MAIMON WORKING PAPER No. 6 March 2024

TOWARDS A NEW PARADIGM IN HEALTH TECHNOLOGY ASSESSMENT: THE CRITICAL IMPORTANCE OF ABANDONING THE QALY AND SIMILAR FALSE CONSTRUCTS

Paul C Langley, Ph.D., Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis MN and Lecturer, School of Pharmacy, University of Wyoming, Laramie WY

Abstract

The current approval of H.R. 485 in the House of Representatives, although squeaking through with Democrats opposed (211 vs 208), is unlikely to achieve any success in the Senate unless the various submissions lift their game and address the fundamental question: what succeeds the QALY? Certainly, we can agree on arguments that the existing QALY (or QALYs to be more accurate) can be dismissed in terms of their failure to meet the standards of fundamental measurement; but that only takes us to first base. It leads nowhere unless there are proposals from QALY abolitionists for a new start in health technology assessment (HTA) that emphasizes the importance of meeting the standards of normal science and fundamental measurement. At the same time the abolitionists must recognize that the current HTA belief in imaginary cost-effectiveness claims is driven by the QALY. Failure to abolish the QALY in Federal programs will give the green light to paper mills and others to promote the pseudoscience of non-evaluable false claims for cost-effectiveness.

INTRODUCTION

The literature is replete with criticisms of the QALY. Unfortunately, the impact of these complaints is predicated upon the QALY meeting required fundamental measurement standards: a single attribute, linear, interval and invariant patient reported outcome (PRO) measure. The QALY fails to meet these standards [1]. This means the arguments for bias and discrimination take second place, as detailed in two previous commentaries on this legislation, to the more substantive rejection of the QALY in terms of fundamental measurement.[2] [3]. In this note the focus will be on the apparent failure to emphasize the failure of the QALY as a measure and, of equal importance, the steps which need to be taken in effecting a transition to measurement in quality of life that meets required standards

If there is a single issue on which those looking to support abandoning the QALY might raise is the apparent lack of understanding by those in value claims assessment of the need to meet the standards of Rasch or fundamental measurement [4]. Leadership asks for knowledge; in this case knowledge of the standards for measurement which have been readily available and adopted globally over the past 80 or more years. This is sadly lacking. It is unlikely to improve at any time soon. It is this failure which will determine whether the *Protecting Health Care for All Patients Act of 2023* is seen as a catalyst for change in health technology assessment (HTA) or an inconsequential tempest in a teapot. This lack of understanding goes a long way to support the presumption that the QALY abolition will fail in the Senate. It is not a defense of the QALY; there is none. As it stands the present commitment to HTA in the US is nothing more than a commitment to assumption driven modelling and measurement pseudoscience [5].

QALYs AND SIMILAR MEASURES

Taken at face value the language of H. R.485 *to prohibit the use of quality-adjusted life years and similar measures* (emphasis added) is, from a fundamental measurement perspective, quite straightforward. It would have been possible to be more explicit, for example, to include the in the legislation a more the wording *QALY and other PRO claims that fail the standards for fundamental measurement,* but the existing wording is perfectly acceptable. The point is that the QALY fails the standards of fundamental measurement. It is, strictly speaking, not a true measure with required interval or ratio properties. It was not developed applying the Rasch mathematical model where data items are selected to fit the model for a patient reported outcome (PRO) instrument. Instead, the instrument is developed by trying to fit the required model or equation to the data derived from a community sample of preferences for health state description. This model was intended to create time trade off (TTO) weights to yield, given category responses from the QALY questionnaire, defined in clinical terms, utility or preference scores in the range 0 to 1, where 1 is perfect health and 0 is death. It is a failed exercise. The preferences for health states and TTO valuations are ordinal scores for composite health state descriptions. The algorithm to generate utilities or preferences produces only ordinal scores, with utility decrements from perfect health overshooting zero to produce negative values or states worse than death. On all of these counts the classical techniques applied to produce a utility or preference score algorithm creates only ordinal values. This is why the QALY which relies on the impossible mathematical construct of time multiplied by an ordinal score is a failure; it is pseudoscience. If the QALY is a failure in terms of fundamental measurement then so is any proposed or existing measure which fails to recognize the Rasch framework as providing the necessary and sufficient rules for transforming observations to measurement.

