# Value Assessment in Cystic Fibrosis: ICER's Rejection of the Axioms of Fundamental Measurement

*Paul C Langley, PhD*
*College of Pharmacy, University of Minnesota*

**Abstract**
*One of the features of the ICER stakeholder involvement in the development of ICER evidence reports is the ability for public comment. Unfortunately, and this may just a miscommunication, the replies from ICER to public comments frequently miss the point or fail to provide backup for their claims. The purpose of this commentary is to review ICER's responses to public comments by the author on the just released final evidence report on cystic fibrosis. The message is quite simple: the ICER value assessment framework lacks credibility. It fails to meet the standards of normal science. This is seen in ICERs apparent ignorance or rejection of the axioms of fundamental measurement which point quite clearly to the mathematical impossibility of creating QALYs from generic multiattribute utility scores. The ICER report also fails standards by creating a model from prior assumptions; there is no logical basis for constructing a value assessment claim. Either ICER should withdraw its value claims or admit the dubious basis on which the model is built, as a duty to its readership.*

**Keywords**: fundamental measurement, ICER beliefs, impossible QALYs, rejecting value, fantasy claims

_____

## Introduction

The construction of assumption driven imaginary worlds to support incremental cost-per-QALY claims for pricing and access recommendations is the hallmark of the Institute for Clinical and Economic Review's (ICER) business model. ICER has issued two evidence reports on cystic fibrosis. The first report, a final evidence report was released in 2018; the second report, a draft evidence report, on 20 February 2020 [1] [2]. Following the draft evidence report release a commentary was published in INNOVATIONS In Pharmacy pointing to the manifest shortcomings in the ICER value assessment framework [3]. Following the release of the draft evidence report, the opportunity presented itself to try and gauge ICER;'s beliefs in respect of fundamental measurement and the construction of imaginary cost-per-incremental QALY worlds. A series of questions were presented with a response to each question posted to the ICER website.

The purpose of this commentary is to consider and respond to the replies received ICER. This is a useful exercise because it is quite clear from these responses that ICER either does not appreciate or possibly chooses not to understand the axioms of fundamental measurement. This is a critical shortcoming because it points to ICER not understanding that the utility scales that are generated by multiattribute instruments such

as the EQ-5D-3L, EQ-5D-5L and HUIMk3, are only able to generate ordinal or manifest scores. The utilities cannot be used to create QALYs because a manifest score cannot support the fundamental four arithmetic operations of addition, subtraction, multiplication and division. Absent multiplication a QALY is an impossible construct. Formulary submissions that rely on QALYs should, therefore, be rejected out of hand. It is not a question of approximate information but of 'information' that is pure fantasy [4].

The impossibility of creating QALYs was made clear in the covering letter accompanying the list of questions posed for ICER response [5]:

> *The EQ-5D has only ordinal properties, it is a manifest scale, and should not be used to construct QALYs. If your staff are unaware of measurement properties for instruments in the social sciences for non-physical attributes, I would be pleased to explain this to them. Unfortunately, apart from the lack of scientific merit in constructing lifetime imaginary models, the misapplication of the EQ-5D-3L utilities means that your reference case model collapses.*

As a first step, however, the scene needs to be set with a brief review of the axioms of fundamental measurement.

## Fundamental Measurement

In the physical sciences, the creation of instruments with the appropriate measurement properties is central to hypothesis testing and the discovery of new facts. The same standards should apply to the social sciences, hence the importance of conjoint simultaneous measurement and Rasch Measurement Theory (RMT) in instrument development [6]. For our purposes,

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

we can focus on the axioms as they relate to basic arithmetic operations.

Our starting point must be to point out that multiattribute utility scales are ordinal or manifest scores. This has been recognized for the past 30 years. This may be considered heresy, but the key to unraveling the technology assessment belief system is to make clear that multiattribute utility scales have neither interval nor ratio properties. Analysts may believe they have; they may also believe in fairies at the bottom of the garden or even the Easter Bunny. But that is irrelevant.

