell disease. The questions with the ICER response are detailed in Table 2. Due to the limited nature of ICER responses, which in a significant number of instances evaded the question, Table 3 presents a response to set alongside the ICER response

TABLE 2

REVIEWING ICER RESPONSES IN SICKLE CELL DISEASE

Question	ICER Response	Comment
There are a large number of potentially competing models for the application of a reference case framework. Why did ICER choose this particular model framework (Section 5.1)? Outcomes (1) ICER points out (Section 5.1) that SCD 'has a large impact on patient's psychosocial wellbeing'. If this is the case, why did ICER chose to model with the EQ5D-3L system which only captures five symptoms with 3 response levels? Outcomes (2)Why did ICER utilize the EQ-5D-3L system when there are others to choose from, including the EQ-5D-5L (introduced in 2009) which is considered to be more sensitive in capturing responses? Outcomes (3) Given the number of alternative generic utility instruments why did not ICER caution readers that choice of an alternative system (e.g., EQ-5D-5L) could lead to different QALY measures?	*The rationale for the model's structure and assumptions are provided in the report text describing the model. The EQ-5D is a widely used generic QoL/utility measure, and the values used in the model were subjected to sensitivity analyses. We did not find suitable EQ-5D-5L value sets to use for this population. We have noted in the limitations section of the report that different instruments could lead to different QALY estimates.*	ICER fails to support the choice of model as one of many that leads to many options, including the need for a lifetime construct of an imaginary world. The fact that the EQ-5D is widely used is no reason for ICER's acceptance of a generic instrument that lacks fundamental measurement properties. Although ICER takes the view that it has interval scale properties, this is incorrect. It is an ordinal instrument with ordinal scales. The application of sensitivity analysis is irrelevant. If we accept constructing imaginary worlds then ICER admits that different models will yield different recommendations. There is no reason to accept the ICER imaginary worlds over a multitude of others. This is the downside of believing decision makers will accept imaginary world recommendations. How are decision makers to 'prefer' one imaginary world; one 'Harry Potter' creation over another?
In early 2019 ISPOR published a good practice for outcomes research task force report	*ICER conducts extensive literature searches for its reviews of clinical effectiveness and cost-*	This is sleight of hand. If ICER is to have credibility then it should follow accepted

