MAIMON WORKING PAPER No. 11 June 2024

ISPOR AND THE QALY: EVIDENCE IS INVENTED NOT DISCOVERED

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

Abstract

In response to ongoing attempts with HR 485 to eliminate the QALY from value claims ISPOR issued a commentary in defense of the QALY. The tone and content of this defense, the aura of a divine revelation to support the allocation of health care and denial of access to new therapies, must give rise to concern because the view of the QALY is entirely positive with no hint of its failure to meet fundamental measurement standard. That is, all value claims for competing therapies must be for a single attribute with linear, interval and invariant properties. The QALY is an illusory composite metric. Patient reported outcomes (PROs) are a special case requiring the application of the unique Rasch rules to transform observations to measurement. If not, the PRO must be rejected as a metric for therapy response. Denying the QALY means denying the CEA reference case modelling. This is invented evidence; claims driven by assumptions to support simulations which are designed to create non-evaluable imaginary outcomes. Judged by the standards for demarcation, this is non-science. Denying the QALY as a measure, as a mathematical impossibility is the end. Attempts to defend the QALY as a sometimes-useful metric strains credulity. While the QALY construct with its basis in community preferences is certainly discriminatory, the argument is irrelevant as the QALY is a will o'the wisp. The QALY may be ableist in terms of states worse than death but the argument is immaterial as the QALY fails to be a defensible measure. The purpose of this brief note is to make the case that the QALY should certainly be consigned to the outer darkness, not only because it is mathematically impossible construct but because its application to reference case CEA models helps that they are an analytical dead end. We have, in practice, wasted 35 or more years in the belief that evidence is better invented than discovered.

INTRODUCTION

The QALY is an integral, element in the current commitment to cost effectiveness analysis (CEA) in health technology assessment (HTA). If the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), the mother ship of *Value in Health,* was to announce that the QALY was of questionable and unsustainable value, the CEA belief system would collapse together with the current commitment to HTA. The central role of the QALY, indeed its indispensable place in CEA, has been argued in a recent paper by Willke et al, in their attempt to counter claims that the QALY is discriminatory and should not be abandoned out of hand [1]. The argument that is presented here is that the Willke et al case is deeply flawed and that CEA and the QALY have no value and should be abandoned, discrimination is only a subsidiary reason [2]. The principal reason, as demonstrated by the many ISPOR practice guidelines, is the lack of interest in the application of Rasch or fundamental measurement to patient reported outcomes (PROs) as well as a complete disregard for the standards of normal science [3] . As a result ,ISPOR and the HTA meme are irrevocably committed to inventing rather than discovering evidence for therapeutic benefits; the

commitment to what Popper described as the evolution of conjectural or objective knowledge is absent [4].

The fundamental measurement or Rasch PRO standard that is overlooked is that all value claims for a new therapy must be unidimensional with linear, interval and invariant properties [5 6]. Instead, ISPOR is putting its efforts as a professional journal with a global readership behind the QALY and the role of assumption driven modelled simulations that produce imaginary cost effectiveness claims. For ISPOR and its advocacy of CEA, evidence to support health system decisions for competing therapies is invented not discovered.

The commitment to the QALY is not only evidenced by the recently released and widely promoted CHEERS 2022 guidance for facilitating the creation of imaginary modeled claims for journal acceptance, but as noted above, the more Willke et al recent attempted defense by ISPOR of the QALY against charges that it is discriminatory [7]. This is a response by ISPOR to the current ongoing assessment in the Senate of the Protecting Health Care for all Patients Act 2023 (H.R. 485) which proposes prohibiting the use of QALYs and 'similar measures' in coverage and payment determinations under Federal healthcare programs [8 9] .

The pivotal anticipated role of H.R.485 is now been overtaken with discrimination prohibition amendments to the Rehabilitation Act of 1973 which take effect July 8, 2024. These amendments to Section 504 of the Rehabilitation Act have established prohibitions on discrimination by value claims for treatment decisions that ensure the demise of multiattribute ordinal preference scores, such as those created by the fatally flawed EQ-5D-3L instrument and the application of the mathematically impossible QALYs [10]. The Section 504 rule states: there should be no value claim entertained for medical treatment decisions by those that receive Federal financial assistance from the Department of Health and Human Services (HHS) if it is based on biases or stereotypes about individuals with disabilities, judgements that an individual with a disability will be a burden to others, or beliefs that the life of an individual with a disability has less value than the life of a person without a disability. This represents a subtle approach to overturning the current HTA belief system, no doubt inadvertently, by hanging it by its own petard: the application of community preferences for health states to support discrimination in the allocation of health care resources.

