MAIMON WORKING PAPERS No. 24 NOVEMBER 2021

RECONSIDERING VALUE CLAIMS FOR COST-EFFECTIVENESS

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

Abstract

To claim that a therapeutic intervention is 'cost-effective' presupposes that the measure reflects a single or combination of attributes and is credible evaluable and replicable. That is, both the numerator and denominator in the expression are single attributes with ratio level measurement properties; a true zero and invariance of comparisons (an interval scale). This follows from the axioms of fundamental measurement where a ratio measure allows division, as well as the other standard arithmetic operations of addition, subtraction and multiplication. Unfortunately, in calibrating cost-effectiveness, claims in health technology assessment the construct falls at the first hurdle; nether the claim for costs nor the claim for effectiveness in the popular cost-per QALY claim meet the standards of fundamental measurement; the cost-effectiveness ratio claim is meaningless. Of course, if claims are invented through assumption driven lifetime simulations they are on all counts an impossible construct; by design the overall cost-effectiveness claim is purely imaginary. For those who recognize the need to meet standards for normal science, claims for cost effectiveness based on ICER modeling is to be dismissed out of hand. ICER cost-outcomes claims fail, not only because they are hypothetical lifetime constructs, but because even in this context the QALY is an impossible mathematical construct and the costs fail to be linked to units of resource use. Claims for cost-effectiveness must be within a meaningful timeframe, not invented on assumption driven hypothetical lives of target patient populations. A further issue relates to the criteria for accepting or rejecting a cost-effectiveness claim? Apart from its imaginary nature, is there any meaning for a threshold value for cost-effectiveness defined in terms of cost-per-QALY? The purpose of this brief commentary is to illustrate that there cannot be a blanket 'cost-effectiveness' claim for a product as this fails the standards of normal science, notably in measurement theory and the role of single attributes. At best, we can make claims for individual attributes relevant to the impact of a given product in a therapy area and associated resource units, possibly defined in an agreed schedule of unit costs.

Keywords: impossible claims, cost-effectiveness, measurement standards, impossible QALY

INTRODUCTION

A perennial question, by those with little grasp of the manifest limitations of assumption driven imaginary simulations in health technology assessment, is whether or not a new therapy is cost-effective. Judged by the typical context in which the claim is made, the simple response is that the claim is meaningless. For the past 30 years health technology assessment has come to a common agreement to reject hypothesis testing and the agreement on research programs to discover

provisional new facts regarding comparative therapy interventions, in favor of the expeditious invention of evidence by assumption to support lifetime ersatz cost-effectiveness claims [1]. This is clearly an absurd position from the standpoint of distinguishing science from non-science (metaphysics and pseudoscience); yet believers persevere [2]. It is, of course, obvious that if the assumption driven simulation model is evaluating outcomes for a hypothetical lifetime of a target population the outcomes are, by design, non-evaluable; rather the point is that taken on their own terms the various ersatz modeled claims for cost-effectiveness are meaningless from the perspectives of the standards of normal science, and in particular fundamental measurement as it relates to single attributes as the basis for value claims [3].

PROTOCOLS AND TIMEFRAMES

The argument that claims for cost-effectiveness and other associated claims should be accompanied by a protocol that describes how a manufacturer, if challenged, is to demonstrate how the various claims are to be assessed is not new; it was first put forward some 17 years ago in draft proposals for formulary submission guidelines as well as detailed in the various versions of the Minnesota formulary guidelines [4] [5]. The rationale is straightforward: it should not be the responsibility of the formulary committee or other health system decision maker to undertake an assessment of the value claims but that of the manufacturer. The added caveat is that the claim should be evaluated and reported to the formulary committee in a meaningful time frame.

