MAIMON WORKING PAPERS

Working Paper No. 8 March 2020

POST NUBES LUX: BEYOND REFERENCE CASE IMAGINARY WORLDS TO REAL WORLD EVIDENCE IN HEALTH TECHNOLOGY ASSESSMENT

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

Abstract
One of the unusual features of health technology assessment that has been endorsed over the past 20 to 30 years by professional groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and health departments in single payer systems such as the National Health Service in England through the National Institute for Health and Care Excellence (NICE), has been the commitment to base decisions on imaginary modeled worlds. Creating imaginary information to support formulary decisions may seem an odd commitment. A paradigm, if that is the right word, that places fabricated evidence ahead of the scientific method. A rejection of claims for competing products that meet standards for credibility, evaluation and replication in favor of assumption driven child-like paracosms that are best described as pseudoscience or, more colorfully, pure bunk. This rejection of hypothesis testing is defended on the grounds that health technology assessment is focused on providing 'approximate information'. If we accept this novel approach to decision making then we must accept, presumably, that the truth is out there. We may never know when we might approximate to this X-file ideal or whether we are providing approximate disinformation. The fabrication of imaginary worlds, the ISP0R meme, not only fails to meet the standards of normal science, but at its core suffers from a fatal flaw: the quality adjusted life year (QALY) fails to meet fundamental measurement standards; mathematically it is a nonsense creation. It fails because the utility scales which are used to 'adjust' time spent in a disease state are ordinal scales. They lack the required properties, not only of interval measurement but more importantly, ratio measurement. The QALY is a logically impossible creation. It might exist in the memetic ISPOR imagination, but must be rejected when it is used to create imaginary incremental cost per QALY claims. The purpose of this commentary is to consider what options we have to support formulary decisions if we are to subscribe to the standards of normal science; where decisions reflect a commitment to real world evidence. What should be our next steps when we progress from imaginary to real world evidence? Post nubes lux.

Keywords: deconstructing imaginary worlds, pseudoscience, ICER, ISPOR, real world evidence, paradigm shift

---

Introduction

Insinuate: to introduce by stealthy, smooth, or artful means (Merriam Webster)

Commentaries by the present author over the past few years, as detailed on the author's website (www.maimonresearch.net), have pointed to the lack of scientific merit in the construction of imaginary, modeled-by-assumption, reference case worlds to support health technology assessment. Outside of professional groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 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 cost-per-QALY lifetime imaginary simulations which claim to provide a framework relevant to health system decision makers for pricing and access for pharmaceutical products and devices. As detailed in a number of commentaries the ICER modeling approach, its value assessment framework, fails to meet the standards of normal science; the discovery of new facts [1]. It is best characterized as pseudoscience (i.e., bunk) [2]. Constructing imaginary worlds to support pricing and access recommendations has certainly characterized health technology assessment of the past 30 plus years. Indeed, ISPOR makes clear that it is not interested in hypothesis testing or the discovery of new facts in treatment impact [3]. ISPOR sees its principal role in generating 'approximate information'; imaginary world evidence, created by its focus on lifetime incremental cost-per-quality adjusted life year (QALY) estimates and willingness to pay thresholds. Unfortunately, the QALY is a logically impossible creation. This absurd focus on imaginary evidence is in contrast to real world evidence where meaningful claims for therapy impact and quality of life in disease areas can be evaluated from evidence platforms.

Of particular interest is the willingness of decision makers to accept claims based on the construction of imaginary worlds. Whether this reflects their lack of understanding of the scientific method or an easy way out in formulary decision making, is a moot point. The fact is that in media presentations, the claims by ICER are taken at face value; products at the market entry price proposed by manufacturers are labeled 'not cost-effective'. It is doubtful if the media commentator understands what the term means; a position shared by a number of health economists. Nevertheless, the ICER recommendations for price discounts and access are treated as gospel truth. This has to be challenged.

The purpose of this commentary is to ask decision makers to abandon imaginary constructs; to put imaginary information, the mainstay of the ICER business model, to one side in favor of a program to support formulary claims driven by real word evidence. But first, we must demolish the ISPOR/ICER reference case framework. This is quite straightforward. But second, and more problematic is to consider how, given our paradigmatic shift, we can most effectively meet the standards for formulary submission that are driven by real world evidence.

Deconstructing the ICER Value Assessment Framework

Deconstruct: analyze (a text or a linguistic or conceptual system) by deconstruction, typically in order to expose its hidden internal assumptions and contradictions and subvert its apparent significance or unity (Oxford Dictionaries).