Four main types of measurement scale are recognized: nominal, ordinal, interval and ratio. Each satisfies one or more of the properties of: (i) identity, where each value has a unique meaning; (ii) magnitude, where each value has an ordered relationship to other values; (iii) interval, where scale units are equal to one another; and (iv) ratio, where there is a 'true zero' below which no value exists. Nominal scales are purely descriptive and have no inherent value in terms of magnitude. Ordinal scales have both identity and magnitude in an ordered relation but the unknown distances between the ranks means the scale is capable only of generating medians and modes. The interval scale has identity, magnitude and equal intervals. It supports mathematical operations of addition and subtraction. A ratio scale satisfies all properties, supporting the additional mathematical operations of multiplication and division.

The case for multiattribute utility scales failing the standards for interval, let alone ratio measurement is that they rely on preference weights attached to ordinal response levels for the symptoms captured by the instrument. The EQ-5D-3L, for example is constructed from five symptom levels each characterized by three response levels (no problem, some problems and extreme problems). These responses can be ranked but we have no idea of the difference between them. You can attach community preferences or weights to the various response levels, add these and create a single utility, but you will still have an ordinal or manifest score; a scale that fails the axiom of invariance of comparisons. Just because you set an algorithm that is supposed to generate utilities on a dead = 0 and 1 = perfect health does not mean that the 0 is a true zero or that the space between 0 and 1 has interval properties. Instruments have to be designed to meet measurement standards; not assumed to have them *ex post facto*.

The ratio scale has a true zero where the value of a variable has no value at all below zero. Because of this it can support all arithmetic operations. A zero point is an essential characteristic. To measure a ratio between any two variables is impossible in the absence of zero, the reference point for all calculations. In the absence of a zero you cannot say that George weighs twice as much as Donald (300 lbs vs 150 lbs: a ratio of 2). We need a zero point to determine the distance from zero to support multiplication and division, as well as interval properties to support addition and subtraction. These

attributes are lacking in the multiattribute utility systems as well as the majority of patient reported outcomes (PRO) instruments. This includes, for example the most frequently used instrument in cystic fibrosis, the Cystic Fibrosis Questionnaire (CFQ) [7]. It fails the standards for fundamental measurement; it is an ordinal measure. This applies to all versions of the CFQ for adults, pediatrics and caregivers, together with the various subscales. The CFQ may meet the standards for classical test theory; but fails when assessed against the required standards exemplified by Rasch measurement theory.

It is important to note that over the past 25 years considerable attention has been given to the problem of negative utilities (from both time trade off [TTO] and the EQ-5D -3L) as well as to the possible transformation from the ordinal response or ranked manifest scores of the EQ-5D to cardinal or interval measures [8]. So far, these efforts have failed to produce any concrete results. This is not surprising. We have techniques for translating ordinal to interval scores but this requires the application of Rasch Measurement Theory; the creation of ratio scales is more complex. The result, therefore, is that groups such as ICER continue to apply the EQ-5D utilities as if they were on a ratio scale (to include interval properties). Unfortunately, the audience for ICER may not share these insights. ICER may believe; on the other hand ICER may be well aware of the 'assumption', knowing it is false [9].

**Questions to ICER**
A total of 20 questions, with a covering letter detailing key references, were submitted to ICER for their consideration and response. Fourteen were selected to provide comments on the ICER response. These follow:

1. **Questions to ICER: EQ-5D Absence of Ratio Property**

Question: It appears that many people building simulated imaginary lifetime models (e.g., ICER Value Assessment Framework) believe that it is appropriate to consider the EQ-5D-3L (used in the cystic fibrosis model) as having ratio properties (i.e., a true zero). As this is incorrect, would you explain why you persist? If you are unsure of the meaning of measurement scales, a full description of their mathematical properties is included in file:

///C:/Users/Paul/Downloads/Working%20Paper%20No.%205%20March%202020.pdf

You might also refer to the Bond and Cox reference on Rasch measurement theory.

