

MAIMON WORKING PAPERS No. 17 SEPTEMBER 2021

MORE NONSENSE ON STILTS: THE INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW'S QALY COMPANION - THE MYFFIC EQUAL VALUE OF LIFE YEARS METRICS

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

Abstract

The Institute for Clinical and Economic Review (ICER) must be complimented on its commitment, if not obsession, with producing simulated assumption driven outcome measures which fail the standards of normal science; measures which lack credibility, empirical evaluation and replication. As a companion artifact to the mathematically impossible quality adjusted life year (QALY) ICER has made extensive use in its evidence reports of the equally impossible progeny: the Equal Value of Life Years Gained charade with the Equal Value of Life Years (evLY) and Equal Value of Life Years Gained (evLYG) metrics. As with the ICER reference case modelling and its devotion to the mathematically impossible quality adjusted life year (QALY) , ICER continues in its crusade for impossible non-evaluable claims in health technology assessment . If analysts believed, quite correctly, that the QALY was a step too far, ICER's latest obsession with the Equal Value of Life Years Gained (evLYG) metric, shows the resilience of pseudoscience. The purpose of this brief commentary is to deconstruct these metrics, with the recommendation they be both rejected. We are looking to a reformation in health, or in more modern terms, a new paradigm that respects the standards of normal science.

INTRODUCTION

The Institute for Clinical and Economic Review (ICER)) is obsessed, if that is not too strong a word, with creating imaginary claims for therapies to support absurd recommendations for pricing and access. While the term relativist is probably a step too far in ICER's understanding, it is to relativism that they cling. Truth, as advocated by ICER, is relativistic: knowledge claims admit all types of analysis and we cannot make the case that one source of knowledge is superior to another ¹. This is the ICER bolthole; we cannot make claim to superior evidence as any analytical framework, or absence of one, is equally valid. Science is not necessarily a way to come to grips with reality. Evidence is never discovered, but constructed within a particular social community. For ICER, therefore, science is about

rhetoric, persuasion and authority. The more people believe, mostly in ignorance, in the ICER invention of evidence the more influence ICER will have. This is the antithesis of what science does which is to show a consensus view must be abandoned when it at odds with the evidence; with ICER models and claims this is impossible. The ICER claims are self-referential; there is no referral to an external authority. ICER can never be wrong unless ICER decides to change assumptions.

The manifest deficiencies of the ICER reference case are well known and widely recognized. ICER is committed, not to progress and the discovery of new yet provisional facts, but to the creation of assumption driven simulations to support cost-per-QALY claims for pricing and access recommendations, ICER is opposed to the scientific method; relying instead on re recycling of old evidence and dubious assumptions to support imaginary claims.

The Equal Value of Life Years and Life Years Gained metrics (evLY and evLYG) are squarely in this ICER relativistic tradition of inventing non-verifiable claims. In this imaginary world, the EVLYG is intended to value all gains in life years at the full value of a healthy life year, defined by an ordinal score in the range 0 to 1. Gain, according to this ICER construct, is a metric that holds regardless of age, gender, disability or illness. All gains are valued equally across all possible disease states. Unfortunately, the entire concept is misplaced. There are no redeeming features; the entire concept is a mathematical charade. Applying the ICER methodology can lead to a multitude of competing evLYG claims; all you need to is to change assumptions and maintain your steadfast believe in the impossible QALY².

The purpose of this note is to deconstruct the evLY and the evLYG, a relatively simple undertaking. Unfortunately, including ICER and a number of academic modelling groups, there will be, as in any odd belief system, a hard core of acolytes whose fidelity to the ICER belief system will be undiminished.