(Brazier et al, Value Health 2019;22:367-75) for the identification, review and use of health state utilities for cost-effectiveness models. ICER appears to have ignored this practice recommendation. Why? Where is ICER's systematic review (pg. 70)? There is no reference to the Brazier paper.	*effectiveness model inputs, including utility values.*	research practices as detailed in the Brazier paper. ICER's response obscures a key point: if you believe in the roll of utilities then you should not rely upon one or two studies. This is what ICER does and the ISPOR standards explicitly reject this as an insufficient evidence base. ICER still fails to present its review of the literature. This is shoddy research practice.
In respect of 5 (above) would ICER consider withdrawing its evidence report until a systematic review is presented to justify it choice of the EQ-5D-3L? If not, why not?	*After review of the published utility values for this population, we selected what we believe to be the inputs best fit for the model*	Again, no review. Nor is there any justification for this selection. There is no evidence that such a review was actually undertaken. Perhaps these were the only utilities available so ICER latched on to them (port in a storm). This approach is rejected in the ISPOR proposed systematic review. More than one or two studies are required to establish this assumption. ICER gives no reason or criteria for this selection of manifest scores.
Given ICER's choice of utilities for constructing QALYs, could ICER demonstrate from one or more empirical assessments that the EQ-5D3L has interval fundamental measurement properties for the target SCD hypothetical population in its reference case modeling?	*Please see the comment regarding the EQ-5D's properties above.*	This is a complete evasion of the question. In commenting on the measurement properties of the EQ-5D, ICER makes an unwarranted assumption that those in health technology assessment, who build imaginary cost-utility imaginary worlds, think that the EQ-5D has interval measurement properties. This is quite false. We have known for 30 years that the EQ-5D is only capable of ordinal measures (i.e. where difference in scores are known (higher vs. lower) but the differences are not. Even if we assumed that the utility scores had interval properties was 'true' that would be insufficient as the creation of QALYS requires the score to have ratio properties (a true zero), which they do not as QALYS under the EQ-5D can have negative values.
In respect of 7 (above), if there is no evidence would it be reasonable to assume that ICER has simply assumed that the EQ-5D-3L for the hypothetical SCD has interval measurement properties? 1. Outcomes (8) If the EQ-5D-3L for the target SCD population has only ordinal measurement properties, how would ICER justify constructing QALYs?	*Please see the comment regarding the EQ-5D's properties above.*	Again ICER misses the point (whether deliberate or by happenstance). If ICER wishes to utilize the EQ-5D for utility scores in an imaginary cost-per-QALY imaginary sickle cell world, then the EQ-5D utilities must have ratio measurement properties for the hypothetical sickle cell population. Otherwise, we have a nonsense QALY construct. ICER has evaded the question. ICER in its evidence report cannot justify this application. ICER presents no evidence to support the proposition that the EQ-5D has other than ordinal properties.
In Table 5.16 presents utility estimates 'by assumption'. How does ICER justify this? Are they assumed to have interval measurement Where SCD-specific utility data were not available, assumptions were made that favored ©Institute for Clinical and Economic Review, 2020 properties? 1. Outcomes (10) In the report on oral semaglutide ICER introduced two separate utility systems (EQ-5D-3L and HUI Mk2) into its model? How can ICER justify this when the systems are quite different?	*Where SCD-specific utility data were not available, assumptions were made that favored the treatments. In all cases, utility values are subjected to sensitivity analyses over plausible ranges.*	This is both evasive and nonsensical. ICER confirms, again, that its reference case value assessment frameworks are nothing more than a series of assumption regarding imaginary model structure and imaginary assumptions. The issue is one of believing that assumptions held in the past will hold in the future (a nonsensical logical positivist position – Hume's induction problem). Apparently, for ICER to create its recommendations, any assumptions (or outright guesses) on utilities are acceptable. The fact that they are subject to