THE HEALTH TECHOLOGY ASSESSMENT BELIEF SYSTEM

To understand ISPOR's defense of the QALY and its apparent lack of interest in the standards of normal science and fundamental measurement, a useful starting point is to consider the ISPOR and HTA belief system or meme in relativist terms. This belief system as formalized by ISPOR in its program of good practices for research over the past 20 years, is unique among the sciences and social science: it pays no regard to the standards of normal science and fundamental measurement. Despite these oversights, it has a loyal global following, notably among health economists and social pharmacists, with deeply held inner convictions that their beliefs are true, or as Dawkins puts it: infected by a high transmission mind virus without owing anything to evidence or reason [11]. In the US an ardent advocate is the Institute for Clinical and Economic Review (ICER) which holds to the bizarre belief that health economists are convinced that the ordinal preference scores to support QALYs are ratio measures in disguise [12] . For ICER, its commitment is perhaps better

understood as a key part of its business model: recommendations for pricing and access based on assumption driven reference case simulation models.

To understand the staying power over 35 or more years of the focus on CEA in the HTA belief system a useful starting point is Wootton's critique of relativism in his new history of the scientific revolution [13]. We begin with the mystery of CEA, and the willingness of individuals to believe in the pre-eminence of inventing rather than discovering evidence. Accepting assumption driven simulated reference case models requires denying belief in the standards of normal science and fundamental measurement. This is entirely reasonable if we take a relativist position. In the UK the so-called strong program in science and technology studies argues for a sociological or relativist interpretation of what is described as science: all perspectives are equally valid [13].

This so-called symmetry principle insists that the same sorts of explanation must be given for all types of knowledge claims, whether or not they are successful. We cannot say that one belief is right or even that there is strong evidence for it; this provides CEA or HTA with a claim to relevance although judged by the standards that emerged with the scientific revolution of the 17th with the emphasis on discovery, evidence and measurement, an intellectually barren retrograde claim. Consider, as Wootton notes, the motto of the Royal Society (1662) 'nullius in verba' (take no person's word for it) [13]. A position that is the antithesis of CEA modeled claims and the insistence that we take the modeler's word for claims access and pricing in therapeutic decisions based on invented evidence. The spreading of the QALY belief and the acceptance of invented evidence has reached Biblical proportions with a continuing flood of QALY references; over 26,000 currently on PubMed.

But in reference to CEA the strong program insists that science is not a program to come to grips with reality*: evidence is never discovered it is always constructed within a particular social community….success of a of a scientific research program thus depends not on its ability to generate new knowledge but on its ability to mobilize the support of a community.* The discovery of new evidence is not what we should be concerned with; science is only about reason, persuasion and authority. This is the only possible common denominator between various belief systems where science and non-science take equal billing. Demarcation and a commitment to falsification of value claims has no place; there is no systematic test of experience or the experimental method. Attending ISPOR conferences can be eye opening in the staged commitment to the authority of the QALY and CEA belief system.

The question of interest is whether CEA and HTA fit into this relativist position where truth is consensus driven by rhetoric and authority. The parallels are compelling; all believers agree that the QALY is the exemplar and CEA the basis for accepting imaginary claims driving health system decisions for new therapies. To say it is at odds with standards and appeals to evidence are brushed aside: facts are invented not discovered. There is no awareness of the scientific method or the standards of fundamental measurement to support assessment of value claims. If evidence is invented it cannot be challenged, other than through changing CEA model assumptions and inventing a revised invented evidence outcome. This is clearly absurd: without appeals to superior evidence, the discovery of new facts to challenge the old, progress is impossible [13] .

THE FUNDAMENTAL UNFORCED ERROR

Since its introduction some 50 years the QALY has been endorsed by a succession of global expert groups who have reaffirmed its critical role as a generic metric to support health care resource allocation. Apparently, no one amongst these disparate groups ever thought about the standards of fundamental measurement that had been accepted in the sciences since the scientific revolution of the 17th century. This lack of awareness seems to be a characteristic of CEA and may be one factor in its continued acceptance.

