

**MAIMON WORKING PAPER No. 7 MAY 2024****THE UNITED STATES OUTLAWS MULTIATTRIBUTE PREFERENCE SCORES AND THE QALY**

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

**Abstract**

*After almost 40 years the quality adjusted life year (QALY) has finally been laid to rest in the US; effective July 8, 2024. Surprisingly, it is not because of critics pointing to its manifest deficiencies, its failure to meet the standards of fundamental measurement, but on its intended purpose: to discriminate between individuals on the basis of their health status as determined by community preferences. Amendments to Section 504 of the Rehabilitation Act of 1973 have established prohibitions on discrimination with value claims for treatment decisions that ensure the demise of multiattribute preference scores, such as those created by the EQ-5D-3L instrument and the creation of mathematically impossible QALYs. 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. Multiattribute instruments with the focus on the efficient allocation of health care resources meet these criteria for discrimination as they represent community preferences for access to and denial of care as well as assigning a significant proportion of respondents to states worse than death. The purpose of this obituary for the QALY is to consider the importance of fundamental or Rasch measurement in establishing required standards for value claims and the creation of disease and target patient population specific instruments to capture patient reported or patient centric outcomes. The latent construct of interest is quality of life expressed in terms of need fulfillment. With the passing of ordinal composite preference scores and QALYs, multiattribute instruments are dismissed together with assumption driven modelled simulations for imaginary cost-effectiveness claims. Ongoing Pythonesque attempts to create a Holy Grail composite generic metric to guide health care resource allocation are a wasted effort. The obituary concludes with recommendations for a new start in health technology assessment, a new paradigm, one that respects the standards for normal science and fundamental measurement that puts the concerns of patients and caregivers front and center.*

**INTRODUCTION**

History is replete with beliefs and applications that, seen in retrospect, are nonsensical or just plain silly. Health technology assessment is no different. Effective July 8, 2024 composite multiattribute value claims and QALY constructs, are effectively outlawed in the US. Amendments to Section 504 of the Rehabilitation Act of 1973 have established prohibitions on discrimination with value claims for treatment decisions that ensure the demise of multiattribute preference scores, such as those created by the EQ-5D-3L instrument and the creation of mathematically impossible QALYs<sup>1</sup>. 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 health technology belief system, no doubt inadvertently, by hanging it by its own petard: the inadvisability of applying community preferences for health states to discriminate in the allocation of health care resources. This is no doubt a valid approach for those with a social welfare perspective, where public preferences trump individual health conditions and concerns with its unfortunate eugenic implications<sup>2</sup>. These are seen in the fact that the application of community preference weights results in significant proportions of respondents to multiattribute health status instruments being assigned to states worse than death. Yet no thought was or continues to be given to the manifest deficiencies of the approach<sup>3</sup>.

The Section 504 rules have also ensured that there is no extended ‘end of life’ debate over the measurement status of multiattribute instruments, the creation of preference algorithms, the merit of composite ordinal preference scores and a defense of the QALY on the grounds that health economists have confidence that the composite preference scores and hence QALYs are truly interval or even ratio measures. Such a debate should never occur once the various contributors to the debate accept the long recognized fundamental or Rasch standards for measurement: observations are not measurement and all measures must be for single or unidimensional attributes with the measure having linear, interval and invariant properties<sup>4</sup>. In the case of patient reported outcome (PROs) the Rasch rules for transforming observations to measurement are unique, they are necessary and sufficient.

The Section 504 amendments to the Rehabilitation Act neatly sidestep a protracted debate over fundamental or Rasch measurement which, to be honest, few in health care decision making would understand, let alone professionals in health technology assessment. This is evidenced in the recent debate over HR 485 in the House of Representatives, the Protecting Health Care for all Patients Act 2023 which proposed prohibiting the use of QALYs and ‘similar measures’ [undefined] in coverage and payment determinations under Federal healthcare programs. It just passed the House (211 vs 208) and is now in the Senate (Committee on Finance). Objections to the Act by the Democrats, who saw it as an attack on the Affordable Care Act, clearly had no idea that the QALY was an impossible construct; a position shared with the Republicans, in kicking a long dead sheep. The Section 504 amendments, as detailed here, mean the passage of HR 485 in the Senate, viewed as uncertain with a Democrat majority, is unnecessary

At long last, the QALY in the US has no part as a gatekeeper in pricing, access and allocative health care decisions. While this is long overdue, it is important to consider key issues related to fundamental measurement and the application of the QALY to create imaginary, false and non-evaluable claims with reference case assumption driven simulation models. This sets the stage for a discussion of the standards for patient-centric value claims as the necessary next step fill the value claims vacuum created by the demise of the ordinal utility or preference scores and the QALY.