		sensitivity analyses is garbage. They are quite different generic ordinal systems. It seems absurd to undertake sensitivity analyses on utility 'guesses'. Why not other 'guesses'?
1. Imaginary Worlds Is the reference case imaginary lifetime model intended to generate credible, evaluable and replicable claims for cost-effectiveness? If not, why not? 2. How much credibility should be attached to the ICER model when it is only one of many that could create imaginary claims in SCD for the products assessed? What sets the ICER model apart from others?	*Descriptive and predictive models are a mainstay of economic analyses, as well as most other scientific disciplines. We use transparent models that follow standard practices and are subjected to multiple scenario and sensitivity analyses.*	Again, ICER evades the question. ICER is not interested in hypothesis testing and discovery; the notion of claims being credible, evaluable and replicable is quite foreign. Hence the conclusion that ICERs modeling is pseudoscience; sharing the stage with intelligent design. The assertion for descriptive and predictive models is a nonsense unless the model (which ICER's is designed not to emulate) produces credible and evaluable claims. The difference with normal science and in positive economics is that imaginary, fantasy constructs with no credible claims are put to one side. The models may be transparent in their assumptions, but that is irrelevant if they have no credible claims. THE ICER models, as well as those published over the last 30 years in health technology assessment, fail to reach the standards of normal science. At best they are marketing exercises, as with ICER, to support a decision that might appeal to health care decision makers. ICER needs imaginary and unchallengeable models to pursue its price reduction agenda.
1. ISPOR: Approximate Information In the 2018 ISPOR task force report on health Economics (Neumann et al, Value Health 2018;21:119-25) it is determined that economic ICER's value framework recognizes that decisions need to be made using evidence available at the time, no matter how approximate or uncertain. Our reports discuss in detail the variance and uncertainty around the available evidence for the clinical effectiveness of treatments. Our economic analyses explore uncertainty via scenario and sensitivity analyses, including probabilistic sensitivity analyses over plausible ranges of values. evaluations are intended, not to test hypotheses, but to inform decision makers of the approximate value of interventions in terms of incremental cost-per-QALYs gained. Does ICER subscribe to this view? 2. ICER: Approximate Information In respect of 14 (above) how would ICER define 'approximate value'? 1. ICER Approximate Information How would ICER differentiate 'approximate information' from 'approximate disinformation'? 2. ICER Approximate Information Where different utilities and model structures are presented in SCD lifetime modeled claims, how would ICER propose that their modeled 'approximate information' is more 'approximate' than other modeled	*ICER's value framework recognizes that decisions need to be made using evidence available at the time, no matter how approximate or uncertain. Our reports discuss in detail the variance and uncertainty around the available evidence for the clinical effectiveness of treatments. Our economic analyses explore uncertainty via scenario and sensitivity analyses, including probabilistic sensitivity analyses over plausible ranges of values.*	It is difficult to see what the term 'approximate information' as opposed to 'approximate disinformation' actually means. In health technology assessment the ISPOR (and ICER) model is to retreat from the normal science of hypothesis resting and discovery, into a fuzzy landscape of approximate (i.e. fuzzy imaginary) information; although fuzzy approximate information makes no sense. If we consider the Latin term "proximus notitia" then we could consider a report that "the barbarians were possibly north or south of the Alps" as Approximate information. A transmission of approximate information that ICER would no doubt embrace an exemplar of its unchallengeable imaginary world. A more adult approach would be for ICER to put approximate (dis)information on modeled imaginary assumptions (limited evidence base) to one side and consider a research program to discover new facts; rejecting intelligent design for natural selection. If we make fuzzy (non-evaluable) projections it is difficult to see how sensitivity analyses do anything more than increase 'fuzziness' and confuse decision makers.

claims for 'approximate information'?		
Lifetime Model: The ICER SCD model takes a lifetime perspective? How, therefore, are we to interpret the 'assumptions' driving this lifetime construct? Are they a 'realistic guess' or what?	*The report clearly states the assumptions used for the lifetime horizon model, along with the rationale*	Again ICER evades the question. All we can say is that ICER has a vision of the future, stretching up to 30 plus years, possibly including the results of major horse races, that through their crystal ball assures us of the realism and reasonableness of the assumptions and guesses that drive the ICER vision of the future. ICER judges which assumptions are relevant to the imaginary claims (supported by model groups at various universities in the US and Canada). The concern must be that, in their construction of ICER modeled worlds, their misunderstanding of measurement standards and the limitations of the EQ-5D makes their contribution problematic.
Model Assumptions: ICER's models look to a future imaginary world for SCD intervention that extend decades into the future for a hypothetical SCD population. This analysis relies on assumptions that have held in the past (from the literature), and are presumably assumed to hold into the future. Given Hume's induction problem, how does ICER justify a model built on assumptions?	*As pointed out above, descriptive and predictive models are used throughout modern scientific analysis*	Again evading (or not understanding) the question, ICER is obviously somewhat limited (if not completely foreign) in their understanding of the debates in the philosophy of science. Hume is quite clear, he is long departed by the way, building a model on assumptions drawn from past observations is nonsense. Yet ICER persists (or is quite ignorant of the question). This issue was resolved some 80 years ago by Popper in his emphasis on discovery; conjecture and refutation. The poverty of the ICER value assessment framework is all too evident.
Rasch Measurement It has been recognized since the 1960s (and in in health technology assessment since the 1990s) that if we are to capture the patient voice in therapy assessments, we require a needs based QoL instrument to capture therapy impacts with interval measurement properties. Why has ICER continued to apply generic measures of HRQoL defended by what many see as a bogus population perspective argument? Could ICER provide their case for non-patient centric HRQoL measures?	*Please see the comment regarding the EQ-5D's properties above [This has now become a standard response to working around the question].*	Again evading the question. ICER has no clue about Rasch Measurement Theory (RMT). This is understandable; one doesn't want to be faced with a proposal for disease specific quality of life measures that destroy the ICER generic utility measurement business model. The fact that RMT has been used for 60 years and in health technology assessment over the last 20 years is of no interest (or ICER is oblivious to this development). Evading the question: ICER is wedded to the EQ-5D, irrespective of its lack of fundamental measurement properties and its assumed role in creating mathematically nonsensical QALYs. This is ICERs weak point – apart from the overriding need, or compulsion, to build imaginary worlds.
Hypothetical Population What is the case for modeling to create non-evaluable claims based on a hypothetical SCD population? Why was this particular population selected for creating the ICER imaginary lifetime SCD world? Does the EQ-5D have unidimensional interval scale properties for the target SCD population?	*The rationale for the population used in the model is described in the report. Please also see the comments regarding modeling and the EQ5D's properties above.*	Again, the question is evaded. There are presumably other models and hypothetical populations which will come to different imaginary conclusions. The question of the unidimensionality of the EQ-5D is unanswered: the reason the EQ-5D does not have unidimensional properties is because it was not developed to capture a single latent construct. It is just a mishmash of symptoms and ordinal responses; each of which might