*ICER Response: We (and most health economists) have the understanding that the EQ-5D (and other multi-attribute utility instruments) do have ratio properties. The EQ-5D value sets are based on time trade-off assessments (which are interval level)*

Comment: If most health economists do, which I doubt, then they are deluding themselves. The EQ-5D cannot have ratio properties as it lacks a true zero; the EQ-5D-3L algorithm generates negative utilities (lowest is -0.59) with an artificial starting point of unity. Given the absence of ratio properties the EQ-5D utility cannot be used to create QALYs as this requires multiplication (i.e., a true zero with no negative values). Can ICER demonstrate that the time-trade-off value sets have interval properties (i.e., invariance of comparisons?). Certainly, the time trade off (TTO) creates a raw score but this can take negative values for fates worse than death (ratios of time spent); but this does not mean that scale has interval properties [10]. The TTO also has a slight problem in dividing by zero for preferred immediate death. Are you familiar with the lead-time/lag-time literature on transforming TTO scores to avoid states worse than death? The TTO does not have either ratio or interval properties. Think: relative differences rather than raw time trade off scores! In any event that is irrelevant as the EQ-5D algorithm also generates negative utilities. You might consider reading the references provided. I recommend the Bond and Cox which points to the inherent difficulties of creating a ratio scale together with the Rasch transformation of raw scores to an interval scale.

### 2.    Questions to ICER: Absence of Interval Properties

Question:  It has been recognized for almost 20 years that the EQ-5D-3L utilities are an ordinal manifest score as the basis for creating their utilities are responses on an ordinal scale for five symptoms with three response levels for each symptom. If ICER believes this is not the case, in continuing to use the EQ5D-3L, could ICER explain why they take this view? If you are unaware of this literature please consider the references below by Grimby et al, Tennant et al, McKenna et al (2 papers) [11] [12] [13] [14].

ICER Response: We (and most health economists) have the understanding that the EQ-5D (and other multi-attribute utility instruments) do have interval-level properties. The EQ-5D value sets are based on time trade-off assessments (which are interval level), with preference weights assigned to different attributes. We fail to see why this should be considered as an ordinal (ranked) scale.

Comment: Again, read the references provided. We have known for over 20 years that the EQ-5D does not have interval properties. It was not designed to have interval properties (because no one asked the question). Note that the five symptoms that characterize the EQ-5D rest on ordinal scales for symptom response (i.e., we don't know the difference between response levels: no problem, some problems, extreme problems). If you attach weights or just integers to ordinal responses you end up with an ordinal scale (e.g., attempting to add up Likert scale values across question items; see Bond and Cox Ch. 6). Even if the EQ-5D had interval properties you could not generate QALYs because an interval scale only supports addition and subtraction, not multiplication. For this you need

a ratio scale, which you do not have. You can create an interval scale from ordinal ranks; but that is not what occurs with the EQ-5D. Again read Bond and Cox (pp30-31), noting the contributions of Thurstone in the 1920s [6]. The purpose of RMT is to translate ordinal responses to an interval scale. Again read the references.

### 3.    Questions to ICER: Invariance of Comparisons

Question: If ICER rejects the notion of the EQ-5D-3L as an ordinal manifest score, could ICER demonstrate that, if we consider the interval measurement scale, that the EQ-5D-3L for the cystic fibrosis population has invariance of comparisons? Could ICER discuss this in the context of floor and ceiling effects? Is the utility difference between 0.4 and 0.45 equal to that between 0.8 and 0.85?

ICER Response: The EQ-5D multi-attribute utility function is designed so that a utility difference of 0.05 is considered equivalent regardless of the starting point.