## FUNDAMENTAL MEASUREMENT, ORDINAL SCORES AND QALYS

Statistical analysis dominates the social sciences to the almost complete exclusion of the concept of measurement. What is overlooked is that measurement must precede statistical analysis; unless a scale has fundamental measurement properties standard statistical analysis is invalid. Importantly, while mention may be made of Stevens' four levels of nominal, ordinal interval and ratio scales, there is no discussion of measurement, particularly in non-physical attributes<sup>5</sup>. The seminal contributions are Rasch in the 1950s and Luce and Tukey in 1964 with in their concept of simultaneous conjoint fundamental measurement as a new type of fundamental measurement that subsumed existing measurement categories for physical objects and derived measures with a framework for detecting measurement structures in non-physical constructs<sup>6 7</sup>.

Composite or multiattribute utility or preference scores are not measures<sup>8</sup>. For a measure to be meaningful it must be unidimensional; referring to a single attribute. Health states as multiattribute constructs are disallowed because latent constructs are by definition unidimensional (e.g., temperature, pain, quality of life). The items in an instrument must measure only a single construct. If this has not been demonstrated the total or composite score has no meaning as a measure; it is merely an ordinal scale lacking construct validity.

Certainly, classification and seriation are important, but this is not measurement. The requirement that is overlooked is that the distinctive feature of lineal measurement must be an *arbitrary yet meaningful unit of difference that can be iterated between successive lengths*. This is the function of Rasch measurement. We cannot assume that a series of numbers have fixed unit interval properties; this has to be demonstrated in the construction of measures. For mathematics achievement, cognitive development or quality of life, to give some examples, we need to translate numbers or observations into a fundamental measure<sup>9</sup>. This can only be achieved if we apply a set of rules that map from observations to a numerical structure such that the relationships among numbers reflect the relationships among the property of the objects being measured. Rasch measurement is unique because it provides the necessary and sufficient set of rules to translate observations to measurement. Unless we can demonstrate that those rules have been applied, we cannot assume that we have a meaningful measure of, for example, an attribute of the latent construct we may call quality of life, that meets the required fundamental measurement standards of unidimensionality with linear, interval and, invariant properties. A further question can then be addressed: translating an interval measure, which can take both positive and negative values, to a ratio measure with a true zero capped at unity. This is possible only after we have constructed, following Rasch rules, an interval measure.

The Rasch model offers a further critical feature that is central to translating observations to measurement in the social sciences: a probabilistic formulation of the relationship between the difficulty of an item selected for a questionnaire and the ability of the respondent to succeed on that item. In Rasch terms the probability of success is a function of the difference between the ability of the person and the difficulty of the item. This is a key element in patient-centric measures in health technology assessment. To measure response to therapy on a unidimensional, linear, interval and invariant scale, items have to be ranked by difficulty, set against a distribution of respondent ability or a random sample of the target patient population.

It is not the purpose of Rasch measurement to replace the application of standard statistical techniques but to produce genuine interval measures that can support those techniques. Traditionally in the social sciences the observed data have primacy and the focus is on fitting a model to the data: the model is exploratory and descriptive with, often ad hoc, attempts to account for all the data. Against this, the Rasch model has primacy. The model is confirmatory and predictive as the data have to fit the model. Focusing on the size and structure of residuals, the choice and ranking of items are central to the application of probabilistic conjoint measurement, indicating whether or not the results can justify the claim that we have a measure with invariant, interval and invariant measurement properties. If this can be justified then we have an instrument, a set of ranked items, that capture response to therapy through the sum of successful integer responses.