It is surprisingly easy to demolish the ICER value assessment framework. This rests on two avenues of attack: (i) to demonstrate that the model driving the value assessment framework lacks scientific credibility and (ii) to demonstrate within the model framework the attempt to create imaginary QALYS is a fool's errand. Of course, such an assault (to extend the military metaphor) applies in equal weight to the ISPOR commitment to imaginary worlds as the gold standard fantasy construct. As note in previous commentaries, ISPOR survives through the acceptance of its health technology assessment meme; a unit of cultural transmission with such fidelity that successive waves of newly minted PhDs will continue to subscribe to this paradigm[4] . This is seen, for example, with ICER contracting to university groups in Canada and the US to construct imaginary value assessment worlds. Typically, the recommendations for price discounting are substantial, but all rest on an assumption driven cost-per-QALY fabrication which fails to meet the standards of normal science.

Deconstruction: Abandoning Imaginary Reference Case Models

If a claim is to be made for cost-effectiveness, then there needs to be some agreement on the framework for assessment. The health technology solution is to construct imaginary reference case value assessment frameworks that extend decades into the future as the most appropriate framework. This has a positive spin: claims made (even if discounted) can never be evaluated empirically; the claims can never be disproved and were never intended to meet evaluation standards. ICER can never be wrong. Certainly, different imaginary scenarios, supported by the inevitable tornado diagram and probabilistic sensitivity analysis, can be presented. This gets us nowhere as the model and its scenarios is only one of a potential multiverse of models for products in a disease area; claims in many cases driven by marketing considerations. Claims lack credibility, are neither evaluable and nor replicable. They are opposed to any notion of the scientific method and the discovery of new facts. This of course, unlike in the physical and other social sciences, built on the last 400 years of 'conjecture and refutation' that has characterized scientific advancement since the 17th century, puts such endeavors to one side [5]. But we follow, slavishly, the direction of the 'leaders' in health technology assessment.

Deconstruction: Rejecting Assumption Based Imaginary Worlds

The ICER/ISPOR value assessment frameworks rest on assumptions and, more to the point a belief that assumptions drawn from the literature can drive claims for product impact. This is nonsense. As noted in previous commentaries, ISPOR/ICER face Hume's problem, of induction. Any number pf prior observations cannot justify the assumption that the prior observations can drive future assumptions. ICER and others 'justify' their construction of imaginary worlds on their choice of assumptions. This is unacceptable. An assumption may be judged 'realistic' and applied to a future scenario, but simple logic dictates that, in recognizing Hume's problem, that *the fact that all past futures have resembled past pasts, it does not follow that all future futures will resemble future pasts* [6] . We cannot ask clients in health care to believe that prior assumptions will hold in the future. It is logically indefendsible.

Deconstruction: Disregarding 'for approximate information' claims

Apparently, the objective of health technology assessment is to create assumption driven approximate information; not hypothesis testing. This term is never defined. It could, presumably, be considered 'approximate disinformation'. If the truth is out there and the approximate information is intended to drive us towards this nirvana, then in the absence of any notion of 'the truth' it is difficult to say whether or not we are at the Greyhound depot or an unknown number of stops from our destination driving in the opposite direction. It is a meaningless phrase. The defense that the assumptions that drive the 'approximate information' are 'believably realistic' is a non-starter. Assumptions have relevance in a model with credible and evaluable claims; with the possibility that the claims will fail. This is the essence of the scientific method. This can lead to a reassessment; a valid basis for reassessment. Imaginary assumption driven worlds deny us this feedback capacity. This is even more of a mystery when key imaginary assumptions, the utility constructed QALY, clearly fail the standards of fundamental measurement. It is not a debate over which assumption is more 'believably realistic' but that building imaginary models on a series of assumptions is just absurd.

Deconstruction: Reject Multiattribute Generic Manifest Scores

The term unidimensionality, a latent construct that drives a measure of response, is absent from the ICER/ISPOR lexicon. Their focus on generic utility measures such as the EQ-5D points to their acceptance of multiattribute scales (or instruments) to support claims from imaginary worlds. Unlike standards in normal science, where the attribute of interest is unidimensional (e.g., temperature), ISPOR/ICER are satisfied with a multiattribute instrument that lacks fundamental measurement properties. Once we admit multiattribute instruments then we abandon fundamental measurement; we abandon the focus in the sciences on measurement tools that meet required axioms of invariance of comparisons and sufficiency. If an instrument is to have the required properties then this has to be clearly stated and accommodated in the process of development.