The fundamental unforced error was to let the data have primacy where the data inputs were composite health state bundles valued by the community [6]. These were valued using the time trade off (TTO) procedure that yielded ordinal observations with both negative and positive values. These TTO values were modeled and tweaked to give the best fit and preference algorithms developed to yield ordinal preference scores. There appeared to be no thought that these algorithms had to yield interval scores. This would be impossible in any event as the algorithms are multiattribute. The ordinal preference scores were capped at unity (perfect health) with utility decrements for more 'adverse' bundles and hence no fixed lower bound. This yielded negative scores or 'states worse than death', with novel religious implications. As these were ordinal scores, decrements were, of course, disallowed. The problem here is subtraction; as it is disallowed for ordinal scores, states worse than death are impossible.

Unlike item response theory and traditional statistical analysis where the data have primacy, the unique contribution of the Rasch rules, developed in the 1950s, is the selection of items for a final questionnaire that fit the Rasch model [5 6]. If the results indicate an appropriate fit, then we can claim that the instrument has captured a single attribute with linear, interval and invariance properties. The Rasch framework is unique: it is the necessary ad sufficient condition to transform observations to measurement for patient centric measurement of therapy response [14].

In place for over 70 years ago and used globally, the Rasch model is patient centric and disease specific [15]. As it supports single attribute PRO value claims it avoids the opprobrium that now attaches the CEA reference case with their imaginary non-evaluable claims with QALY options for discrimination. Those developing multiattribute instruments had no apparent concept of fundamental measurement[16] . If you want to deconstruct time to create a perfect health equivalent then the tool applied must meet Rasch measurement standards. The algorithms that support multiattribute preferences produce nothing more than ordinal scores which is even a misnomer as they are composite constructs. In any event, they are not single attribute linear, interval and invariant measures which can support by further transformation a ratio measure. There was no awareness of the established Rasch rules for creating PRO measures and the impossibility of multiattribute metrics. A lack of awareness that continues.

INVENTING LIFETIME EVIDENCE

Those decision makers who take seriously the QALY and CEA cost-per-QALY claims to support price negotiations and reimbursement, should recognize that the cost-per QALY claims are based on assumption driven lifetime simulation models that fail the standards for models in normal science: a model should create high information content value claims that are falsifiable in a

meaningful time frame not imaginary outcomes. These so-called QALY reference case models are designed to create imaginary claims for cost-effectiveness; essentially fairy stories with the QALY as the good fairy, and objective knowledge as the bad fairy. This failure to meet demarcation criteria means the reference case models join intelligent design and astrology in the non-science camp. It is perhaps unusual for health system decision makers in the US to rely on imaginary cost-effectiveness claims but the audience seems mostly unaware of the standards for normal science and fundamental measurement, let alone Rasch rules for PROs. These standards and rules are certainly not part of typical CEA curricula, discussed in HTA textbooks or referenced in ISPOR practice guidelines where CEA and the invention of evidence is supreme [17] [18].

Criticism is absent. All too many graduate with none or only a limited awareness of standards for normal science and fundamental measurement. This reinforces the transmission fidelity of the current belief system and the acceptance of invented evidence. It also ensures a continuing attachment to multiattribute instruments and the search for the Holy Grail single metric to allocate health care resources.

Of course, it is understandable for health economists and decision makers to yearn for such a metric; it makes the central planning of allowance/disallowance of therapeutic support so much easier and, more importantly even, justifiable politically. An allocative appeal to a higher measurement authority such as the National Institute for Health and Care Excellent (NICE) in the UK that allows us to brush aside eugenic criticisms or unpopular denial of drug coverage. The appeal of a false science. After all, creationism is a staple belief for a sizable proportion of the US population.

Unfortunately, the Holy Grail is a Pythonesque not an unattainable goal; the QALY is a total failure from the standards of fundamental measurement. If there is to be a true preference score then this score should be a Rasch ratio scale for a single attribute that is linear, interval and invariant [19]. That is, it should be a proportion which ranges from zero to unity or an interval scale with a true zero. There must be a meaningful relative distance between values on the scale; we can compare the ratios of scale values as there is a true zero. Uless this is made clear from the outset the result is instruments such as the EQ-5D-3/5L that fail to meet the standards for fundamental measurement in the insistence on community valuation or ordinal preference scores, both positive and negative, for arbitrarily defined health states. If this is all that is attainable with CEA then we should consign it to the recycle bin of the philosophy of science.