If the standards of fundamental or Rasch measurement had been recognized when multiattribute instruments had been first proposed to value health states then the proposal should have immediately been rejected. If we want to make a value claim then, in the first instance, the claim should be unidimensional, linear, interval, invariant and open to empirical evaluation. In the case of multiattribute instruments such as the EQ-5D-3L, the algorithm to create a score yields only a composite ordinal scale. This means the scale is incapable of supporting standard arithmetic operations. All the scale provides is a ranking of responses. Applying the utility or preference scores to support quality adjusted life years (QALYs) is disallowed; the QALY is therefore, as noted above, an impossible mathematical construct. Attempting to create incremental cost per QALY value claims is simply a waste of time.

The lack of awareness of fundamental or Rasch measurement has a further critical implication: the translation from observations to measurement for instrument or item responses that are patient-centric. Community valuations of health states has no part in the Rasch model. The focus is on the patient perspective. This has obvious implications for target patient groups with disability. In Rasch terms, where the objective is to create value claims that are patient-centric, there is no place for multiattribute instruments; nor for community valuations of health states. All instruments must be for a target patient population, reflecting the concerns of that population. Hence the term patient-centric.

Once the imperative of patient-centricity is recognized, any misguided attempts to create generic instruments must be avoided. Multiattribute instruments, the EQ-5D-3L is the most popular, are designed to discriminate between individuals in terms of their health state. Community preferences for, usually limited, health state descriptions are the basis for discriminatory decisions: allowance versus disallowance. We can think of multiattribute instruments as central planning tools for discrimination, applying not only to those with disabilities, but across the spectrum of health care experience. Tools that are necessary if the focus is on efficient health care allocation. The concerns of patients are immaterial and fundamental measurement requirements are avoided in creating composite scores.

## **DISCRIMINATION: THE 504 AMENDMENTS**

The amendments are quite specific when it comes to denial of access to medical treatment (§ 84.56): no qualified individual with a disability shall, on the basis of disability, be subjected to discrimination in medical treatment under any program or activity that receives Federal financial assistance, including in the allocation or withdrawal of any good, benefit, or service. This general prohibition includes three specific prohibitions for denial of treatment: (i) bias or stereotypes about the patient's disability; (ii) judgements that the individual will be a burden on others due to their disability; and (iii) a belief that the life of a person with a disability has lesser value than the life of a person without disability, or that life with a disability is not worth living.

Translating these prohibitions to value claims for therapies and other interventions for target patient groups, means that any instrument (e.g., the EQ-5D-3L) that discriminates between health states as a basis for the access to and denial of medical care is prohibited. But this is precisely what these various multiattribute, generic instruments are intended to do. The community valuation of health states yields criteria, utility or preference scores, to justify or deny access to therapies. There is no guarantee that the more adverse the health state valuation that the individual or target patient group will be allocated resources on the same basis as those with less adverse health states as judged by community preferences. The Section 504 amendments to the Rehabilitation Act are, therefore, the legislative standards for outlawing composite utility or preference scores, and the QALY. This holds irrespective of views on the failure of composite scores to meet the required standards for Rasch or fundamental measurement. In this context, an obvious implication is to disallow any metric for value claims which is based upon community valuation of health states.

These prohibitions do not deny the application of value claims to support approval for new therapy interventions and the circumstances for that approval. Issues such as the quality of care and cost containment are relevant but only if the measures applied to support approval and access are non-discriminatory. In terms of prohibition (i) above, the various multiattribute generic instruments are clearly stereotypical in attempting to apply a minimalist global health state description (e.g., five health dimensions and three problem levels for the EQ-5D-3L) to classify winners from losers. In respect of prohibition (ii) value claims based on these instruments are intended to measure the community valuation of the burden of disease, assuming that a single metric is sufficient to distinguish a more burdensome disease state from one which is closer to the defined perfect health state, ranking health states in the process (243 health states for the EQ-5D-3L and 3,125 health states for the EQ-5D-5L). Finally, in respect of prohibition (iii) these instruments yield negative scores which, as considered in detail below, is interpreted as the community valuation of a health state as worse than death, with the implication that this lesser value denotes a life not worth living. A conclusion that is not, apparently, shared by patients who are assigned to that category <sup>10</sup>.