It mut be emphasized that the argument rests on the failure to understand, in the assessment of patient report outcomes or observations, that these are not measures. A measure has a to have clear meaning in terms of the standards of fundamental measurement. On these criteria the QALY is not a measure; it is a score or set of numbers that have ordinal characteristics. Further, as detailed in previous commentaries, few grasp the basis on which a QALY is constructed where ordinal preference or utility scores are combined with time spent in a disease state. This results in an impossible mathematical construct. Once this is accepted it points to the case that should be made for fundamental measurement and the application of Rasch standards for instrument development which have been accepted for over 60 years. It is time to move on.

VESTED INTERESTS AND FALSE CLAIMS

One of the most disturbing features of the current HTA belief system is the central role played by the QALY in creating assumption driven modeled imaginary claims. Model claims regularly published by groups such as the Institute for Clinical and Economic Review (ICER) promote the creation on non-evaluable cost-effectiveness claims with associated claims recommendations for pricing and access to new therapies. The principal reason for this is the commitment in HTA by organizations such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), the premier professional group, to the creation of approximate information [6]. As data for new products is typically limited at marketing approval and entry, then the traditional response in HTA is to produce assumption driven cost-effectiveness claims for the lifetime of hypothetical

target populations. These models fail the standards of normal science in facilitating imaginary and non-empirically evaluable claims for products as well as taking no regard for fundamental measurement, in having the QALY as the key component of the model. They are pseudoscience and join intelligent design in the non-science category. Until the QALY is abolished, the pseudoscience of imaginary claims is an attractive and easy option to create value claims.

If for no other reason, the QALY should be abandoned because it facilitates the ability of unscrupulous manufacturers and others to create false claims for cost-effectiveness at a price acceptable to the manufacturer (incidentally, a monopoly price). This is inevitable and impossible to police because the entire exercise is built on assumptions from trials and the literature which may have been deliberately selected and manipulated. There is ample evidence for this activity [7]. All that can be done is to propose alternative assumptions; this, in the US, is never done. In any event, it is a fruitless exercise.

If the QALY fails to be prohibited this would give a signal for paper mills and other actors to put increased effort into assumption driven modeled claims at favorable prices. Presented as an approved ordinal measure, or at least one that has been allowed to be used in HTA by the Congress, the QALY will continue for decades to produce false and imaginary claims. These claims will include not only submission to Federal programs to justify the sponsors preferred pricing as well as formulary submissions to the private health care sector.

NEEDS-BASED DISEASE SPECIFIC INSTRUMENTS

The focus of Rasch measurement is on the individual, not upon the views and prejudices of community sample populations and their preferences for health states, with their eugenic implications. It is not a question of abandoning the concept of quality of life but to assess the extent to which, from the patient perspective, disease and its treatment prevents the fulfillment of basic human needs; how do impairments and disabilities affect need fulfillment and consequently the quality or value of respondents' lives? This is the construct theory of needs-based instrument measures, developed utilizing the Rasch model, to assess the extent to which needs, as articulated from patient interviews, are currently being met and how interventions may impact those needs [8].

The needs based measure for a specific target patient population rests on deriving potential instrument items directly from patients and selecting, using Rasch criteria, those which give the best fit. This yields, as long as the fit is considered satisfactory, an instrument or questionnaire that will produce a unidimensional, linear, interval and invariant measure that will provide a valid basis for evaluating therapy impact. This has been the focus of a significant PRO research program for over 30 years with some 30 disease specific instruments developed. Disease states for which Rasch instruments have been developed include: pulmonary hypertension, Alzheimer's caregivers, psoriasis, atopic dermatitis, Crohn's disease, herpes, multiple sclerosis, migraine, osteoarthritis, rheumatoid arthritis and psoriatic arthritis [9]. The methodology can be readily applied to any other disease state or target patient population to create a unique measurement of needs fulfillment in therapy response. Issues of bias would not arise as the measure is patient centric.

There is, unfortunately, a slight problem. Over 95% of existing disease specific instruments also fail to meet the standards of fundamental measurement. This has been recognized for decades but

conveniently overlooked by clinicians and others developing these instruments; a more obvious reason is that the developers, principally clinicians, are clearly unaware of fundamental measurement standards.

