MAIMON WORKING PAPER No. 5 MARCH 2024

THE PRESUMPTIVE FAILURE OF H.R. 485 IN THE SENATE: NO ONE UNDERSTANDS THE QALY CONSTRUCT OR ITS MATHEMATICAL IMPOSSIBILITY

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

ABSTRACT

If one wished to open a book on the odds of H.R. 485 passing the Senate the result would heavily favor failure. This failure is due entirely to a monumental inability by Republicans (the sponsors), Democrats, the Congressional Budget Office (CBO) and the myriad consultants and others who have advised and voiced their opposition to recognize that the QALY is a mathematical impossibility. The reasons for the QALY being mathematically impossible and the standards required for value claims after the QALY have been detailed in a recent Maimon Working Paper (No. 3 2024). It comes down to a total lack of knowledge and appreciation of the standards for fundamental measurement and high school mathematics. The fact that these standards have been recognized and applied hundreds, if not thousands, of times in the assessment of patient reported outcomes (PRO) over the past 60 years, is a bridge to far for all of the participants in this debacle. Thet have entirely themselves to blame in failing to ask a simple question: how is the QALY constructed?

The purpose of this brief note is to build on the previous critique to focus on some of the more egregious failures and then propose a way forward for value claims for competing pharmaceutical products. The way forward is well documented with a number of instruments, applying fundamental or Rasch measurement standards. This does not involve a generic QALY equivalent. The QALY concept is dead. The key is capturing as value claims the needs of patients and caregivers in target disease states and stage of life. This note concludes with a short statement of three reasons why the QAY and other measures with the same failings should be eliminated with amendments to the Social Security Act. These are: (i) The QALY fails the standards accepted in the sciences of fundamental measurement and must be disallowed; (ii) All patient reported outcomes instruments that fail the standards for fundamental measurement should be disallowed; (iii) All patient reported outcomes instruments that support negative QALY scores for states worse than death should be disallowed. Which, in the simplest terms, comes down to disallowing all patient reported outcome instruments that cannot be demonstrated to meet the required standards for fundamental measurement. A requirement that is some 40 years overdue.

INTRODUCTION

The theme from the previous critique of H.R. 485 is that the generic multiattribute quality adjusted life year should have been smothered at birth [1]. The fact that it has been actively promoted over the past 50 years with over 20,000 QALY references in PubMed is a sad commentary on the standards or belief system in health technology assessment (HTA) [“QALY.” 25,775 hits 16 March

2024] and its continued pre-eminent status in HTA as evidenced by a recent laudatory paper is a monument to a failure to understand fundamental or Rasch measurement [2] [3]. To this, as detailed in the previous critique, is the central role played by the QALY in assumption driven modelled simulations and their nonsensical claims for cost-effectiveness. This, as detailed in another recent Maimon Working Paper, is due to the acceptance of seven key false premises by model builders in HTA [4]. These false premises are: (i) the pre-eminence of imaginary claims; (ii) the irrelevance of demarcation; (iii) the apotheosis of cost-effectiveness; (iv) the rejection of induction; (v) the rejection of fundamental measurement; (vi) the acceptance of composite ordinal scores as single attribute ratio measures in disguise; and (vii) the acceptance of the cost-per-QALY thresholds to drive resource allocation.

The conclusion to this further critique of the HTA belief system is that the QALY and the creation of imaginary cost-effectiveness claims is, to apply a phrase accepted by the philosophy pf science, bullshit with scientific pretensions [5] [6] . Unless we understand why the Rasch model is the only way of transforming observations to measurement then, by default, the impossible QALY will maintain its position [7].