Comment: Really! If it was designed to have an interval scale, invariance of comparisons, then it has failed miserably. What does 'consideration' mean? It may have but may not? The lack of interval scaling properties has been remarked for the last 20 years (floor effects, ceiling effects, bunching at extreme values, negative utilities). Does going from 0 to 0.05 the same as going from -0.4 to -0.45? What does this mean? Can ICER demonstrate that the EQ-5D-3L scale has this property? If it was designed to have interval scaling properties then it must have been designed to have negative interval scaling properties! How do you go from five symptom levels with ordinal properties to a utility scale with interval properties where the algorithm creates negative utilities? Again, in any event, you need a ratio scale not an interval scale to create QALYs.

### 4.    Questions to ICER: The Dead State

Question: If ICER accepts that the EQ-5D-3L has interval properties and moves to ratio properties, can ICER demonstrate that the EQ-5D-3L has a 'true zero'? How would ICER reconcile this to the fact that with the EQ-5D-3L preference algorithm the lowest utility value allowed is -0.59? Would ICER agree that this invalidates the notion of a 'true zero'?

ICER Response: ICER believes that the dead state represents a natural zero point on a scale of health-related quality of life. Negative utility values on the EQ-5D scale represent states considered worse than dead.

Comment: Clutching at straws here! ICER might believe this; ICER might also believe in fairies at the bottom of the garden. Belief is irrelevant. What does 'natural zero' mean? Is this a weight of zero on a weigh scale (a true zero) as you cannot have a negative weight? It is gratifying that ICER acknowledges the existence of negative utilities (i.e. If you admit it then the EQ-

5D-3L cannot be a ratio scale) but perhaps ICER does believe this? Are negative utility values the equivalent of negative weights? It is not a true zero; it is just an artifact of the scoring algorithm. You should possibly adjust your preference weighting to make sure that there a no negative utilities but at least one health state that yields a "0" dead state (or possibly slightly higher to show you are at death's door) Again, you should really read the references. However, if ICER truly and deeply believes the dead state to be a true zero, then so be it. But ICER should make its audience aware of its firmly held belief.

### 5. Questions to ICER: Please see above

Question: If ICER cannot demonstrate that the EQ-5D-3L has ratio properties (let alone latent measurement properties) how can ICER persevere with its value assessment framework and recommendations for pricing and affordability? If the EQ-5D-3L algorithm allows for negative utilities (which it does) then this is conclusive that there is no 'true zero' and the notion of a QALY collapses because multiplication is disallowed.

*ICER Response: We disagree. Please see the responses above.*

Comment: *Please see above* is an ICER stock response. If ICER disagrees then a stronger case should be put forward for the belief that the axioms of fundamental measurement do not apply in their lifetime cost-per-QALY imaginary worlds. It is not a question of disagreement; it is a question of the axioms of fundamental measurement formulated by Stevens in 1946 [15]. ICER needs to start from recognition of the importance of fundamental measurement and not by unsupported belief in the value, if any, of the EQ-5D or similar utilities as the basis for QALY claims. ICER has yet to demonstrate the EQ-5D-3L or 5L have interval or even ratio properties; let alone the other multiattribute generic measures. Of course, as the EQ-5D-3L yields negative utilities, that means that ICER is prepared to recognize negative QALYs. Perhaps ICER can tell its audience how negative utilities are accommodated in its value framework? Can the modeled lifetime health path accommodate patients moving between negative and positive QALYs? Are lifetime QALYs the aggregate of the negative and positive time states? What happens when a hypothetical patient group only experiences negative QALYs? Can we have cost per negative QALY?

### 6. Questions to ICER: An Abundance of Assumptions

Question: Is ICER prepared to argue that while the EQ-5D-3L fails the standards of fundamental measurement, this is immaterial in its construction of imaginary value assessment frameworks as they are only driven by assumption anyway?