The argument for failure is straightforward. The majority of these disease instruments rely upon scoring and adding integers assigned to ordinal responses on a scale for each questionnaire item (technically a Likert scale: e.g. 1 = never to 5 = all the time). This summation of integers relies on two assumptions being met: first, all items are of equal difficulty and the thresholds between the steps are of equal distance or equal value. These requirements are never considered. This means that if there is an argument that following disallowance of the QALY there is a safe harbor in disease specific instruments then, if we apply the 'similar measure' criteria, these would be equally disallowed. For such measures to be allowed it would have to be shown that they met Rasch criteria for a unidimensional, linear, interval and invariant scale. If the QALY fails to be abolished, then as with the multiattribute QALY, they will continue to be applied in pricing and access decisions.

TRANSITION TO FUNDAMENTAL MEASUREMENT

The current belief system in HTA rests in large part on the commitment over the past 30 to the creation of imaginary cost-effectiveness claims with assumption driven simulations. In this modelling the QALY plays a central role. Outcomes of modeled therapy interventions are expressed ib incremental-cost-per-QALY terms with claims for cost-effectiveness judged by cost-per-QALY thresholds. Clearly, if the QALY is rejected as a measure, then the modeled simulations cease to have any relevance for pricing a product access. Claims for bias in the treatment of disabilities can be put to one side.

Abolishing the QALY means the end of the HTA meme or belief system. More accurately, it points to the conclusion that we have been following a will o'the wisp in our assessment of the benefits of therapy interventions by stage of disease and for target patient populations. This is why passing H.R. 485 is so important: it is the catalyst that will allow us to transform from a failed HTA to a new start where the required evidentiary and measurement standards apply. At the same time, this would put US healthcare in the forefront of standards for meaningful patient centric therapy choice.

The transition to a new HTA framework is quite straightforward; there is no extensive retraining or agreement on guidelines to support formulary submissions and the assessment of new product impact [10]. All value claims for pharmaceutical products and devices, whether these are presented in clinical, PRO or resource utilization terms must rest on premises:

- All value claims must refer to single attributes for defined patient populations that meet the demarcation standards for normal science: they must be credible, evaluable and replicable
- All value claims, notably for patient or caregiver reported outcomes. must be consistent with the limitations imposed by the standards of fundamental measurement: they must be unidimensional with linear, interval and invariance properties
- All value claims must be supported by an agreed protocol detailing how they are to be assessed in a meaningful timeframe

The first premise eliminates assumption driven modelling with imaginary cost-effectiveness claims; the second premise eliminates the QALY and successor attempts to create multiattribute instruments; and the third premise requires all claims to be empirically evaluated and monitored to eliminate false claims and support reproduction of claims in different target patient populations (the test for the invariance of an instrument). These standards for a new approach to HTA are detailed in a recently released Certificate Program from the University of Wyoming *A New Start in Health Technology Assessment* [11] .

QUESTIONS AND ANSWERS

The proposed new paradigm in HTA will not sell itself; a belief system or meme with exceptional transmission fidelity that has been sustained for over 30 years, will not surrender easily. Those advocating abandoning the QALY must be prepared to make a substantive case for a new paradigm.

If we make the not unreasonable assumption that the Act will face substantial opposition and not only in the Senate, then those supporting the legislation must be in a position to make the strongest case possible. This is best accomplished by making the case that passing this Act is the first and necessary step to what we may call a new paradigm in HTA. A framework for evaluating patient needs in health care that relies upon meeting the standards of normal science and fundamental measurement.

The obvious first question is to ask those presenting arguments what is meant by the term 'fundamental measurement' ? Followed by a supplemental question as to why interval and ratio measures are critical in evaluating response to therapy? As to the unique role of fundamental measurement, the answer is that the application of rules, Rasch rules, allows us to transform observations to measures with single attribute, linear, interval and invariant properties.

There will be a question inevitably on the term 'similar measures". The answer is that this is a shorthand for measures that fail the standards for fundamental measurement that the QALY does. A more explicit wording could be QALY and other PRO claims that fail the standards for fundamental measurement'.

An obvious second question is what are these 'standards of normal science. This response is quite clear: All value claims or hypotheses must be proposed in a form that allows empirical evaluation? As an example, claims that a product is cost-effective are imaginary if they are based, as they usually are, on assumption driven modelled simulation from organizations such as the Institute for Clinical and Economic Review (ICER).