		support a latent construct. It is not clear if ICER even understands the notion of unidimensionality.
Multiplicative Assumption In respect of 9 (above), if ICER cannot demonstrate that the EQ-5D-3L has interval properties, rather than ordinal, how does ICER justify the statement that HRQoL is multiplicative (Table 5.4)?	*Please see the comment regarding the EQ-5D's properties above [Note ICER never gives details on the EQ-5D properties in terms of measurement scales].*	Whether this is evading the question or just ignorance of measurement scales, the reader must judge. The point is that if you wish to apply a utility score to construct utilities for time spent in a disease stare, then you require a utility measure that has ratio (not interval) properties. This has been recognized for the past 300 years. So far, ICER has defended its imaginary constructs in terms of an assumed 'interval' measurement property. This is fine for addition and subtraction from a 'score' but cannot be assumed. As ICER refuses to demonstrate for the target sickle cell population that the EQ-5D has interval, let alone ratio, properties (an impossible task for a single latent construct unless microdata are available) modeled claims must be rejected. Unless ICER can demonstrate that the EQ-5D, in any version, has ratio properties then the ICER sickle cell model collapses. Welcome to the real world of normal science.
1. Assumptions. Why should we' believe' ICER's assumptions as opposed to the assumptions underpinning competitor models? 2. Assumptions Why should treatment effects from clinical trials be transferable to a reference case imaginary world	*The assumptions used in our models are clearly stated, along with the rationale for each, and are tested in scenario and sensitivity analyses. Please also see the comment regarding analytic modeling above.*	Again, ICER evades the questions. There are a multiverse of competing models; each would, given their assumptions, generate alterative cost per imaginary incremental QALYs and equally imaginary recommendations for price discounting and access. If we consider effectiveness as opposed to efficacy, external validity, the absence at this early stage in product launch would argue against efficacy modeled claims. Why not wait? Instead of imaginary and easily demolished imaginary, sub-contracted models to university groups, substituted for real world evidence. Surely we can move from ICER's imaginary evidence to real world evidence. Or why do we have to rush to judgement on flimsy evidence and unsupportable assumptions?
1. Interval Scoring In respect of the utility estimates in Table 5.16. Can ICER show that each of these has interval scoring properties (including the NICE report)? This applies also to the 'calculated' utilities.	*Please see the comment regarding the EQ-5Ds properties above.*	Again, ICER evades the question. ICER nowhere demonstrates that the EQ-5D or other instrument has interval scoring properties or that they have ratio properties. Explain negative EQ-5D scores? Is a tariff difference between 0.4 and 0.5 equal to the tariff difference between 0.6 and 0.7? This has to be demonstrated for the sickle cell population. How can the EQ-5D have ratio properties (by assumption) when the EQ-5D tariff can create negative utilities. There is no true zero.
1. Validation Why has ICER ignored empirical validation (pg. 75) of claims? 2. Validation In respect of 14 (above). Why has ICER	*The report includes comparisons suggesting that predicted prevalence in the model is very similar to that of the Medicare population in*	This is evasion on a grand scale. The reason is that the modeled claims fail the test of credibility. They cannot be evaluated