HOUSE DEBATES

The gage the lack of appreciation of how the QALY is constructed and its failure in terms of the underlying utility or preference scores having only ordinal properties is abundantly apparent in the complete absence of these facts on the various House committee encounters. On February 7th, 2024 , Frank Pallone Jr (D-NJ) expressed his opposition to H.R. 485 in terms of his presumption that this Republican bill was a trojan horse to undermine the progress made in lowering prescription drug costs and, inevitably, the destruction of the Affordable Care Act. Apart from the histrionics, a key point made was in respect of the vagueness in the language of the proposed amendments to the Social Security Act. Pallone focused, rightly, on the phrase ‘QALYs and similar measures ‘ . This is a meaningless requirement: if we want to eliminate similar measures to QALY then we have to ask what are the elements of the QALY construct that are objectionable? This was not raised. But the answer is straight forward: all value claims for pharmaceutical products must meet the standards for fundamental measurement: the instrument that supports the claim must meet Rasch standards for a single attribute, linear, interval and invariant measure. This is the only acceptable metric for pricing and access negotiations. This is not ‘discrimination’ but a standard to apply and avoid imaginary cost-effectiveness claims. As detailed here and in the previous critique, the patient or caregiver voice is paramount: we must endeavor to meet the unmet needs of target patient populations. There is no universal cost-per-QALY metric; we must make decisions for specific diseases and patient needs.

Once the standards of normal science and fundamental measurement are recognized and put in place as an information target, then it is up to manufacturers to meet these standards for product value claims supported by protocols for the assessment of those claims in a meaningful time frame.

It is doubtful if the Democrats would be agreeable to being defined as the party that supports mathematically impossible measures; but that is what the rejection of the H.R. 485 will amount to if they maintain their support for the QALY. The issue is one of failing to ask the right questions

CONGRESSIONAL BUDGET OFFICE

The Congressional Budget Office (CBO) appears to lack any awareness of the mathematical impossibility of the QALY [8]. This undercuts their estimates of the estimated budgetary effects of H.R. 485 by failing to recognize that in terms of federal programs, including Medicare and Medicaid, they are proposing to value the impact of imaginary claims. If the QALY is unable to assess the relative value of medical interventions the analysis should stop right there. They have no conception of the standards of fundamental measurement. The present prohibition on use of the QALY are recognized but the CBO expects, for no good reason for an imaginary construct, why the QALY should continue to be used. As detailed later, the CBO seems quite unaware that the impossible QALY facilitates the scoring of states worse than death with dubious ethical implications for the use of the QALY to support resource allocation and the denial of care to patients who are scored as less deserving.

INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW

The Institute or Clinical and Economic Review (ICER) has long been one of the most stalwart supports of the assumption driven simulation modelling fraternity; indeed, such models with recommendations for threshold cost-per-QALY pricing, access to care and budgetary impacts are central to their business case. In defense of their stand against H.R. 485, a proposal by Epstein, Becker, Green PC (EBG) provides the traditional defense of the QALY in terms of its role in comparisons across disease areas in the allocation of health care resources; the ICER model [9].

The QALY according to EBG 'can help distinguish the value of drugs based on what matters to patients' . This is patently incorrect for two reasons: first, the QALY is based on a community judgement of what is a preferred not on any reference to the interests and needs of patients and, second, EBG have no idea of the standards of normal science with evaluable claims or of the standards of fundamental measurement. Certainly, it represents the received wisdom of the HTA belief system in the QALY, but that is why the contribution of EBG is nothing more than special pleading for a false metric with its unfortunate eugenic implications of deciding who and who should not have access to therapies. Perhaps EBG should have been told that the algorithms to produce the utility and preferences scores can generate negative values for states worse than death and, by construction, negative QALYs. In the US negative scores account for 20% of EQ-5D-5L health states in the range 1 (= perfect health) to – 0.573 [10].. It can be appreciated that EBG are trying to make, from essentially a legal perspective, the best case possible for the QALY. Unfortunately, the horse has bolted, it left some 80 years ago, with the formalization of measurement classification [11] . While EBG must have missed the memo they should be aware of what is high school mathematics. Constructed from multiattribute instruments with composite ordinal utility and preference scores, the QALY is mathematically impossible; a fact of which ICER is certainly aware.

STATES WORSE THAN DEATH

Apart from the failure to meet the standards of normal science and fundamental measurement the fact, as just noted, that the community preference weights that attach to QALYs typically produce negative scores in up to 20% or more of the health states, bundles of health dimensions and

symptom levels, that are created by the scoring algorithm. The latest value set for the US reports 20% negative scores for the health states defined by 5 health dimensions and 5 response or problem levels.