## **THE REDUNDANCY OF HR 485**

Introduced to the House of Representatives, HR 485 Protecting Health Care for all Patients Act 2023, was proposed to amend Title XI of the Social Security Act to prohibit the use of QALYs and 'similar measures' [undefined] in coverage and payment determinations under Federal healthcare programs. While this at the time was long overdue, it was pointed out that those proposing the HR

485 (Republicans) and those opposing (Democrats) had little idea of the limitations on the QALY and, more broadly, the concept of fundamental measurement. For the Democrats it was a gut reaction with HR 485 seen as an assault on the Affordable Care Act set against a genuine concern by Republicans for unequal treatment of patients where the QALY is a metric. Further confusion was evidenced by the use of the term ‘similar measures’ which no one could define or give examples.

It seems clear from the amendments agreed for Section 504 of the Rehabilitation Act of 1973, that fundamental measurement played no part in the development of the criteria. This is unsurprising as the belief system that has support health technology assessment over the past 30 years, in the emphasis on assumption driven modelled simulations with non-evaluable value claims, is similarly bereft of an awareness of fundamental measurement. A lack of awareness which has also characterized the HR 485 debate, where neither party, as noted, has any idea of the limitations imposed by fundamental measurement. If that had not been the case then we could have put to one side the question of disability bias in favor of an appeal to the standards of fundamental measurement.

Somewhat surprisingly, the House Republicans voted in favor (211 to 208) with HR 485 sent to the Senate where it will likely fail as the Democrats have the majority. With the amendments to the Disabilities Act of 1973, HR 485 in the Senate is redundant legislation.

## **ABANDONING MULTIATTRIBUTE INSTRUMENTS**

The Section 504 specific prohibitions, as has been emphasized, clearly derail the application of the EQ-5D-3L or other multiattribute instrument in their support for a value claim. Implementation of the rules means that generic multiattribute instruments are effectively outlawed in the US. While this may shock those who, for 30 or more years, have subscribed to instruments such as the EQ-5D-3L/5L, once the standards of fundamental measurement are recognized, the demise of these instruments and their claims for universal application, is seen as long overdue. The fact that it has been such an extended period of dominance, with ordinal scores supporting the mathematically impossible QALYs, is a salutary lesson for the rejection of both the standards of normal science and fundamental measurement in the present health technology assessment meme. From the patient perspective the concern must be for the potential for adverse decisions in pricing and therapy access due to the application of these modeled false decision criteria.

All multiattribute instruments must be outlawed. These include not only the EQ-5D-3L/5L but the Health Utility Index (versions HU12 and HU13), the Quality of Well Being (QWB) scale and the SF-36 with and its various progeny including the SF-6D. More recently, but in development for 10 years, is the EQ-Health and Wellbeing (EQ-HWB) measure designed to soften the image of the austere clinical EQ-5D-3L/5L instruments by tagging on a supplementary list of questions<sup>11</sup>. The focus is generic to capture broader values to calculate QALYs as a common outcome across health and social care settings. Again, with the focus on resource allocation the end product is designed to be a composite score anchored at 0 and capped at unity. As such, the EQ-HWB would still be subject to the 504 prohibitions as well as failing the standards for fundamental measurement as a multi-attribute confection.



While these measurement failings are well established, there is a dogged insistence at the center of the current technology assessment belief system, that these reference models are a mainstay. The latest evidence for this is the CHEERS 2022 guidance for submitting modeled imaginary claims for publication in leading journals. The issue of fundamental measurement is not addressed. To do so would cut the ground out from under the guidelines in their attempt to encourage the creation of more imaginary cost-effectiveness claims and the increased likelihood of their acceptance in leading journals who subscribe to such claims<sup>12 13</sup>.

## THE CURSE OF NEGATIVE NUMBERS

The failure of multiattribute utility and preference numbers is demonstrated by the fact that these numbers can be negative. That is, the number reflects a state worse than death as valued by community preferences. Implicit in the disability case for abandoning ordinal preference and utility scores and hence the QALY, is the fact that the inputs to utility or preference algorithms are based on a community sample valuing hypothetical health states. Asking a random sample of a country's population to value each of a sample of hypothetical health states defined by a handful of disease dimensions, each offering a problem level, is asking for a eugenic interpretation. In the case of the most popular multiattribute instruments, the EQ-5D-3L/5L, the problem is not only the community valuation or weighting of these health states, interpolated from a sample of responses, but the fact that 34.6% of health states yield negative scores when the estimated community weights are applied to create ordinal scores<sup>14</sup>. Capped at unity by design the range open is from 1 to -0.594. In the case of the EQ-5D-5L, the range in scores is 1 to -0.285, with 5.1% of health states valued as worse than death.