COMMUNITY PREFERENCES AND STATES WORSE THAN DEATH

Central to the concept of measurement is to measure one attribute at a time. A measure is invariant for a single attribute *where the readings will remain invariant across all suitable contexts; and for any one context, all suitably calibrated devices will yield invariant readings* [6]. Composite measures must be avoided. The outcome is pre-ordained. Community preferences for bundles of health state descriptions not only fail Rasch standards for fundamental measurement but provide a modern twist on false eugenic discrimination. Community preferences for one health state over another defines discrimination. It cannot be avoided although the ISPOR Willke et al defense, of the indefensible, argues that if properly adjusted any discriminatory claim can be avoided.

But the issue of ISPOR and discrimination is more concerning: the avoidance in this defense of the QALY of any mention of states worse than death. There is no reference to a recent paper, whether by oversight or design, which labels the QALY ableist, examining the unethical implications of health states worse than death [20]. In other words, extending the lives of people in these community determined negative QALY health states will reduce overall population health even if the medical cost of this is zero. This argument could be extended to any group with an ordinal preference or QALY score lower than a pre-assigned cut-off. Looking at how certain regimes have treated patients with disabilities point to the QALY as not only being ableist but potentially lethal. Just as the ISPOR defense argues that we could adjust the QALY implications to avoid discrimination so we could adjust to apply a discriminatory cutoff more widely to capture a higher proportion of a population for denial of health care.

Putting the metric impossibility to one side, care has to be taken in making the case against the QALY preferences for states worse than dead. The numbers vary by instrument. A recent review of values for the UK found that for the EQ-5D-3L, 34.6% of health stats had negative scores (range 1 to -0.594) while for the EQ-5D-5L 5.1% of health states (range 1 to -0.285) were states worse than death [21] . Note, however, that the EQ-5D-3L produces 243 health states with 84 worse than death while the EQ-5D-5L produces 3,125 heath states (yes, really) with 159 worse than death. As the preference scores are ordinal, we can rank these respondent states but only report non-parametric statistics such as medians or modes. There can be no such animal as an average score which hides the negative preference scores.

This does not mean that these proportions will always be observed. Depending on the distribution of health states in a population the result could be no respondent with a state worse than death or others where all respondents are in that category; presumably to the surprise of the respondents themselves when given the individual survey results. Would they be denied insurance cover?

Whatever techniques are used to create multiattribute scores the fact that, apart from a multiattribute score being illusory and lacking construct validity, the preference score is achieved by decrements from unity (the community definition of perfect health). This means, inevitably, that whatever techniques are used for the scores they will undershoot or overshoot zero (the community definition of death). This is true whatever convoluted mathematical framework is employed. There is no way that there is by construct a true zero, let alone score with interval, linear and interval properties; a ratio measure in the range 0 – 1.

The problem for the multiattribute advocates and their fixation on the concept of a QALY is that it started out and continued, in ignorance of fundamental measurement, seeking a Holy Grail metric for allocating health care resources following community preferences for health states (few of which the community valuers had ever experienced). From the perspective of fundamental measurement, valuing health states is not only a waste of time but an analytical dead end; made more confusing by the variety of multiattribute instrument currently in play and contradicting each other so that health decision makers have a menu of impossible QALYs to choose from. The current debate over the EQ-5D-3L versus the EQ-5D-5L which is now in its fifteenth year is a salutary warning. The focus should be on the patient and the assessment of single attributes. The Rasch rules as they are applied to patient groups are, by definition, non-discriminatory. One group

is not being compared to another (with different QALY metrics). The QALY is designed to compare groups. As such it is discriminatory, and quite correctly falls foul of Section 504 prohibitions. It ensures the patient is in second place. Not surprising when patients who are in health state worse than death are asked to comment, the response is that they consider their lives worth living; until, presumably, the QALY police end them .

