

**MAIMON WORKING PAPER No 26 NOVEMBER 2020****WHY ICER HAS FAILED**

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

The Institute for Clinical and Economic Review's (ICER's) failure to provide credible claims for pharmaceutical product pricing and access was guaranteed from the beginning. ICER chose to create approximate information to support its claims rather than the application of the standards of normal science to meet evidence gaps and provide formulary committees with robust and testable claims for product performance <sup>1</sup>. The result was inevitable. None of the evidence reports for product pricing and access published over the past decade by ICER have any claim to credibility. They are impossible and entirely imaginary.

ICER's failure can be detailed as follows:

1. *Rejection of the Standards of Normal Science*: ICER has rejected, in its adoption of the reference case simulation framework, any attempt to create credible, empirically evaluation and replicable claims for value assessment. These are the standards of normal science to fill evidence gaps and discover new facts about products and patient response to therapy. Instead, ICER relies on evidence creation. The reference case is a lifetime modeled simulation of the assumed process of a disease in a hypothetical patient population. The model is populated by assumptions, building on limited clinical data from phase 2 and phase 3 trials. These assumptions are expected to hold for decades, with time lines for claims that stretch 30 or more years into the future. This is entirely a fantasy construct. We don't know if ICER is right, if ICER is wrong and we will never know. In fact, we were never intended to know. It is pseudoscience (cf., intelligent design vs. natural selection) <sup>2</sup>.
2. *Reliance on Assumptions*: ICER's non-evaluable claims are built on assumptions drawn from clinical trials and the wider technology assessment literature. ICER believes that assumptions that held in the past will necessarily hold for a simulation that 'represents' the future; ICER's presumed 'best 30 year guess'. This is logically absurd. ICER ignores Hume's induction problem: 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*' <sup>3</sup>. Certainly, assumptions

are important in developing models; the difference is that these models produce credible, empirically evaluable and replicable claims for target treating populations.

3. ***Axioms of Fundamental Measurement:*** ICER's reference case modeling framework ignores recognized limitations on measurement claims in favor of measurement assumptions that are absurd. In the physical sciences, and the more rigorous, and aware, social sciences such as economics, an understanding of the axioms of fundamental measures are recognized and are considered essential in instrument development <sup>4</sup>. Following the formalization by Stevens and others in the 1930s and 1940s, the axioms of fundamental measurement are well understood <sup>5</sup>. The measurement scales used in statistical analysis are nominal, ordinal, interval and ratio. Each scale of measurement has one or more of the following properties: (i) identity where each value has a unique meaning; (ii) magnitude where ordered values on the scale have an ordered relationship with each other but the distance between is unknown; (iii) invariance of comparison where scale units are equal to each other in an ordered relationship and known; and (iv) a true zero where no value on the scale can take negative scores. The implications for the ability to utilize a scale to support arithmetic operations (and parametric statistical analysis) are clear cut. A nominal scale is just a set of unique meanings but nothing else (e.g., gender). An ordinal scale has identity and magnitude in an ordered relationship but we do not know the distance between the values (i.e., it cannot support arithmetic operations, only non-parametric statistical evaluations, modes and medians). An interval scale has known differences but no true zero and can support only addition and subtraction (i.e., it can change the point on an integer line but only relative to other points). A ratio scale can support the additional operations of multiplication and division because it has a true zero (i.e., change the point on an interval line relative to zero).
4. ***Measurement Implications:*** Apart from a refusal to accept the standards of normal science, ICER's failure stems from its lack of awareness of measurement theory. This is a fatal oversight and guarantees the impossible nature of ICER's modeled claims. An understanding of measurement axioms is fundamental to recommendations for dismissing ICER modeling from formulary decisions.
5. ***Ordinal Utilities.*** Creating evidence from lifetime simulation relies on the application of utilities. These are, hypothetically, scores in a range from 1 = perfect health to death = 0 that can be applied to time spent in a hypothetical disease to 'create' perfect years of life equivalence <sup>6</sup>. To accomplish this the utility scale must have ratio properties to support basic arithmetic operations; in this case multiplication. The ratio property requires a true zero. No utility value can take negative values. Unfortunately, the EQ-5D-3L utility scale, widely used by ICER in modeling, has a range from -0.59 to 1.0. It does not have a true zero. This is common for all constructed utility scales. It is not a ratio scale and, therefore, cannot support multiplication (or addition,

subtraction and division). It is, in fact, an ordinal scale. ICER understandably argues that this is not true; there are alternative facts.