sidestepped the question of empirical validation of the cost-effectiveness claims made in their reference case model? 3. Validation Are any of ICERs claims in SCD capable of empirical assessment?	*terms of chronic disease prevalence.*	empirically or replicated. Comparisons with selected observations are not evaluation. We require formal hypotheses with instructions on how these are to be tested empirically. ICER sidesteps the question of empirical assessment because the model was designed to do exactly the opposite; to fudge the question, to ask people to believe in imaginary/fantasy constructs None of ICER SCD claims merit the attention of anyone who subscribes to normal science. There may, perhaps, be a receptive audience in the one of many science fiction conventions in the US. ICER may even consider an imaginary outcomes booth.
1. Base Case Results Would ICER agree that if it cannot be demonstrated that the utilities for the hypothetical SCD population fail to demonstrate interval properties then the QALY base case results for the SCD products (Table 5.22) are meaningless? [i.e., lifetime QALYs are a mathematically meaningless fabrication]	*Please see the comment regarding the EQ-5D's properties above.*	Again, the standard evasion terminology. Obviously ICER would not agree (the reason for me asking the question in the first place) to putting aside an imaginary construct and the opportunity to strut the stage of WAC price discounts. As ICER cannot demonstrate interval properties or, in the absence of a true EQ-5D zero, a basis for creating imaginary QALYs, the answer is that the model is meaningless.
1. Thresholds Would ICER agree that if it cannot demonstrate the integrity of lifetime cost-per-QALY claims (meeting fundamental measurement standards) then recommendations for pricing and affordability based on such claims are questionable?	*Cost-effectiveness analysis using QALYs are considered the gold standard in the field of health economics. Again, please see the comment regarding the EQ-5D's properties above.*	Again, evading by hiding behind "…. Yet everyone else does it". If imitation is the sincerest form of flattery then ICER has embraced not a gold standard but fools' gold. Presumably, again to make a point, ICER is rejecting any thought of modeling and hypothesis testing that meets the standards of normal science. The ISPOR meme is an analytical dead end. Claims are not tested by assumptions. As a professional economist, I can assure ICER that my colleagues and I would reject out of hand the construction of imaginary worlds (think: positive economics)
Media Releases: Many companies have concerns that the media release ICER's recommendations for price discounting and affordability without detailing the model and its assumptions. Would ICER consider adding the following caution when its reports are released? *ICER wishes to emphasize that any conclusions and recommendations made in this report are specific to the model structure and its assumptions devised by ICER. Alternative models could be developed with their own structure and assumptions and come to possibly quite different conclusions and recommendations*	*ICER's reports and other communications always strive to include appropriate details of, and to point out any limitation, of its assessments.*	A very bureaucratic and ingenuous response. ICER may strive (presumably to persevere) but the fact remains that the ICER press releases (few actually read the report or have an understanding of the limitations of the imaginary world) give few details that would lead health system representatives to question the recommendations. It is worth noting that in the public comments presented in addition to this one, no one recognizes the imaginary nature of the ICER model and its failure to meet the standards of normal science. Hopefully this will change.
Health Related Quality of Life Construct ICER makes clear (to a certain extent) the difference between health related quality of life (HRQoL) and the broader concept of quality of life (QoL). In adopting the EQ-5D-3L as the utility measures, could ICER make clear as to what is the latent construct (if any) or other construct(s) that ICER is unable to provide this	*As pointed out above, the EQ-5D has a broad history of use in health technology assessments, as well as an extensive literature on its measurement properties.*	The final evasion or, to be more charitable, ICER probably did not understand the question. These issues are central to any coherent and intellectually responsible attempt to develop disease specific instruments that capture non-physical latent constructs, support unidimensionality and meet fundamental measurement standards.