ABSENCE OF AGE DISCRIMINATION

To illustrate the lack of awareness of fundamental measurement and the constraints it imposes on statistical evaluations, it is important to remember that we can only perform parametric assessments if the data of interest are unidimensional, interval, linear and invariant, a recent paper by Xie et al (referenced and described by ISPOR) is instructive [22]. The purpose of this analysis was to assess whether there was any evidence for QALY age discrimination in published cost-effectiveness analyses published between 1976 and 2021. The first point to note is that the authors selected the published studies from the Tufts CEA database which makes no assessment of the fundamental measurement properties of the listed studies. Second, there are some 10,000 CEA imaginary claims that have been dissected an entered into the database over some 50 years; none meet the standards for interval measurement. Third, the study analyses some 4,445 CEAs to assess the distribution of the incremental cost-effectiveness ratios for those modeled respondents under 65 years and 65 years and over. As such the study has no relevance to support the absence of age discrimination because the CEA models fail both the standards of normal science and fundamental measurement. The latter means that any statistical analysis is disallowed, even if focused on imaginary claims.

A RIPPLE EFFECT: CASUALTIES

The demise of the QALY as a proposed measure for therapy response will have, like a house of cards, a ripple effect for those advocating a suite of item supplements to the QALY to make it more socially palatable. The first casualty is the EQ-Health and Wellbeing (EQ-HWB) Index [23]. Some 10 years in development it is now a Pythonesque monument to efforts to revisit community preferences while failing to understand the standards of fundamental measurement . Yet assessments of the EQ-HWB continue, with claims, particularly for the 9-item short form, as a useful metric. Unfortunately, in its design as a discriminatory tool, an add on to the EQ-5D-3L, it will, at least in the US, fall foul of the discriminatory prohibitions of the Section 504 of the Rehabilitation Act of 1973. It has nowhere to go.

The aspect that is surprising is that those developing the EQ-HWB appear to have no notion of the standards for fundamental measurement in evaluating therapy response; they are an illusion. Composite scores are disallowed. Any instrument must have (once again) unidimensional, linear, interval and invariant properties). These are not an afterthought but an integral part of instrument development. After 10 years of development we are still left with a composite score for both variants of the EQ-WB which is ordinal. It cannot support QALYs. The entire exercise should have been abandoned as soon as first proposed. The resources devoted to this could have been more productively spent in developing Rasch standard disease specific measures which would give a defensible estimate of therapy response.

Despite attempting to bolster the case for the QALY with some 42 references some going back decades, all this defense of the QALY demonstrates is a pervading failure in health technology assessment to come to grips with fundamental measurement, let alone standards of normal science. Unless response to therapy is captured by PRO instruments that follow the Rasch rules in their development, then no PRO claim is valid. Rather than a concept of value that takes a health economics (i.e., clinical) perspective we must look at value from the patient or caregiver perspective. After all, attempting to value from a health economics perspective has been a failure with community preferences driving imaginary modeled discrimination.

Further casualties are those in the 'value of hope' camp who look to further QALY enhancement by tacking on 'novel' elements. The ISPOR value flower has wilted. To save face, the easy way out, of course, is to cling tenaciously to the QALY arguing that whether or not it is a valid measure, the debate continues, possibly indefinitely, and we must be restrained in any attempt to undermine its authority [24]. Fortunately, the debate is over; it should never have been entertained in the first place. The impossible QALY has no authority.

PATIENT VALUE AND NEED FULFILLMENT

If we are concerned with the benefits of a new therapy then we should not be asking the community to value the benefit but address the patients, in terms of their needs and concerns directly. This a challenge: to find a non-Pythonesque measure that meets the measurement standards for a ratio measure with units as proportions. This is achievable with Rasch measurement: creating a unidimensional, linear, interval and invariant measure and transforming it to a ratio measure with the focus on need fulfillment [25].

The starting point in Rasch PRO modelling is to identify a latent construct of interest [5]. The next step is to consider what manifestation of that latent construct can be observed and translated to a measure that meets Rasch standards. In the Rasch PRO instruments that have been proposed to date the latent construct is quality of life with the manifested attribute of interest the extent to which the needs of the target patient population are met. Need fulfillment quality of life emerged in the 1990s. It is not health related quality of life which categorizes health status rather than assessments of patient value, where the object is to maximize the patient value of every dollar spent. The focus shifts from measuring physical attributes of health status to direct measures of patient value which will vary across disease status and target patient populations. Thus, *the needs model hypothesizes that the value of individual lives is dependent on the extent to which their human needs are fulfilled* [24] . Disease and its treatment are, particularly in chronic disease, the major influence on need fulfillment. But it is impossible to separate clinical factors that are present in, say, multiattribute instruments, from additional factors such as supporting social and family care.