Deconstruction: Recognize the Standards of Fundamental Measurement

It is critical to emphasize and re-emphasize the lack of appreciation of the standards of fundamental measurement in health technology assessment modeling. [7]. Previous commentaries have defined this standard. Recognition of standards for ordinal, interval and ratio measurement are integral to instrument development. This is the basis of Rasch measurement theory where instruments (e.g., a needs-fulfillment quality of life questionnaire) are designed in their selection and ordering of items to meet fundamental measurement standards. This yields interval measures that give an index of response to therapy [8 9]. Unfortunately, apart from the objections to the notion of constructing non-evaluable reference case value assessments, a lack of appreciation of such standards in instrument development, relegates the EQ-5D-3L, EQ-5D-5L and other generic or patient reported outcome (PRO) instruments to the category of ordinal scales. Scales where the only possible measures are for median and modal values. It is mathematically illogical to multiply time spent in some modeled disease stare by an ordinal metric on a presumed 0 to 1 utility scale to create a QALY. Of course, ISPOR/ICER could assume this is possible; that would add further doubts to the merits of the exercise in the first place. We can only go so far in denying the relevance of the scientific method to support claims.

Deconstruction: Reject QALYs as an Analytical Impossibility

It is a puzzle as to why QALYs were proposed as an outcome measure (or <u>the</u> outcome measure) in health technology assessment? The answer is simple: no one thought about the fundamental measurement properties required of generic utility measures and even so-called disease specific measures. If time spent in a disease stage is to be translated to QALYs, then the utility measure must exhibit two properties: first it must be on a ratio scale (with a true zero) to support multiplication and division (not interval properties which allow only addition and subtraction) on a number line and, second, it must have a cap of unity. Apart from the demonstrable fact that the EQ-5D-3L fails to meet the true zero requirement (the algorithms allow negative utilities: lowest 'score' is -0.59) with an artificial cap of unity (as the utility score is created by subtracting from unity). The utility fails to meet the axioms of fundamental measurement. The utility is an ordinal manifest score [10]. It was never designed to have fundamental measurement properties. The QALY, despite its central place in modeling for 30 years, is a mathematically impossible construct.

The Tenacity of Memes

Dog's Breakfast: a complete mess; so fouled up as to be utterly useless (1930s UK and Commonwealth slang)

Irrespective of the case for abandoning incremental cost-per-QALY reference case imaginary worlds, there seems little doubt that organizations such as ISPOR, not to mention ICER will dig their heels in to argue for the 'scientific merits' of their paradigm. After all, as made clear in a recent ISPOR report: *Leaders in the field of economic evaluation in health care have long recommended that analysts seeking to inform resource allocation decisions approximate the value of interventions in terms of incremental cost-per-QALY gained* [emphases added] [3]. This is a difficult position for the paradigmatic leaders to step back from; to acknowledge that for 30 years and thousands of cost-per-QALY modeled claims, all apparently to inform decisions, all fail the demarcation test [11]. As noted here and in previous commentaries and reports we a truly dealing with a dog's breakfast of 'real' and 'imagined' assumptions to drive recommendations by groups such as ISPOR, that may have untoward and unjustified implications for price discounting and access. Unsupported recommendations that may cost manufacturers, notably those with orphan drug populations, millions of dollars but send entirely the wrong signal to potential investors. Whether the meme leaders recognize, or are even prepared to reconsider their support for QALY claims, is a moot point. Perhaps we can look forward to an extended debate over whether or not EQ-5D utilities meet ratio scale standards and the importance on continuing to model from a lifetime perspective.

From Imaginary to Real World Evidence

Paradigm Shift: a fundamental change in approach or underlying assumption

Introduced by Thomas Kuhn (1922-1996) in the early 1960s, a paradigm shift is defined as a fundamental change in the basic concepts and experimental practices of a scientific discipline [12]. To the extent that health technology assessment, the commitment to the construction of imaginary worlds and a rejection of hypothesis testing would be considered, at least in its present incarnation as a 'scientific' discipline, the refocus from imaginary to real world evidence may be considered to merit this label. Deconstructing the reference case 'paradigm' has illustrated its critical weaknesses; an analytical 'for approximate information' framework that fails the demarcation test between science and pseudoscience (or bunk).

If we are to accept, possibly grudgingly, a new paradigm, then it must be made clear that the focus of the past 30 years on imaginary incremental cost-per-QALY imaginary worlds has been an analytical dead end. While this may have a seismic effect on the ISPOR health technology assessment meme, the fact is that the approach fails the standards of normal science; not only in the lack of credible, evaluable and replicable claims (it was never a standard that it aspired to) but, critically, a reliance on the notion of a 'gold standard' QALY driven by utilities but with only ordinal, manifest scale properties. This rejection holds irrespective of criticisms that have been directed at QALYs and attempts to argue that, after all, cost-per-QALY is an element, possibly in a multiple criteria decision or similar value framework. Perhaps we should concur that we can reject fundamental measurement and insist, against all evidence, that the EQ-5D utility has ratio properties. Attempting to bring QALYs in through the back door won't work. Even if it were possible to create a generic 'utility' that had unidimensional, interval properties it would still be impossible to create a QALY as this requires a ratio scale (a ratio scale is virtually impossible). Even then, there would still be the fundamental objection to the pseudoscientific status of imaginary reference case lifetime modeled claims.