*ICER Response: As stated above, we do not accept the premise of this question.*

Comment: What premise? That the ICER value assessment framework is simply a set of (one among many) assumptions? If ICER staff had reviewed the references provided, they might have appreciated the fact that that you cannot assume that what has been observed in the past can be used to support assumptions about the future. This belief fails to recognize Hume's problem of induction. How do we justify the prediction of *instances of which we have no experience resemble those of which we have had experience?* [16] Or. as Magee puts it: *The whole of our science assumes the regularity of nature – assumes the future will be like the past in all those respects in which natural laws are taken to operate – yet there is no way in which this assumption can be secured. It cannot be established by observation, since we cannot observe future events. And it cannot be secured 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* [17]. You cannot assume it will hold in the future; even if shrouded by scenarios and sensitivity tests. So why create imaginary worlds? Who will believe you? Hugo awards for science fiction?

### 7. Questions to ICER: Descriptive and Predictive Models

Question: Is the reference case imaginary lifetime model intended to generate credible, evaluable and replicable claims for cost-effectiveness? If not, why not?

*ICER Response; 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.*

Comment: Again, ICER fudges a response. The terms are not defined. Certainly there is a role for descriptive models to define a structure and possible relationships as a step to formulating hypotheses. Predictive modeling raises the more pertinent question of whether ICER believes in testing hypotheses? The ICER models are certainly not intended to meet the standards of normal science: to generate credible, evaluable and replicable claims. ICER has instead embraced the creation of predicted 30 year imaginary claims which fail to meet the predictive standards of normal science. They are pseudoscience, sharing the Dover courtroom with intelligent design [18]. If the model is a non-evaluable fantasy construct then no amount of scenario analysis and probabilistic claims will save it. Of course you can claim transparency in your choice of one model structure and set of assumptions among many other possible modeled worlds; this is your prerogative.

### 8. Questions to ICER: My fantasy model is better than your fantasy model

Question: How much credibility should be attached to the ICER model when it is only one of many that could create imaginary

claims in cystic fibrosis for the products assessed? What sets the ICER model apart from others?

*ICER Response: We produce detailed reports describing the model's structure, assumptions, and inputs so that readers may judge the credibility of the model. At the draft report stage, we also share the actual model with relevant manufacturers for feedback and critique (the manufacturer of the treatments in this review declined to participate). In addition, we compare the model to prior published models in the same therapeutic area*

Comment: So what? It still comes down to a contest 'my model is better than your model'. My model can best represent the next 30 years of cystic fibrosis treatments and responses in target populations! Why? Why bother? You can't validate your model in terms of other models which are also fantasy constructs. The argument is somewhat circular with each model building group validating their model in terms of other models. Of course, if ICER addressed the question of hypothesis testing of claims (impossible) this may give a more useful base for comparing modeled claims. Vertex, the cystic fibrosis product manufacturer has, wisely, refused to participate in this value assessment modeling exercise.

### 9.    Questions to ICER: No Evidence?

Question: In the 2018 ISPOR task force report on health economics approaches to value assessment determined that economic evaluations are intended, not to test hypotheses, but to inform decision makers of the approximate value of interventions in terms of imaginary incremental cost-per-QALYs gained [19]. Does ICER subscribe to this view? How approximate is the modeled information in cystic fibrosis?

*ICER Response: 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.*

Comment: Presumably, in the complete absence of 'information' ICER will still build a lifetime reference model and create recommendations, modeling 30 years into the future? Is there a cutoff for determining whether or not there is sufficient information to justify assumptions and create an imaginary world? Of course, we can fall back on that catch-all term 'uncertainty'; but the focus is on claims that are neither credible nor evaluable. Lack of evidence is not a problem; with an abundance of assumptions capturing the 30 year value assessment, a few more assumptions (true, sort of true, false) will hardly make any difference. How do you gauge 'plausibility' of an assumption when you are venturing 30 years into the future? Is there a criterion for 'plausibility"? What about entry

of new products in the disease area? Is it 'sort of realistic' according to your model building team? If nothing changes this is what we think will happen? The ICER claims are still safe from any presumptuous attempt to match these to observations – which are in the future anyway. The induction problem, if ICER has even recognized its import, can be quietly ignored.