The focus of needs fulfillment instrument design, identifying items for needs measures, is on patient or caregiver interviews. Questions are framed around how the life of the patient has been affected by the disease in question, with questions relating to functional limitations and their effect on respondents [24]. Detailed assessment of responses yields a preliminary list of needs, ranked in order of importance or difficulty, for the target interviewee group. Assessments of face and content

validity with a large sample establishes a final item set (usually 30 items or less) to establish reliability and validity. In need fulfillment, as with other RASCH PRO instruments, there is an important focus: the Rash model is probabilistic [4] [5]. The likelihood that a respondent will, following a new therapy, meet more of the articulated needs, expressed in terms of relative difficulty, will be a function of the difference between item difficulty and patient ability . This conjoint interaction is completely absent from virtually all disease specific PROs. This process is supported by access to Rasch software to establish a final item set that is unidimensional, linear, interval and invariant. Since the 1990s some 30 disease specific needs fulfillment measures have been developed [26].

CONCLUSIONS

Asking the community sample to value health states of which few have any experience in clinical terms has unfortunate eugenic implications. Those with the least valued health states can be discarded, refused health care, and by the logic of the QALY the overall health status is raised. The QALY is the ultimate in population measurement; it means all things to all persons as long as the appropriate adjustments are made to avoid claims for discrimination. This is hardly a decisive argument for the QALY. The QALY can be applied to discriminate; this is precisely why Section 504 is focused on disability prohibitions. If discrimination is optional, it can always be possible.

In this respect, it is not surprising that the ISPOR position is that it *is an imperfect tool (an understatement) for measuring health benefits as an input to healthcare decision making. to population health-care .... and with appropriate use it will not be discriminatory* [1]. Rather than banning the QALY and depriving manufacturers and others of an important tool to invent the benefits and harms of a treatment the truth is it is a tool that should never have been developed in the first place: it fails to meet the standards of fundamental measurement. When put in the context of the unnecessary baggage of CEA reference case models, the incremental cost-per-QALY claims are just meaningless. They are imaginary, lacking replication and the possibility of falsification. It is perhaps unusual for health care decisions to be based on invented imaginary claims and undefined appropriate use. Unless CEA can adjust to the required standards then it should be abandoned as an unfortunate and misleading anachronism in the history of science, together with the QALY..

There is a fundamental disconnect between CEA and meeting patient needs. CEA is rightly rejected because it is discriminatory. Cost per QALY is not a starting point, as ISPOR practice guidelines insist, but an end point. . It cannot inform decision makers because of its failure to meet standards. for normal science and measurement. Reference case modeling is a wasted effort; imaginary non-evaluable outcomes should have no role in health care decisions.

If community weighted preferences for composite health state bundles fail standards for fundamental measurement, then why continue to extol their virtues? One reason is the obsession with attempting to attain the Pythonesque Holy Grail. A mystery metric which can be accepted even with its manifest measurement failures. This is a contradiction in terms. The quest is futile. Expert opinion may insist on a generic, universal metric but this is just a reflection of a lack of knowledge or even interest in fundamental measurement by such experts. Of course, if you are in

a hole, the sensible response is to stop digging. In the case of the QALY and CEA the ISPOR response is quite the opposite: digging has a unique role and a rewarding future if the failure to meet the standards of normal science and fundamental measurement are put to one side.

We need a new start in HTA. To this end, a recently released on-line Certificate Program from the University of Wyoming (see below) offers a way forward consistent with the standards of normal science and fundamental measurement. Evidence is not invented.

UNIVERSITY OF WYOMING CERTIFICATE PROGRAM

A NEW START IN HEALTH TECHNOLOGY ASSESSMENT

For those who are interested in following up the arguments presented here for Rasch standard patient centric value claims, the recently released on-line University of Wyoming Certificate Program: A New Start in Health Technology Assessment is recommended.