If an analyst or groups such as the International Society for Clinical and Economic Review (ICER) and the International Society for Pharmacoeconomics and Outcomes Research (ISPR) subscribe fully to the relativist position that imaginary non-evaluable lifetime claims for products are as credible as hypothesis testing, then the argument stops right there. Protocols to evaluate invented claims are impossible. There is no need for a protocol as the claims were never meant to be empirically evaluable; they are outcomes of the modeler's imagination and those of other modeler's creating alternative imaginary claims. They must be taken at face value because the belief in assumption driven simulations dominates, irrespective of the fact that there can be a myriad of potentially competing imaginary model claims. As it is illogical to believe that observations and assumptions from the past will justify claims on the future, there is no basis on which to declare one model as opposed to the other as the winner, even though hypothetical life time models are themselves of no merit whatsoever. To declare, on this basis, that one product is cost-effective because one model delivers these assumption driven results is clearly ridiculous. The denial of a role for evaluation protocols is, therefore, the sign that the current modeling approach is bankrupt. The only basis, it is claimed, for modifying an imaginary claim is to change to more 'realistic' assumptions and come up with an alternative imaginary claim. In this belief system, testing and a process of conjecture and refutation to discover new yet provisional facts is irrelevant; It is so much easier to invent evidence for cost-effectiveness claims and commit to non-science [6] .

MEASUREMENT REQUIREMENTS

If a claim for cost-effectiveness is to have merit, then both the numerator and denominator must have ratio properties reporting on a single attribute or an evaluable ratio of attributes. As an example, if there is a claim for cost-effectiveness involving direct medical costs and quality adjusted life years (QALYS) then it is only allowed if both numerator and denominator are ratio measures each defining a single coherent empirically evaluable attribute. The reason for this is quite clear: as a ratio involves division then it is only ratio scales with a true zero and invariance of comparison (interval) property that can support division. No other level of evidence (nominal, ordinal or interval) will suffice. This requirement not only excludes QALYs as a metric for effectiveness, but all of the other outcome claims that characterize the ICER evidence models [7]. The QALY is rejected because it is a mathematically impossible measure constructed with ordinal preferences [8]. The preference scores (EQ-5D-3L/5L) are ordinal, comprising a cluster of ordinal preferences for symptoms that, if aggregated, lack dimensional homogeneity and construct validity. When these algorithms to produce scores (with TTO weights) were developed no though was given to required measurement properties. The result is an ordinal scale that can take negative values and no true zero. Even if attempts were made from real world data or observational studies to emulate a cost-per-QALY claim it would fail at this fundamental level. It follows that any other outcome measures that embodies ordinal preferences and hence impossible QALYs in its construction must be similarly rejected.

The costs assumed by ICER and the expert academic groups supporting ICER, are similarly invalid. Not only because they were never designed to be empirically evaluable, just accepted at face value. If single attribute claims for costs are part of a value claims package, then we should step back and focus on evaluable resource units. This requires value claims expressed in terms of CPT and NDC codes. These can be tracked from any number of large data bases, with the tacking facility defined as part of the protocol. Given assumptions made in the ICER attempt to set budget limits in a 5-year time frame, the value attributes can be defined for product uptake, product substitution and compliance. These are straightforward to identify and can be reported to a formulary committee in a meaningful timeframe.

Similar objections apply in the case of disease specific patient reported outcomes (PRO) instruments if these are intended to support value claims. Once again, the various authors have failed to recognize the limitations imposed by the axioms of fundamental evidence. Many, if not the majority, of these PRO instruments rely upon Likert scales. The Asthma Quality of Life Questionnaire is a prime example with few appreciating that the end product is a multiattribute ordinal scale, not a ratio scale that has been assumed for the past 20 years or more in evaluating response to therapy; the AQLQ is incapable of supporting claims for therapy response [9]. The scale itself comprises 32 items (Likert scales) designed to assess the physical, occupational, emotional and social qualities of adults 17 to 70 years exhibiting mild to moderate asthma. It is a multiattribute instrument with four domains: symptoms (12 items), activity limitation (6 generic and 5 patient-specific items), emotional function (5 items), and environmental stimuli (4 items). Each item response is on a 7 point Likert scale with responses ranging from 1 = maximal impairment to 7 = minimal impairment. The items are in the form of

questions with each of the scale points anchored on a word or phrase and not just the extreme values; descriptors include "totally", "extremely", "very", "moderate", "some" "a little". What is overlooked is that each Likert scale is an ordinal ranking of responses with integers attached; we could equally well have A, B, C, etc. We cannot claim that the psychometric distance between integers 1 and 2 is that same as the distance between 3 and 4. It is not an interval scale. Given all 32 items have the same ordinal property we cannot add the various integer items for each scale to create sub-domains and an overall (even if rescaled) AQLQ score. In other words, if traditional or classical statistical operations are to be attempted with an instrument such as the AQLQ, then the developers need to demonstrate: (i) that all items are of equal difficulty and (ii) that the spaces between each Likert item are of equal distance [12]. If not, then these instruments and any claims based on them must be rejected.