### 10.    Questions to ICER: Providing approximate Information or 'the truth is out there'

Question: In respect of 12 (8) (above) how would ICER define the 'approximate value' of its cystic fibrosis modeling for incremental cost-per-QALY gains? How is this to be distinguished from 'approximate disinformation'?

*ICER Response: See response above (actually not very helpful)*

Comment: it is not clear how ICER reconciles its commitment to modeling for approximate information to its professed commitment to predictive modelling. Perhaps the predictions are no meant to be evaluated empirically (yet 30 years out); a class of imaginary predictions If so, it is clear that ICER is committed to pseudoscience with the endorsement of the ISPOR commitment to generating approximate information. But this is odd; approximate information in respect of what? An unknown 30 year 'the truth is out there' model where the reference point to define 'approximate' is non-existent; a truth constructed from impossible QALYs yielding an impossible incremental cost-per-QALY 'master' scenario? Would other 'approximate information' models have their own 'the truth is out there' unknown and unknowable reference point for their alternative universe?

### 11.    Questions to ICER: Latent Unidimensionality

Question: Could ICER detail whether or not the EQ-5D-3L, as a health related quality of life measure, has a latent unidimensional construct? If not, how are we to characterize the 'construct' (if any) that supports this instrument?

*ICER Response: As above, please see the literature on multi-attribute utility theory (again, not very helpful).*

Comment: Another fudge. It is not clear as to whether or not ICER understood the question. A central tenet of measurement theory is that only one attribute should be captured by an instrument (e.g., temperature, needs fulfillment). The multiattribute generic instruments are, frankly, a dog's breakfast of different attributes or latent constructs. This is unfortunate, because the ability to capture change is attenuated (see Bond and Cox). The solution, long recognized in the physical science and in education (to a lesser extent psychology) is to capture one attribute at a time. There is no latent construct for the EQ-5D; it is simply a collection of clinician determined symptoms (pain, mobility, depression, etc.) each of which should be a measure in its own right. The

catch-all label health related quality of life (HRQoL) is then attached. It is not clear what attributes are being captured and how we should interpret these aggregate responses as an ordinal manifest score. What is driving a change in (hypothetical) utility score? Is it relevant to that disease state? Is it relevant to patients who would be in a more defensible position to respond on their own?

### 12. Questions to ICER: The Patient Voice

Question: It has been recognized since the 1960s (and 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?

ICER Response: *The quality-of-life weights we used in calculating QALYs were derived from EQ-5D responses from CF patients in a prior published study. Appropriate data on HRQoL from the relevant clinical trials were not available. We encourage manufacturers and researchers to include disease-specific and generic measures of HRQoL/utility in future studies.*

Comment: A walk around the question. ICER appears to put generic HRQoL claims ahead of patient centric disease specific claims. Presumably this is to maintain the 'integrity' of their imaginary value assessment framework. ICER will never countenance a shift to disease specific modelling of evaluable quality of life claims which meet the interval standards of fundamental measurement because it would destroy their value assessment framework which relies on multiattribute generic cost per QALY calculations. Yet, such instruments exist and meet the standards of RMT [20]. Again, ICER avoids the question (and fails to read references)

### 13. Questions to ICER: The ICER methodology is not flawed (?).

Question: In the ICER modeled case for cystic fibrosis, there is a clear case, based on fundamental measurement, to reject the modeled cost-per-QALY claims? Given ICERs persistence with this flawed methodology, why should we take these threshold cost-per-QALY claims and pricing recommendations seriously? How does ICER defend these recommendations?