The Certificate Program is in three parts:

- Part I: Required evidentiary standards for product and therapy assessment
- Part II: The failure of approximate modelled information for therapy decisions
- Part III: Formulary submission value claims and protocols for a new start in product evaluation in health system management

The Certificate Program package includes extensive notes (overall for the 14 modules 85,000 words), audiovisual presentations and a short true-false and multiple-choice assessment for each module. The cost of the Certificate Program is $875 USD with 20.5 hours of ACPE credit. For those who do not need ACPE accreditation, the University of Wyoming will provide a Certificate of Completion. Following interest already expressed, for those introducing the proposed new start standards for technology assessment there will be a program of one- and two-day workshops and on-line seminars to support course development and alternative program structures to meet local needs. There will also be a series of working papers to explore specific aspects of the new start program.

The link to register in the Certificate Program is:
https://www.uwyo.edu/pharmacy/resources/certificate-program-a-new-start-in-healthtechnology-assessment.html

The Certificate Program was developed by Dr Paul C Langley, a health economist. Dr Langley is currently resident in Tucson, Arizona. If further information on program content is required feel free to contact on langleylapaloma@gmail.com

REFERENCES

[1] Willke R, Pizzi L, Rand L et al. The Value of Quality Adjusted Life Years. *ValueHealth*. 2024;27(6):702-05.

[2] 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

[3] Langley P. Are the practice guidelines of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) fit for purpose. *Maimon Working Papers*. No. 9 June 2024 https://maimonresearch.com/wp-content/uploads/2024/06/MAIMON-WORKING-PAPER-No-9-2024-V5-1.pdf

[4] Popper K. Objective Knowledge: An Evolutionary Approach (Rev. Ed.). Oxford: Clarendon Press, 1979

[5] Andrich D, Marais I. A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences. Singapore: Springer, 2019

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

[7] 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)

[8] H.R.485 - Protecting Health Care for All Patients Act of 2023. 118th Congress (2023-2024) https://www.congress.gov/bill/118th-congress/house-bill/485

[9] 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

[10] US Government. Health and Human Services Department. Nondiscrimination on the Basis of Disability in Programs or Activities Receiving Federal Financial Assistance: Final Rule. Federal Register May 9, 2020. 89 FR 40066 https://www.federalregister.gov/public-inspection/2024-09237/nondiscrimination-on-the-basis-of-disability-in-programs-or-activities-receiving-federal-financial

[11] Dawkins R. A Devil's Chaplain. New York: Houghton Miflin, 2004

[12] Langley P. Peter Rabbit is a Badger in Disguise: Deconstructing the Belief System of the Institute for Clinical and Economic Review in Health Technology Assessment. *InovPharm*. 2021; 12(2): No.20

[13] Wootton D. The Invention of Science: A New History of the Scientific Revolution. New York: Harper Collins, 2015

[14] Wright B, Linacre J. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989; 70(12):857-60

[15] Rasch, G. Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielsen & Lydiche, 1960.

[16] McKenna S, Heaney A. Composite outcome measures in clinical research: the triumph of illusion over reality. *J Med Econ*.2020;23(19):1196-1204

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

[18] Garrison L, Neumann P, Willke R et al. A Health Economics Approach to US Value Assessment Frameworks – Summary and Recommendations of the ISPOR Special Task Force Report [7]. *ValueHealth*. 2018; 21:161-65

[19] Langley P. McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*.2021;12(2):No. 6

[20] Schneider P., The QALY is ableist: on the unethical implications of health states worse than dead. *Qual Life Res. 2022*; 31:1545-52

[21] Mulhern B, Feng Y, Shah K et al. Comparing the EQ-5D-3L and English EQ-5D-5L value sets. *Pharmacoeconomics*. 2018;36:699-713

[22] Xie F, Zhou T, Humphries B et al. Do QALYs discriminate against the elderly? An empirical analysis of published cost-effectiveness analyses. *ValueHealth*. 2024;27(6):706-12

[23] Brazier J, Peasgood T, Mukuria O et al. The EQ-HWB: Overview of the Development of a Measure of Health and Wellbeing. *ValueHealt*h. 2022;25(4):482-91

[24] N Devlin, Drummond M, Mullins D. Quality Adjusted Life Years, Quality Adjusted Life Year Like Measures, or Neither: The Debate Continues. *ValueHealth*. 2024;27(6):689-91

[25] McKenna S, Wilburn J. Patient Value: Its nature, measurement and role in real world evidence studies and outcomes based reimbursement. *J Med Econ*. 2018;21(5):474-80

[26] Galen Research UK. Measures Database https://www.galen-research.com/measures-database/