IMAGINARY COST EFFECTIVENESS

Having spent the last 30 years extolling the virtue if the QALY as the gold standard single metric for effectiveness, the fact that it is a mathematically impossible construct, is little short of embarrassing. Generations of scholars and students have demonstrated a complete lack of understanding of the constraints imposed by the axioms of fundamental evidence, let alone embracing a relativist position of the symmetry of competing analytical belief systems, which extols the virtue for decision making in creating hypothetical lifetime claims. This position also endorses the long discredited approach of logical positivism, Hume's problem, and that it is impossible to assemble realistic (i.e., believable) claims on the future from the fact that all past futures have resembled past pasts [10] [11] [12]. This fails Logic 101, as well as demonstrating a failure to understand the basic tenets of the philosophy of science in separating science from non-science (metaphysics and pseudoscience) in applying a demarcation standard.

Lifetime hypothetical claims for imaginary cost effectiveness are the objectives of the ICER reference system modeling. Whether ICER really believes, along with the various academic research centers that create the impossible QALY models, in the role of imaginary value claims to support pricing and access of pharmaceuticals is an open question. If, as has been proposed, the ICER modeling framework is nothing more than eugenics revisited, where the application of thresholds play a key role in denial of access to new therapies, then there is the question of denying care on imaginary claims [13].

In the ICER imaginary world, and to the wider ISPOR audience of believers, QALYs and lifetime incremental cost-per-QALY imaginary claims are critical to overall imaginary claims for cost-effectiveness at different pricing points and, as the pièce de résistance, the claimed social price of a health benefit price benchmark (HBPB). This ICER devised artifact is the imaginary price which reflects prices aligned with commonly cited long-term imaginary cost-effectiveness thresholds ranging from $100,000 to $150,000 per QALY gained. The price represents discounts or price premiums from wholesale acquisition cost (WAC) that would be required to reach these cost-effectiveness thresholds. A price range, net of any discounts and rebates that supposedly align fairly with a treatment's imaginary added benefits for patients over their hypothetical lifetime. Prices at or below these imaginary thresholds apparently help ensure that the health benefits gained by patients using new

treatments are not outweighed by health losses due to long-term cost pressures that lead individuals to delay care, abandon care, or lose health insurance. This is quite unacceptable as the entire edifice is built on assumption driven simulations incorporating mathematically impossible QALYs.

Manufacturers who engage with ICER (or similar agencies) have the options of subscribing to imaginary ICER worlds as an unavoidable and easily met rite of passage, ignore ICER altogether or engage in a more rigorous critique of the limitations and relevance of the ICER modeling charade and produce their own submission for formulary review. It is not an easy choice given that for those manufacturers who have health economics units, the belief held by the staff is likely to be entirely consistent with the technology approximate imaginary evidence paradigm. There has to be a learning curve, not for those in technology assessment but also for formulary committees and other healthcare decision makers. One possibility is to present formulary committees with a list of questions to address when assessing value claims, as demonstrated in the latest version 3 of the Minnesota formulary guidelines [14] .

PROGRESS VERSUS STAGNATION

There is no evidence that claims for cost-effectiveness over the past 30 years have ever been accompanied by protocols detailing how the claim can be empirically evaluated and replicated in target patient populations. The answer is obvious: no one thought that the claim would be other than imaginary; an approximate endeavor. An indicative claim or one that is approximate, shielded by sensitivity analyses and presented in probabilistic terms was deemed sufficient for formulary committees to make decisions. This is made quite clear in the most cited textbook by Drummond et al for health care decision making, which is most appropriately seen as a primer for constructing multiple assumption driven imaginary simulation models [15]. There was, by agreement in the early 1990s, the decision to abandon hypothesis testing and the provisional discovery of new facts in therapy areas in favor of imaginary approximate information cost-effectiveness claims. As noted in previous commentaries this has led to the situation where health technology assessment as a discipline is the only one in the social and health sciences where its claim to 'fame' and 'relevance' rests on creating imaginary non-evaluable modeled claims. The concept of progress is abandoned in favor of inventing assumption driven 'evidence'. A retreat from Popper's concept of conjecture and falsification to a belief that claims on the future are validated by claims or observations from the past. The audience for these claims is sufficiently gullible (or at least sufficiently uninformed) to take the ICER imaginary claims at face value.