The next question should be to ask why the QALY fails these standards? The answer us that the QALY was not designed to be a measure of response to therapy following Rasch rules. It comprises two components: time spent in a disease stage and an ordinal utility or preference score. Certainly, time is a measure, but the time component in the QALY is merely a modelled, imaginary estimate. The other component is a preference or utility score designed to have values in the range 0 – 1 (1=perfect health; 0 = death). Unfortunately, all that has been created is an ordinal score which can take negative values or states worse than death and, hence, negative QALYs.

As a follow up question there may be need to elaborate on why the QALY is an impossible mathematical construct. The answer is that we cannot multiply time, a ratio measure with a true zero, by an ordinal score which has order of observations but an unknown distance between observations or categories.

Will abandoning the QALY lead to a decrease in the number of false HTA claims? It would, until the more astute paper mills and others find a way around, including false clinical trial claims in peer reviewed journals. In the short term the impact would be significant; be aware of self-serving arguments to retain the QALY. Too many people have too much to lose.

Wrapping up, there will be a question on what is meant by a new start in HTA? The answer is quite simple: when we wish to assess value claims for a new product in a target patient population the value claim must meet the standards of normal science and fundamental measurement.

So how, for the hypothetical final question, does need fulfillment fit into this? Again, the response is straightforward: need fulfillment is a construct that supports fundamental or Rasch measures that assess the extent to which disease prevents the fulfillment of human need, defined at the disease or target patient population level. A needs fulfillment value claim asks the extent to which, given patient ability, a new product increases the extent to which more difficult needs are fulfilled. The needs fulfillment instrument or question comprises items derived from patient interviews that yield a unidimensional, linear, interval measure.

CONCLUSIONS

Legislation to abolish the application of the QALY and similarly false measures gives us the opportunity to overcome 30 years of HTA claims that deny the relevance of both the standards of normal science and fundamental measurement. The transition will put to one side physician adjudicated clinical performance criteria in favor of patient centric measures of need fulfillment. This represents a sea change, rejecting multiattribute simulated imaginary cost-effectiveness claims, in favor of measures that support claims that reflect the true concerns or patients and caregivers. This is an opportunity, which comes once in a lifetime, and one that should not be cast aside in favor of pseudoscience.

REFERENCES

[1] Langley PC and McKenna SP. Measurement, modeling and QALYs [version 1; peer review: 2 approved]. *F1000Research* 2020, **9**:1048 (https://doi.org/10.12688/f1000research.25039.1)

[2] Langley P. Will we finally abandon the mathematically impossible QALY with HR 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-V5-2.pdf

[3] Langley P. The presumptive failure of H.R. 485 in the Senate: No one understands the QALY construct or its mathematical impossibility. *Maimon Working Paper No. 5* March 2024 https://maimonresearch.com/wp-content/uploads/2024/03/Maimon-Working-Paper-No.-5-March-2024-V2.pdf

[4] Bond T, Yan Z, Heene M. Applying the Rasch Model: Fundamental measurement in the human sciences. New York: Routledge, 2021

[5] Langley P. Bullshit with Scientific Pretensions: Assumption driven simulated claims in health technology assessment. *Maimon Working Paper No. 20*. October 2023 https://maimonresearch.com/wp-content/uploads/2023/10/Maimon-WP-No-20-October-V3.pdf

[6] Neumann PJ, 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 https://www.valueinhealthjournal.com/article/S1098-3015(17)33891-3/pdf

[7] Xie F, Zhou T. Industry sponsored bias in cost-effectiveness analysis: registry-based analysis. *BMJ.* 2022:377 https://www.bmj.com/content/bmj/377/bmj-2021-069573.full.pdf

[8] McKenna S, Heaney A, Wilburn J. Measurement of patient reported outcomes . 2: Are current measures failing us? *J Med Econ*. 2019;22:6 (523-30) https://www.tandfonline.com/doi/epdf/10.1080/13696998.2018.1560304?needAccess=true

[9] Galen Research, Manchester UK https://www.galen-research.com/measures-database/

[10] Langley P. Nothing to Cheer About: Endorsing Imaginary Economic Evaluations and Value Claims with CHEERS 22 [version 1; peer review: 2 approved]. *F1000Research* 2022, **11**:248 https://doi.org/10.12688/f1000research.109389.1

[11] University of Wyoming. School of Pharmacy. Certificate Program: *A New Start in Health Technology Assessment.* https://www.uwyo.edu/pharmacy/resources/certificate-program-a-new-start-in-health-technology-assessment.html