It surprising that this has not apparently been raised by the advocates of abandoning the QALY. Certainly, disabilities are a concern (if we believe the metric), but how do these relate to negative utility and preference scores? The fact that negative scores are commonplace is typically glossed over as an unfortunate by-product of the ordinal utility or preference algorithm and usually subsumed in an average score for the target patient group. This raises another error as an ordinal score cannot support the stand range of arithmetic operations, only median and modal values. This aggregation means, presumably, that the false QALY score is lower overall with implications, if it is used for resource allocation, as it includes those who are unlikely to benefit, in these terms, from therapy. Perhaps a filter questionnaire that, is completed in the presence of the physician, will yield an ordinal score to determine access for those with a weighted positive utility or preference score?

As major health care problems, including those related to disability occur in life's later stages, the fact that the utility and preference scores for these groups are likely to be negative, is more than enough reason to abandon the QALY. There is, across the board, an implicit bias against the older age groups, including disabilities and those with disabling rare disease. In a recent paper by Schneider the QALY is described as ableist: the case is straightforward, the QALY with its unfortunate eugenic implications, it discriminates against people in poor health, they are assigned fewer QALYs so that extending the lives of these people reduces overall population health [12]. To improve overall health, therefore, resources should be reallocated from those with negative scores to those with positive scores. While the Schneider analysis is compelling it should be noted that he does not recognize that the utilities and preference that support the QALY are an ordinal score and the QALY is an impossible construct.

It is worth noting that if scores are assigned with an instrument meet Rasch standards, there can be no negative scores. This is by design where the analysis is needs based and assigns therapeutic impact by the extent to which needs a better met for target patient populations capturing the ability of individuals to meet their needs. The needs are determined by patient interviews not aby clinicians who try to bundle and score health states that they determine are appropriate.

COMMUNITY PREFERENCES

The valuation of health states utilizing multiattribute instruments such as the EQ-5D-3L is based upon community preferences; the interests of the patient are of no concern. The health states and the weights attached are purely clinical, which may not be related to the needs of the patients. Schneider points out a Swedish study asked patients to value the health state they are in (using the EQ-5D-3L) with its 5 health dimensions and 3 problem levels. This differed significantly from the usual findings where community weights are applied. The patients did not report any health state worse than death and with a value of 0.34 even the one worse health state had a relatively high value. There are no negative QALYs. This may be contrasted, again for the EQ-5D-3L,in the UK whereof the 243 health states 34.6% take negative values (and negative QALYs). As noted above,

applying the EQ-5D-5L in the US, only 20% of the health states report negative scores. This does not mean that, judged by community preferences, the eugenics implications are less, just that these two instruments produce markedly different results for the same heath state.

CONCLUSION: WITHOUT A CLUE

Oxford Languages' definition of clueless (adjective) *is having no knowledge, understanding or ability;* one that seem singularly apt when applied to the current fiasco over the prohibition of the QALY. While the abandonment of the QALY, even for the wrong reasons, is to be welcomed it unfortunately seems highly likely that with House vote on H.R. 485 (211 to 208) it will fail in the Senate. This is an outcome that is due to the lack of appreciation that a more fundamental reason for abandonment is that the QALY has no scientific status judged by the standards of fundamental measurement. This was never made clear; possibly because those advising had no idea of the required standards of normal science and fundamental measurement. It leads to imaginary claims for cost-effectiveness, which are meaningless, but which are widely accepted, along with a naive belief in the magical properties of the QALY by less informed or clueless decision makers.

It's a sad commentary on the lack of skills in recognizing the standards of normal scientific and fundamental measurement by those who see themselves as advocates of QALY abandonment that the argument for fundamental measurement is the only reason for disposing of the QALY. The argument for disabilities has an unfortunate implication that the QALY is a measure but that it is inappropriate when it comes to disabilities. This misses the point: not only is the QALY not a measure, but an understanding of how the QALY is constructed points to the nonsense of states worse than death which could include disabilities, due to the community preference weighting. Focusing on disabilities without raising these issues seriously weakens the argument. Subsidiary reasons include the focus on assumption driven imaginary simulated claims which would cease to have any merit (they had none to start with) when expressed as cost-per-QALY outcomes and imaginary cost-effectiveness with imaginary thresholds.