ICER Response:  *As outlined in the responses above, we disagree with the premise that the methodology is flawed.*

Comment: I think ICER's responses to the questions raised in the public comment are sufficient to respond to this assertion. After all, ICER is apparently the self-appointed arbiter of technology assessment for new products in the US. As to 'flawed', ICER could hardly admit otherwise. If it did then it would have to withdraw all previous evidence reports and recommendations for pricing a product access.

### 14. Questions to ICER: Any serious evaluable claims?

Question: Apart from the fatal measurement assumptions, ICER asks us to believe that is possible (even with the problematic EQ-5D-3L manifest score) that the claims for a range of outcome measures should be taken seriously? Is there any intent on ICER's behalf that these claims should meet the standards of normal science for credibility, evaluation and replication?

*ICER Response: As mentioned above, descriptive and predictive models such as this are a mainstay of economic analyses, as well as many other scientific disciplines.*

Comment: Another fudged response. We are probably going around in circles. Creating imaginary worlds which lack any pretense to meet the standards of normal science is ridiculous. As a professional economist with over 40 years of experience it was not until I encountered the world of health technology assessment that these imaginary constructs appeared. Yes, creating imaginary ICER-type value assessment frameworks has been a mainstay of formulary admissions in countries with single payer health systems. There is no reason we should emulate them. But imaginary worlds to support non-evaluable claims that stretch 30 years in the future are not an accepted feature of mainstream economics. What ICER can't see is that it is in an analytical dead end. ICER may talk about predictive models but has no interest in developing them. Instead it relies on a bizarre interpretation of the axioms of fundamental measurement to support impossible QALYs in imaginary worlds.

### Next Steps
Assuming there are any; after all, the view is that the ICER value assessment framework is an analytical dead-end. This is reinforced by ICER's response to questions which give little support to those who would defend ICER on its knowledge of fundamental measurement. On its own, the belief, strongly held that all generic instruments have ratio (encompassing interval) properties should be a sufficient red flag for discarding ICER models and their recommendations for price discounts and access. It was pointed out some 30 years ago that it was time to reject misinference from ordinal scales and their misapplication in clinical decision making [21].

But ICER will persevere! It will attempt to shrug off the fundamental measurement qualifications; it will continue to create generic QALYS and model claims for incremental cost per QALY thresholds that fail the standards of normal science. ICER is publishing pseudoscience. Unfortunately, if the belief is that truth is consensus and evidence is constructed not discovered, then there is a surprisingly large audience [22].

## A Duty to Inform

ICER has a duty to inform its readership that it is employing a value assessment framework that, to put it in the best possible light, admits of alternative views. The ICER media releases should at least caution the reader that the construct does not meet the standards of normal science; it fails the demarcation test. It should be seen as pseudoscience in that it lacks claims that are credible evaluable and replicable. Obviously, ICER's assumptions are largely evidence based; from prior studies reported in the literature. There is no logical basis for assuming any of the assumptions will hold into the future. The media and the ICER readership, including those manufacturers supporting ICER, should be advised that there are also a number of critical assumptions that do not stand up to scrutiny. These relate to the axioms of fundamental measure and their properties, or lack of, in utility scales. If ICER believes that the EQ-5D-3L has ratio properties, this position needs to be defended; if ICER recognizes that the EQ-5D-3L lacks ratio properties but, in defiance of the evidence, is prepared to make this assumption to defend the construction of QALYs, then this should be stated clearly..

Failure to inform its readership, including those formulary committees and other health decision makers who take ICER's claims at face value, need to be made aware of the various criticisms directed to the ICER value assessment framework. The claims are imaginary. The argument that other analysts employ the same paradigm is no defense. If ICER subscribes to the belief in providing 'approximate' imaginary information, then this should be made clear. Is it approximate yet, in some sense, relevant? ICER needs to make its position clear.