The denial of the standards of normal science in favor of non-science (metaphysics and pseudoscience) is seen in the failure to propose how claims for cost-effectiveness are to be evaluated and replicated. Formulary committees are at equal fault in failing to insist on evaluation protocols; rather they submit to the imaginary claim belief system. If there are no demands for evaluation protocols them manufacturers are under no obligation to provide them. All concur, apparently, in the relativistic belief that decisions for drug pricing and access can be supported by imaginary claims.

If the only contribution health technology assessment can make is to invent evidence, then we have abandoned the provisional discovery of new facts in therapy interventions in favor of the acceptance of imaginary modelled claims that effectively bar any notion of the standards of progress and the normal science paradigm. All the ICER paradigm has to offer is the prospect of, at some time in the future, modifying assumptions to produce another version of the imaginary outputs. This is hardly reassuring as it reinforces the criticism that that ICER claims are always provisional and can be successively adjusted as assumptions change to invent more non-evaluable claims. It would be more reassuring if there was a commitment to a research program that focused on discovery; but health technology assessment has rejected that commitment.

CONCLUSIONS

The claim that product A is cost-effective compared to Product B, where effectiveness is a composite mathematically impossible metric attempting to embrace a range of attributes lacks all credibility. If the objective, in the ICER imaginary universe, is to argue for cost-per-QALY thresholds to establish a social price for a product, then we are similarly in a different relativist universe; a universe where ordinal scales have magical ratio properties, QALYs exist, and where formulary decisions are best served by the embrace of assumption driven simulations to invent evidence. Either formulary committees recognize and accept the role of invented evidence as critical to formulary decisions or they have failed to grasp the manifest failures in creating assumption driven evidence. The status of ICER claims is clear cut: they are a failure and should not be entertained in formulary decisions.

The way forward, if we are to respect and conform to the standards of normal science, is to establish, for each target population in a disease state, a profile of required attributes. Until we catch up with instruments that meet the required measurement standards defining the required attribute set, physical and latent, these could be designated as primary or secondary, with the latter anticipated as part of an ongoing research program. Each proposed attribute would conform to required ratio measurement properties. Attributes would include clinical outcomes that may already have been detailed and reported in pivotal phase 3 trials or be part of a proposed research program to explore further therapy outcomes. Formulary committees would have discretion regarding protocols for outcomes assessment. At the same time further attributes or outcome might include quality of life such as the N-QOL [16], resource utilization impact following market entry and compliance.

The fact is that neither ICER nor the various academic research centers that act as consultant simulation model builders to ICER can defend their commitment to inventing evidence. ICER is well aware of the manifest deficiencies in their assumption driven simulation paradigm; they are well aware, as are ISPOR, that it is an analytical dead end. The question is: how long do we wait for an acknowledgement of past errors? We should not be overly optimistic.

REFERENCES

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

[2] Langley P. Peter Rabbit is a Badger in Disguise: Deconstructing the Belief System of the Institute for Clinical and Economic Review. *InovPharm*. 2021; 12(2): No 22
https://pubs.lib.umn.edu/index.php/innovations/article/view/3992/2855

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

[4] Schommer J, Carlson A, Rhee T. Validating pharmaceutical product claims: questions a formulary committee should ask. *J Med Econ*. 2015;18(12):1000-1006

[5] 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

[6] Langley P. Abandoning Eugenics and the QALY. *InovPharm*. 2021;12(3): No.20
https://pubs.lib.umn.edu/index.php/innovations/article/view/4291/2939

[7] Langley P, McKenna S. Fundamental Measurement and Quality Adjusted Life Years. *Value Health*. 2021;24(3):461[letter]

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

[9] Juniper E, Guyatt G, Epstein R et al. Evaluation of impairment of health-related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax*. *1992*;47:76-83

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

[11] Russell B. The Problems of Philosophy. London, 1912.

[12] Magee B. Popper. London: Fontana Books, 1974

[13] Pigliucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010

[14] 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

[15] Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes. 4th Ed. New York: Oxford University Press, 2015

[16] Langley P, McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*.2021;12(2):No. 6
https://pubs.lib.umn.edu/index.php/innovations/article/view/3798