Care has to be taken in interpreting these scores: they refer to the weighted community values of the health state description. It does not mean that a study will necessarily encounter states worse than death or find states worse than death dominating a target patient population. The lower proportion of states worse than death for the EQ-5D-5L has to be seen against the number of health state descriptions. The EQ-5D-3L produces 243 health states where 84 are worse than death (34.6%) while the EQ-5D-5L produces 3,125 health states with 159 worse than death (5.1%).

Although standard arithmetic operations such as addition and multiplication are disallowed for ordinal rankings (only medians or modes), ignoring this restriction and creating average numbers means that negative and positive scores are combined yielding a lower value for a QALY than when negative numbers were absent. In the traditional incremental cost-per-QALY calculus, this means a higher cost per incremental QALY as the QALY count is lower. This devalues cost-per-QALY claims where states worse than death are a significant proportion of patient health state valuations, with the possibility of target patient group negative 'average' QALY.

The reasonable question is how to interpret and act on a [disallowed] negative score? Are respondents who have indicated by their response that they are experiencing a community valued health state worse than death, to be denied access to health care? If health care resources are limited why not divert them to target patient groups with positive scores who may benefit? Given different multiattribute instruments with their own ordinal utility or preference scores, which instrument should be selected to assess a value claim? Which instrument should be used to allocate

health care resources? The answer is none; a conclusion driven by the 504 amendments' specific prohibitions.

A recent paper addresses the interpretation of those allocated negative scores, pointing to the conflict between individual and social preferences in respondents with disabilities<sup>15</sup>. In one example, with a new therapy that extends the life of a patient with a negative score but at no extra cost, the result is that while the patient may wish to receive the treatment, society would consider the health state as worse than death and withhold the treatment. While this is a sample of one, by extension to a group it can be shown that therapies involving states worse than death, the average score for the group, would lead to the same conclusion to withhold treatment. More importantly, if there was a mix of patients with positive and negative scores then the mix of states worse than death could still lead to withholding access to treatment for the group overall. There is one qualification: the paper does not recognize the failure of these preference measures. Instead, the paper concludes with options for 'adjusting' the negative scores. This is of little value given the failure of these scores to meet the standards of fundamental measurement.

### **EUGENICS AND THE QALY**

While those advocating QALYs see it as a valid technique for eliciting societies preferences and thus an efficient driver in the allocation of resources, there are issues of ethics which, as a shorthand, we can call eugenic<sup>16</sup>. While the patient may consider their life worth living, the illuminati in charge of health care resource allocation, may believe otherwise. It is a societal decision. Their view predominates, discriminating against those in poor health, including disability health states, is seen as a societal prerogative. Of course, in this central planning model for the allocation of health care resources, this is exactly what was intended.

If we want to avoid accusations that pharmacoeconomics (or possibly better described as pharmacoeugenics) is focused on access to care with an instrument such as the EQ-5D-3L which is arbitrary in its choice of clinical dimensions and threshold cut-offs, then we must think in patient centric terms. The failure to recognize or even be aware of the standards for fundamental measurement doomed the notion of preference scores and the QALY from the first time the instrument was proposed. If we continue to apply it, which seems unlikely in the US, we are merely promoting a eugenics tradition, with all of its adverse historical connotations.

### **A PATIENT CENTRIC VACUUM**

Given the scope or reach of the health care sector in the US with, as the Section 504 amendments detail, some 490,000 providers of ambulatory care and 3,044 hospitals, and with the rule covering all recipients of HHS funding, taking effect in July, there is going to be concern over what value claims are to be allowed. No detail was provided on the number of formulary committees as gatekeepers for value claims assessment. With HHS funding at potential risk, attention should be given to formulary committees and other health system decision makers to ensure formulary submission standards for manufacturer submitted claims meet those of normal science and fundamental measurement.