If there is to be any real progress in developing meaningful value proposition from patients and caregivers in target disease states then the tools are readily available. All that is required is to set out the case for an alternative to the generic QALY and other disease specific measures which deny the standards of normal science. To accomplish this, we must follow the Rasch prescription for transforming observations into measures. Rasch is the necessary and sufficient basis for creating single attribute or unidimensional interval, linear and invariant measures. That is the only requirement for disease specific value claims; there is no basis for generic scores as the focus must be on needs of patients within those target populations. After over 60 years is it asking too much for a new start in health technology assessment that pays attention to scientific credibility and not culpable nonsense in value claims? Have we dropped the ball? The answer is probably yes, because no one in this debate asked: how is the QALY constructed?

If a believable case is to be made for amending the Social Security Act, all that is required is to recognize three key points:

- The QALY fails the standards accepted in the sciences of fundamental measurement and must be disallowed;

- All patient reported outcomes instruments that fail the standards for fundamental measurement should be disallowed; and
- All patient reported outcomes instruments that support negative QALY scores for states worse than death should be disallowed.

Which, in the simplest terms, comes down to disallowing all patient reported outcomes instrument that cannot be demonstrated to meet the required Rasch standards for fundamental measurement. A requirement that is some 40 years overdue [7].

Finally, this is not a debate; it is about incontrovertible facts based on fundamental measurement and the application of patient reported outcomes instruments that meet these standards. We have the tools to do this. This note may have helped for those engaged to have a more coherent understanding of the issues involved.

REFERENCES

[1] Langley P. Will we finally abandon the mathematically impossible QALY with H/R. 485 passing in the Senate. *Maimon Working Paper No. 3* March 2024 https://maimonresearch.com/wp-content/uploads/2024/03/Maimon-WP-No.-3-March-2024-V5-1.pdf

[2] Spencer A, Rivero-Arias O, Wong R et al. The QALY at 50: One story many voices. *Soc Sci Med*;296:114653

[3] Bond T, Yan Z, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th Ed). New York: Routledge, 2021

[4] Langley P. False Premises and Imaginary Claims: A proposed new start for health technology in Australia. *Maimon Working Paper No. 4* March 2024 https://maimonresearch.com/wp-content/uploads/2024/03/Maimon-Working-Paper-No.4-March-2024-V7.pdf

[5] Frankfurt H. On Bullshit. Princeton: Princeton University Press, 2005

[6] Moberger V. Bullshit, Pseudoscience and Pseudophilosophy. *Theoria*. 2020;86:595-611

[7] Wright B, Linacre J. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989; 70(12):857-60 https://www.researchgate.net/publication/20338407_Observations_are_always_ordinal_measurements_however_must_be_interval/link/5563b02408ae9963a11ef326/download

[8] Congressional Budget Office. H.R. 485 Protecting Health Care for All Patients Act of 2023.Cost Estimate. https://www.cbo.gov/system/files/2023-09/hr485_0.pdf

[9] Morris F, Gabay A. ICER Analyses and Payer Use of Cost-Effectiveness Results based on the QALY and evLYG are consistent with ADA protections for individuals with disabilities. Epstein, Becker& Green, PC. Washi9ngton DC. https://icer.org/wp-content/uploads/2021/02/ICER-Analyses-and-Payer-Use-of-Cost-effectiveness-Results-Based-on-the-QALY-and-evLYG-Are-Consistent-With-ADA-Protections-for-Individuals-With-Disabilities.pdf

[10] Pickard A, Law E, Jiang R%$ et al. United States Valuation of EQ-5D-5L Health States using an international protocol. *ValueHealth*. 2019;22(8):931-41

[11] Stevens S. On the Theory of Scales of Measurement. *Science*, (New Series). 1946; 103; 2684 (677-680)

[12]Schneider P. The QALY is ableist: on the unethical implications of health states worse than death. *Qual Life Res*. 2022;31:1545-1552