DECONSTRUCTION: STEP ONE

The centerpiece of the ICER contribution to cost-effectiveness claims in health technology assessment is the reference case model. All ICER metrics stem from this model. The model invents evidence to support cost-effectiveness claims driven by assumptions for the experience of patients exposed to a new therapy. The therapy benefits are expressed as quality adjusted life years (QALYs); the

costs as a subset of direct medical costs (no increase in drug costs from baseline) projected over the same hypothetical lifetime of a hypothetical patient population. That said, it is quite obvious that the reference simulated evidence fails the standards of normal science: the claims lack credibility, cannot be empirically evaluated and cannot be replicated^{3 4}. ICER's claims for cost-effectiveness are the antithesis of the notion of progress in science captured by modified approaches to a process of conjecture and refutation. If we are to take health technology seriously and not engage in fantasy modelling, three criteria must be met:

- All claims must be credible and presented as single attributes within a disease area
- All claims must be empirically evaluable conforming to the axioms of fundamental measurement (ideally with ratio measurement properties)
- All claims must be replicable for empirical evaluation in different target patient populations within a disease area

The ICER reference fails on all three criteria. As the evLYG rests on the ICER reference case model, it cannot meet these criteria as a scientific value claim. It is pseudoscience.

DECONSTRUCTION: STEP TWO

ICER utilizes a 'utility' (more accurately preference) value of 0.851 for the value of a healthy life. This is derived from age and gender adjusted utility from the US EQ-5D-5L preferences⁵. Apart from the fact that the EQ-5D-5L preferences are ordinal scores and can only support non-parametric statistics, including any manipulations to create an ersatz 'perfect health' reference point, the choice is singularly inept because with 5 symptoms and 5 response level for each symptom (with generic perfect health defined solely in terms of these symptoms), the instrument yields 3,125 health states (5^5), but with 20% of the health states (625) defined by negative preferences or states worse than death (preference score range -0.5 to 1.0) The 0.851 is not perfect health as defined by the EQ-5D-5L with 3,124 health states < 1 . It is not clear in this adjusted perfect health for the US population how the states worse than death are accommodated: are they put to

one side or are they used to deflate to 0.851? Or can we live with the claim for population adjusted perfect health at a preference score of 0.851 (not 1.0) which will lead to a deflated new QALY claim. It is not clear how we would value claims based on preferences in the range 0.852 to 1.0? The issue becomes even more bizarre once other instruments (including the EQ-5D-3L) are accommodated. Although they are all ordinal measures, each has a different preference algorithm with varying attributes and response levels. The result is that (ignoring the limitations of an ordinal score) preference instruments will, in principle at least, each have their own equivalent "0.851".

DECONSTRUCTION: STEP THREE

If the magic 0.851 is intended to capture perfect health then ICER might be aware that each multiattribute instrument embodies its own definition of perfect health given the symptoms or attributes covered, the response levels and their descriptions. So, as noted, we would have a range of "0.851" scores which would add to the general criticism the ICER reference case that it is only one of a possible multitude of assumption driven simulations each with their own cost-effectiveness claims. Every assumption driven claim by ICER fails the standards of normal science. We should not even maintain it is 'provisional' subject to future 'clarification' from new assumptions. This would imply that the ICER reference case model as some claim to be a 'meaningful' baseline case. This is just false; it is one of many.

The application of generic multiattribute preferences has long been challenged on the groups of content validity (let alone the umbrella concept of construct validity). Apart from the lack of construct validity, perfect health is a nebulous concept (indeed, if the concept has any meaning) by those who argue for disease specific measures, but defined as single attributes captured as representative of that disease state and valued separately as credible claims. As it stands the application for ICER is to multiply the simulation modeled life extension (which exclude quality of life improvements which are impossible anyway) by the magic 0.851.

DECONSTRUCTING: STEP FOUR

To construct the evLY and the evLYG the following imaginary outputs are required from the assumption driven imaginary simulation for a particular product (s) [A] and comparator(s) [B] for a hypothetical year in a disease state⁶:

- Product A life year (scale 0 – 1) [e.g., 0.90 or 10.80 months]
- Product A QALY (scale 0 – 1 [e.g., 0.63 or 7.56 months]
- Product B life year (scale 0 – 1) [e.g., 0.70 or 8.40 months]
- Product B QALY (scale 0 – 1) [e.g., 0.42 or 5.04 months]
- Community preference score for perfect health 0.851 (from weighted US EQ-5D-3L ordinal preference scores)