6. ***ICER's Mystical Utility Scale:*** Despite the formalization of measurement scales some 80 years ago and the addition of conjoint simultaneous measurement which set the stage for Rasch Measurement Theory in the early 1960s, ICER apparently holds strongly to the belief that there is, to support the application of ordinal utilities a 'mystical' utility ratio scale without a true zero <sup>7</sup>. A parallel universe of measurement theory known only to ICER and its academic consultants. This is bizarre. What ICER fails to recognize is that if you want to have a measurement scale to have particular properties (i.e., interval or ratio) then it has to be designed to have them. In fact, for latent (i.e., non-physical) characteristics or attributes such as quality of life it is, as far as can be judged, impossible to create a ratio scale for a latent attribute <sup>8 9</sup>. At best, given the framework of RMT, construct instruments (i.e., questionnaires) that create interval scales. These capture response to theory but cannot support QALYs.
7. ***The Impossible QALY:*** as the utility score applied to create QALYs has only ordinal measurement properties (and despite ICER's unfounded belief or understanding that it has ratio properties) it is mathematically impossible for a QALY to be created (hence the term I-QALY) <sup>10</sup>. If the I-QALY is disallowed, which it has to be, then the ICER claims for lifetime incremental cost-per-QALY imaginary simulations are a waste of time; they fail the axioms of fundamental measurement apart from, in broader terms, the standards of normal science.
8. ***ICER's Impossible Pricing Thresholds and Fair Prices:*** Central to ICER's belief in its role as arbiter for pricing and cost-effectiveness claims in the US (with accolades from the media and those academic groups who should know better) the cost-per-QALY willingness to pay threshold is also an impossible construct. If the QALY is an impossible mathematical construct, then claims based on costs-per-QALY (e.g., \$50,000, \$100, 000 or \$150,000) are just as meaningless. ICER's claim to generate socially acceptable fair prices and consequent price discounts for formulary acceptance are simply nonsense.
9. ***ICER Rejects Dimensional Homogeneity:*** Multiattribute utility scales as the name implies bring together in a single score a variety of clinical symptoms and response levels for patients to report. This violates the axioms of fundamental measurement <sup>11</sup>. It is impossible to combine different health attributes (e.g., depression, pain, mobility) into a single measure because each attribute has own dimension measured by its

characteristics. Items in any composite scale are only valid if they have the same dimension. Accurate measurement of latent constructs requires unidimensionality, reflecting a single construct or attribute. This requirement is violated with multiattribute instruments such as the EQ-5D-3L. It is dimensionally heterogeneous. It is not only an ordinal scale but an impossible ordinal scale.

10. ***ICER Rejects Single Attributes:*** Interval and ratio scales are only viable if the measure refers to a single attribute. This has been recognized in the physical sciences since the scientific revolution of the 17<sup>th</sup> century. You cannot combine different attributes in a single measure. In product assessment this means that claims must be in single attribute terms, appropriately measured as interval or ratio scales, to cover clinical response, quality of life and resource utilization. Lumping everything together in a blanket cost-per-QALY framework, a mash-up of single attributes, is unacceptable
11. ***ICER's Ignorance of Rasch Measurement Theory.*** Although developed in the early 1960s to provide a framework for measuring latent constructs (e.g., quality of life) Rasch Measurement Theory has been ignored by ICER <sup>12</sup>. This is because ICER is hooked onto an analytical framework that utilizes generic (i.e., non-disease specific) utility measures to create QALYs and modeled simulated evidence to support cost-per-QALY claims. This not only defies normal science but involves assumptions, including denial of the axioms of fundamental measurement. It totally lacks credibility. The Rasch approach is to create QoL instruments specific to target population needs in disease areas. Rasch allows raw scores from interviews to be translated to interval measures providing robust evidence for response to therapy. The Rasch approach is both consistent with the standards of normal science but also the axioms of fundamental measurement. Rasch focuses on single attributes for claims (e.g., improvement in quality of life) and dispenses entirely with QALYs and imaginary lifetime simulations. Rasch claims are credible, empirically evaluable and replicable. We have a range of Rasch measures in quality of life.
12. ***A New Framework for Value Assessment:*** We have the successor framework for true value assessment available <sup>13</sup>. We know how to maintain the standards of normal science and respect the axioms of fundamental measurement. The framework focuses on claims for single attributes proving an evidence base to support continued assessment of product performance to support ongoing disease area and therapeutic class reviews.

**13. Nullius in Verba: The motto of the Royal Society (1662): *Take no man's word for it.* In this case we would be foolish to take ICER's word for pricing or product access under any circumstances. It rejects normal science and ignores fundamental measurement in favor of creating evidence to support fanciful and non-evaluable claims.**

## REFERENCES

- 
- <sup>1</sup> 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>
- <sup>2</sup> Piglucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010)
- <sup>3</sup> Magee B. Popper. London; Fontana, 1973
- <sup>4</sup> Langley PC, McKenna SP. Measurement, modeling and QALYs [version 1; peer reviewed] *F1000Research* 2020, 9:1048 <https://doi.org/10.12688/f1000research.25039.1>
- <sup>5</sup> Stevens S. On the theory of scales of measurement. *Science*. 1946;103:677-680
- <sup>6</sup> Drummond M, Sulpher M, Claxton K al. Methods for the Economic Evaluation of Health Care Programmes. 4<sup>th</sup> Ed. New York: Oxford University Press, 2015
- <sup>7</sup> Langley P. The Impossible QALY and the Denial of Fundamental Measurement: Rejecting the University of Washington Value Assessment of Targeted Immune Modulators (TIMS) in Ulcerative Colitis for the Institute for Clinical and Economic Review (ICER). *InovPharm*.2020;11(2): No 17 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3330/2533>
- <sup>8</sup> 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
- <sup>9</sup> McKenna SP, Heaney A, Wilburn J. Measurement of patient reported outcomes. 2: Are current measures failing us? *J Med Econ*. 2019;22(6):523-30
- <sup>10</sup> Langley P. The Great I-QALY Disaster. *Inov Pharm*. 2020;11(3): No. 7 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3359/2517>
- <sup>11</sup> McKenna S, Heaney A. Composite outcome measurement in clinical research: the triumph of illusion over reality. *J Med Econ*. 2020; 23(10):1196-1204
- <sup>12</sup> Bond T, Fox C. Applying the Rasch Model 3<sup>rd</sup> Ed . New York: Routledge, 2015
- <sup>13</sup> Langley P. Value Assessment, Real World Evidence and Fundamental Measurement: Version 3.0 of the Minnesota Formulary Submission Guidelines. *InovPharm*. 2020;11(4): No 12 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3542/2613>