Cost-Effectiveness

Care has to be taken in applying the term cost=effectiveness to competing product claims. The key point to note is that any cost-effectiveness claim must be credible, evaluable and replicable. Irrespective of the outcomes proposed, which in ICER's case would include not only the so-called primary measure of cost effectiveness the cost per-QALY life year gained, but subsidiary imaginary measures of cost per life year gained, cost per equal value life year gained and other costs per unit of health benefit such as stokes prevented. None, in the context of the ICER value framework, are empirically evaluable.

This is not to reject cost-effectiveness models out of hand. After all, if a manufacturer wishes to make short-term claims for some cost-outcomes measure, supported by a claims assessment protocol, then this will yield evaluable claims that, in principle, could be reported to a formulary committee. This may, for example, be a modeled extrapolation from a clinical trial where the costs and outcomes are appropriately powered, capturing both treatment effectiveness and outcomes that meet standards for fundamental measurement. Other claims, from an observational perspective or captured by administrative claims data could be for improved compliance with therapy, reduced emergency room visits and hospitalizations and other elements of direct medical costs. Access to micro-data from a registry could support multivariate analyses to assess the independent contribution of therapy switching as well as quality of life claims if these were captured from an evidence platform and where the quality of life instrument meets fundamental measurement standards.

The issue is, of course, time. We can either build imaginary fabricated worlds that fail the standards for normal science in order to fill a perceived evidence gap with imaginary non-evaluable 'approximate' claims for the more gullible members of the audience to influence product entry pricing and access decisions, or we can treat pricing and access as provisional and put in place a long-term claims evaluation platform. The platform can be designed to produce claims reported for different timescales, supporting ongoing disease area and therapeutic reviews, and possible reassessments by formulary committees. This is not rocket science; elements of tracking and reporting platforms have been applied over the years to support, for example, risk sharing agreements. The difference here is that the evidence platform is integral to the assessment of product claims, including clinical effectiveness, that address patient needs and the extent to which competing therapies meet those needs.

Perspectives in Real World Evidence

Real World Evidence: Clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of real world data. Real world evidence can be generated by different study designs or analyses, including but not limited to, randomized trials, including large simple trials, pragmatic trials, and observational studies (prospective and/or retrospective). (FDA:. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence)

To support a paradigm shift two complementary approaches are recommended. First, from the manufacturer's perspective, a 'global' product dossier and, second, from a health system perspective, formulary submission guidelines that meet the standards of normal science.

A Global Dossier

The purpose of a global dossier is to underwrite clinical and other comparative claims for pharmaceutical products and devices. Integral to the dossier is an evidence base maintained by the manufacturer. This establishes a real world evidence framework for ongoing clinical trials as a long-term research strategy that may involve expanding or refining the definition of, and access to, the target population (e.g., genomic markers), identifying evidence gaps, expanding product indication and linking product claims to external complementary data sets. It is, if you like, a one-stop-shop for product support. The evidence base, with a registry as a classic element, supports claims assessment with feedback to formulary committees. It can support quality of life claims with instruments introduced to the target patient registry that meet fundamental measurement standards.

The global dossier supports submission to formulary committees. Depending on the data elements required, these can be assembled from the dossier (e.g., characteristics of the target population; summary of phase 3 clinical trials, indirect comparisons of clinical trial-based efficacy) together with the case made for selected non-clinical outcomes (e.g., quality of life). As a common resource the dossier can also include data elements from external databases (e.g., administrative claims) together with statements by the company only to support claims that are credible, evaluable and replicable. This is a potentially key element given the number of single payer health systems that continue to subscribe to (and require) the construction of imaginary incremental cost-per-QALY worlds to support pricing and access. Manufacturers must be in a position to challenge the role of imaginary reference case worlds.

The dossier is, of course, a work in progress. It not only supports product entry to the various health systems but provides ongoing reports of clinical trials and claims assessment. It can support a publication program as well as being a resource for marketing and sales.