assurance, it should be made quite clear.		The EQ-5D and ICER's use of it in modeling fail on all counts. There is no extensive literature on EQ-5D measurement properties that ask the question: does it meet fundamental measurement standards? The literature, by those who are versed in measurement theory, is uniformly negative pointing to the failure to meet interval and ratio properties.

ICER has presented a less than convincing case for the value assessment framework. Building the requisite value assessment model puts to one side the scientific method, logic, measurement theory and even the level of credulity required to believe the ICER model framework by health system decision makers. We are faced with a modeling charade; the only defense being that this seems to be the standard in health technology assessment. If this is the case we face a bleak future: formulary pricing and access decisions driven by fantasy models. If the ICER model approach is the 'gold standard' in health technology assessment, then we are dealing with fools' gold:

Imaginary Worlds and Imaginary Pricing Recommendation

Imaginary: existing only in the imagination (Oxford Dictionaries)

As well as releasing the response to public comments, ICER also released its recommendations for price discounting in its final evidence report [3]. These, of course, should not be taken seriously as they are the result of an imaginary world driven by assumptions (including a failure to understand fundamental measurement). The cost-per-QALY claims are, as detailed in previous commentaries, nonsense. The ICER recommended discounts from wholesale acquisition costs (WAC) are, as expected, substantial: crizanlizumab 70% to 74% (Novartis AG); voxelotor 79% to 83% (Global Blood Therapeutics, Inc); and L-glutamine 35% to 40% (Emmaus). As an aside, given ICER's focus on pricing and access, it has always been a puzzle as to why manufacturers support ICER financially when their products, based upon an imaginary world, may experience recommendations for substantial price discounts. This is even more of a puzzle when the modelling construct lacks scientific merit; the pseudoscience of imaginary approximate information to sit alongside intelligent design in the pantheon of scientific achievement.

Conclusions

In previous commentaries ICER has been described as an unnecessary distraction. If we are to support the application of normal science in drug discovery and assessments of clinical benefit in treatment practice, then we need to abandon imaginary approximate evidence and look to real world evidence. At product launch, rather than constructing fantasy lifetime models on a limited evidence base, made flesh by the choice of assumptions (and pure guesswork), we need to consider evidence platforms (e.g., registries) to track response to therapy. If quality of life is considered an important attribute to evaluating the impact of a product, then we need to develop instruments with demonstrable measurement properties to assess response to therapy (e.g., a needs-fulfillment construct). The focus must be on the disease state. Abandoning assumptions that a scale can have any properties that we wish it to have, will also lead to abandoning the notion of QALYs and the lifetime imaginary cost-per-QALY calculus.

References

[1] ICER. Sickle Cell Disease: Public Comments https://icer-review.org/wp-content/uploads/2019/08/ICER_SCD_Public-Comments_031220.pdf

[2] Bradt P, Spackman E, Synnott P, Chapman R, Rind D M, Pearson S. Crizanlizumab, Voxelotor, and L-Glutamine for Sickle Cell Disease: Effectiveness and Value. Institute for Clinical and Economic Review, January 23, 2020. https://icer-review.org/material/sickle-cell-disease-draft-evidence-report/

[3] Bradt P, Spackman E, Synnott PG, Chapman R, Beinfeld M, Rind DM, Pearson SD. Crizanlizumab, Voxelotor, and L-Glutamine for Sickle Cell Disease: Effectiveness and Value. Institute for Clinical and Economic Review, January 23, 2020. https://icer-review.org/material/sickle-cell-disease-draft-evidence-report/