It is one approach to act on belief  (rather than logic and evidence), yet another to put to one side critical appraisals of utility scores and QALYs when the model builders are aware of the lack of interval and ratio properties in utility scores, yet apparently choose to ignore them. Perhaps ICER should gracefully withdraw, yielding the ground to normal science. As Tennant et al made clear some 16 years ago: *As long as primitive counts and raw scores are routinely mistaken for measures by our colleagues in social, educational and health research, there is no hope of their professional activities ever developing into a reliable or useful science* [11].

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

## References

[1] ICER. Modulator Treatments for Cystic Fibrosis: Effectiveness and Value. Final evidence report and meeting summary. 7 June 2018 https://icer-review.org/wp-content/uploads/2017/10/CF_Final_Evidence_Report_06082018.pdf

[2] ICER. Modulator Treatments for Cystic Fibrosis: Effectiveness and Value. Draft Evidence Report. 20 February 2020 https://icer-review.org/wp-content/uploads/2019/09/ICER_CF_Draft_Report_022020.pdf

[3] Langley PC. More Unnecessary Imaginary Worlds – Part 3: Cystic Fibrosis and the Institute for Clinical and Economic Review's Draft Evidence Report. *Inov Pharm*. 2020; 11(1):No. 15 https://pubs.lib.umn.edu/index.php/innovations/article/view/3056

[4] Langley PC. Nonsense on Stilts – Part 1: The ICER 2020-2023 Value Assessment Framework for Constructing Imaginary Worlds. Inov Pharm. 2020;11(1): No 12 https://pubs.lib.umn.edu/index.php/innovations/article/view/2402

[5] https://icer-review.org/wp-content/uploads/2019/09/ICER_CF_Public-Comments-Folio_042720.pdf

[6] Bond T, Cox C. Applying the Rasch Model: Fundamental measurement in the human sciences. New York: Routledge, 2015

[7] Quittner A, Buu A, Messer M et al. Development and validation of the Cystic Fibrosis Questionnaire in the United States. *Chest.* 2005;128(4):2347-2354

8 Craig B, Busschbach J, Salomon J. Modelling ranking, time trade off and visual analog scale values for EQ-5D health states. *Med Care*. 2009;47(6):634-41

9 Yang Z, Luyo N, Bonsel G et al. Selecting health states for EQ-5D-3L valuation studies: Statistical considerations matter. *Value Health.* 2018;21:456-61

10 Richardson J, Hawthorne G. Negative utility scores and evaluating the AQoL all worst health state. Centre for Health Program Evaluations. University of Melbourne. Working Paper 113. June 2001  https://www.aqol.com.au/papers/workingpaper113.pdf

11 Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: Time to end malpractice (Editorial) *J Rehab Med.* 2012;l44:97-8

12 Tennant A, McKenna S, Hagell P. Application of Rasch Analysis in the development and application of quality of life instruments. *Value Health*, 2004;7(1 Suppl 1):S22-26

13 McKenna S, Heaney A, Wilburn J et al. Measurement of patient reported outcomes 1: The search for the Holy Grail. *J Med Econ*. 2019;22(6):516-22

14 McKenna S, Heaney A, Wilburn J. et al. Measurement of patient-reported outcomes. 2: Are current measures failing us? *J Med Econ.* 2019;22(6):523-30

15 Stevens S. On the theory of scales of measurement. *Science*. 1946; 103:677-680

16 Stanford Encyclopedia of Philosophy Archive. Summer 2016 Ed. https://stanford.library.sydney.edu.au/archives/sum2016/entries/induction-problem/

17 Magee B. Popper. London: Fontana, 1973

18 Piglucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010)

19 Neumann P, Willke R, Garrison L. A health economics approach to US value assessment frameworks – Introduction: An ISPOR Special Task Force Report [1]. *Value Health*. 2018;22;119-123

20 McKenna S,  Wilburn J. Patient value: its nature, measurement, and role in real world evidence studies and outcomes-based reimbursement. *J Med Econ*. 2018;21)5):474-80

21 Merbitz C, Morris J, Grip J. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil*. 1989;70:308-7

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