Formulary Guidelines

An obvious first step, if we are to move to a superior real world evidence paradigm, is to consider the extent to which this is bought into by formulary committees for the various health system and insurers. In the case of ICER the reference case sets the value assessment framework with contracting university groups following the assessment requirements. Apart from this, the formulary submission guidelines developed by the Academy of Managed Care Pharmacy have received widespread acceptance. Again, these follow the technology assessment paradigm of building imaginary worlds [13]. Outside of the US, the majority of single payer health systems have followed the lead of NICE in developing their own versions of guidelines. With few exceptions (e.g., Germany) all give central place to claims expressed in modeled incremental cost-per-QALY terms with value thresholds to support pricing and access

decisions. None have addressed the issue of fundamental measurement or the lack of scientific merit in the fabrication of imaginary worlds.

This continued adherence to the incremental cost-per-QALY health evaluation paradigm has been, albeit unsuccessfully, challenged by guidelines proposed by the present author where the emphasis has been, not on imaginary constructs, but on requiring any claims for competing products to be empirically evaluable [14]. This was initially proposed in 2005 for the WellPoint (now Anthem) health system, where manufacturers were asked as part of the formulary decision, and in future requests for re-assessments under disease area and therapeutic reviews, to propose how their claims were to be evaluated and reported to the formulary committee in a meaningful time frame. More recently, this same standard was proposed in successive versions of the proposed Minnesota formulary guidelines and posted to the website of the College of Pharmacy, University of Minnesota [15]. These are currently under review for a third version.

If formulary committees are to re-focus on real word evidence for the assessment of competing product claims, then guidance should be given to support revised formulary submissions. While it is not the purpose here to propose a detailed template for re-drafting formulary submission guidelines, there are key evidence issues that should be addressed. It goes without saying that imaginary pseudoscientific reference models that rest their case on incremental cost-per-QALY frameworks that fail the standards of fundamental measurement should be ignored.

Two evidence issues are, at this stage, worth noting. First, the formulary committee should appreciate that a product approval and launch the evidence base for treatment effect claims is limited. This does not mean, however, that this gap is 'plugged' with the construction of an imaginary reference case world. Second, the formulary committee should also appreciate that if evidence to support claims is required than it will take time. Claims, as noted above, should be credible, evaluable and replicable; reported to formulary committees in a meaningful time frame. Manufacturers must be able to demonstrate how these claims are to be assessed, which brings us back to the product dossier with its evidence base as the key element.

Conclusions

Silver Bullet: a simple and seemingly magical solution to a complicated problem.

The often limited evidence and the obvious absurdity of trying to fill gaps with imaginary evidence as determined by ISPOR/ICER mean that decisions over pricing and access for new products must be provisional. Rather than attempting to shortcut this process with willingness-to-pay thresholds driven by nonsensical cost-per-QALY lifetime models, parties must recognize that there is no silver bullet. There are a number of data elements that will be factored into negotiations. It is up to the formulary committee to judge their relative 'worth' in disease areas and target populations. Of particular relevance, given ongoing commitment by health systems to creating real world evidence as a complement to the wealth of real world data that already exists, a disease specific platform commitment would appear a necessary step. Technically it is straightforward; there is ample experience with registries and the minimum standards that should be met. At the same time, a focus on registries would provide a further necessary link between manufacturers, or a consortium of manufacturers, and patient advocacy groups. A registry could be seen as a public good, a commons if you like, which provides (to mix metaphors) a level playing field for claims evaluation.

References

[1] 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

[2] Piglucci M. Nonsense on Stilts: How to tell Science from Bunk. Chicago: University of Chicago Press, 2010

[3] 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;21:119-25

[4] Dawkins R. The selfish Gene. 30th Anniversary Edition. New York: Oxfod University Press, 2006

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

[6] Magee B. Popper. London: Fontana/Collins, 1973

[7] Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: Time to end malpractice? (Editorial). *J Rehab Med*. 2012;44;97-8

[8] Bond T, Fox C. Applying the Rasch Model. New York: Routledge, 2015

[9] 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

[10] Tennant A, McKenna S, Hagell P. Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value Health*. 2004;7(Suppl 1):S22-S26

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

[12] Kuhn T. The Structure of Scientific Revolutions. 1962

[13] Langley PC. Modeling imaginary worlds: Version 4 of the AMCP Format for Formulary Submissions. *InovPharm*. 2016;7(2): No. 11

[14] P C Langley, Recent Developments in the Health Technology Assessment Process in TR Fulda and A I Wertheimer, *Handbook of Pharmaceutical Public Policy*, New York, Haworth Press, 2007, pp. 457-477.

[15] Langley PC. Nullius in Verba: Version 2.0 of the University of Minnesota, School of Social and Administrative Pharmacy Program, Proposed Guidelines for Formulary Evaluation. *Inov Pharm.* 2016;7(4): No 16