The algorithm to create the evLY or the equal value of life year for product A is:

- $$\begin{aligned} \text{evLY (A)} &= ((0.63/0.90)*0.70) + (0.851* (0.9 - 0.7)) \\ &= 0.66 \end{aligned}$$

The evLYG is then calculated by subtracting the product B QALY (0.42 which is defined as the evLY for B) from the evLY for product A (0.66):

- $$\begin{aligned} \text{evLYG} &= \text{evLY(A)} - \text{evLY(B)} \\ &= 0.66 - 0.42 \\ &= 0.24 \end{aligned}$$

These values can be computed for each Markov cycle in the imaginary assumption driven model and then aggregated (and discounted) over the hypothetical model lifetime.

Reporting on evLYs and evLYGs is now a standard element in the ICER repertoire of imaginary outcomes. As an example, the recent draft evidence report for mavacamten in hypertrophic cardiomyopathy, presents results for evLY and

evLYGs⁷. While this modelling is entirely fictional and to be taken seriously, the numbers produced are illustrative (Tables 4.3 and 4.4). For the lifetime hypothetical model the base case for mavacamten with standard first line therapy QALYs yields 13.51 QALYs, life years are 16.58 and evLY 13.51 years compared to standard first line therapy alone with QALYs 12.54, life years 16.58 (no mortality assumed) and evLY 12.54 years. With imaginary costs of \$1,568,000 and \$434,000 respectively, costs per QALY and costs per evLY can be imagined. To extend the fictional analysis, incremental costs per QALY can be estimated with cost per QALY gained of \$1.2 million, which is the same as the cost per evLY gained. Note in this model QALYs and evLY are the same so the evLYG is zero.

CONCLUSIONS

The evLY and evLYG metrics have no scientific status; they are invented with no chance of ever being empirically evaluated. It is questionable what the evLYG means. The calculated evLYG is nothing more than the results of bringing together imaginary assumption driven outcomes from the simulation model; as such they are meaningless unless one believes in rejecting fundamental measurement. The QALYs are impossible mathematical constructs as the preference scores are ordinal and the life years are created by the model structure and the state transition matrices. Welcome to the imaginary world of ICER where belief in the standards of normal science and fundamental evidence are rejected. Instead, we have a fantasyland of misinformation and falsehood which observers might find difficult to come to terms with, yet with a hard core of believers, who are prepared to deny care on ICER's terms. Perhaps ICER persists in creating one imaginary modelled evidence report after another because, as a mystery, where for the faithful mystery is a virtue and, as Tertullian asserts, '*certum est quia impossibile est* (it is certain because it is impossible)⁸

REFERENCES

¹ Wootton D. The Invention of Science: A new history of the scientific revolution. New York: Harper Collins 2015

² Langley P. The Great I-QALY Disaster. *InovPharm*. 2020;11(3): No 7
<https://pubs.lib.umn.edu/index.php/innovations/article/view/3359/2517>

³ Langley PC. Nonsense on Stilts –Part 1: The ICER 2020-20234 value assessment framework for constructing imaginary worlds. *InovPharm*. 2020;11(1): No. 12
<https://pubs.lib.umn.edu/index.php/innovations/article/view/2444/2348>

⁴ 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>

⁵ Pickard A, Law, Jiang R et al. United States Valuation of EQ05D-5L health states using an international protocol. *ValueHealth*. 2019;22(8):931-41

⁶ O’Day K, Mezzio D. Demystifying ICER’s Equal Value of Life Years Gained Metric. *ISPOR Value & Outcomes Spotlight*. 2021;7(1)

⁷ Wasfy JH, Walton, SM, Beinfeld M, et al. Mavacamten for Hypertrophic Cardiomyopathy: Effectiveness and Value; Draft Evidence Report. Institute for Clinical and Economic Review, August 18, 2021. <https://icer.org/hypertrophic-cardiomyopathy-2021/>.

⁸ Dawkins R. *A Devil’s Chaplain*. New York: Houghton Mifflin, 2004