It is time to step back from what has become a Pythonesque search for the Holy Grail of a single metric to allocate health care resources on the basis of a community valuation of minimalist health state descriptions which, for the majority of the respondents, involves valuing health states they have never encountered or can conceptualize in disease specific and patient centric terms. The answer, as the Section 504 amendments make clear, is to eliminate the present focus on resource allocation and a naïve belief in the QALY that have entrenched discrimination as an unfortunate by necessary by-product of the search for global efficiency in health resource allocation. The Section 504 amendments have eliminated preference scores and QALYs; but this is only a first step. The challenge is to define patient centricity and establish standards for value claims. that address the concerns and needs of target patient populations.

### **PATIENT CENTRIC VALUE CLAIMS,**

Three criteria are proposed for a formulary committee to consider in the assessment of a value claim. These are:

- All value claims for therapy impact, whether for clinical endpoints, PRO, drug and resource utilization must meet the standards of normal science for credibility, empirical evaluation and replication;
- All value claims must meet Rasch of fundamental measurement standards, with outcome value claims for patient centric instruments supporting successful needs fulfillment as a single attribute that meets Rasch conjoint measurement standards for respondent ability and item difficulty
- All value claims must be supported by a protocol detailing how the claim is to be assessed and reported.

As far as the standards of normal science are concerned, the focus is on demarcation <sup>17</sup>. We have to distinguish science from non-science claims such as those from reference models. Empirical evaluation is essential, both to value the merits of the claims as well setting the stage for a program to build on our information base; the evolution of objective knowledge <sup>18</sup>.

### **PATIENT VALUE AND NEED FULFILLMENT**

Rasch models are by definition patient-centric; they meet the standards of normal science in creating, if the Rasch fitting criteria are met, unidimensional, linear, interval and invariant assessments of response to therapy with the application of standard statistical techniques. The Rasch PRO instrument is developed from extensive interviews with a sample of the target patient population. The model is probabilistic where the likelihood of successfully responding to an item is a function of the difference between respondent ability and item difficulty. A framework that is entirely relevant to disability populations.

The starting point in Rasch PRO modelling is to identify a latent construct of interest <sup>19</sup>. 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 developed to date by Galen Research (UK) 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 <sup>20 21</sup>.

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*<sup>18</sup>. 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. 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. 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.

From the perspective of a therapy or other intervention, given the relevance of the needs identified, ranked in terms of the difficulty perceived by respondents in meeting those needs, response to therapy is assessed by the extent to which the intervention increases the number of needs that are fulfilled in the target population. With an interval scale the response is captured by integer count of changes in the number of successful responses. These can be assessed with traditional statistical techniques as this is a true interval scale with a similar analysis applied if the interval scale is transformed to a ratio scale bounded by zero and unity. As the ratio scale is a proportion it can be applied to create the needs fulfilled equivalent of a target patient group specific N-QOL, but without any inbuilt bias for disability or similar populations (e.g., late-stage cancer treatment) in the attempt to create a universal metric<sup>22</sup>.

Rasch modelling provides the necessary and sufficient framework for making the transition from observations to measurement. It is unique in the techniques that are applied to yield a patient-centric interval and then a ratio measure. The problem is that Rasch modelling is essentially foreign to health technology assessment: we must reject not only multiattribute instruments but also the vast proportion of the hundreds of disease specific PRO instruments. The reason for this disavowal of extant patient-centric PROs is that no thought was given to the development of an instruments that was unidimensional, linear, interval, invariant and empirically evaluable. Rasch modelling is unknown to the majority of those producing disease specific instruments.

Developing a patient-centric Rasch instrument is relatively straightforward. There have been software packages available that provide necessary guidance; software which has been available for some 45 years, plus an extensive global Rasch network of modelling expertise. There is the question of who is to take responsibility for instrument development; clearly, it has to be the manufacturer as part of product and protocol development. Certainly, the manufacturer may delay the investment until it becomes a requirement for a formulary submission. But this will create

delays in market entry. It is far better to assess formulary committee or gatekeeper requirements for a patient centric measure. In turn, this may require education not only of the manufacturer but also the gatekeepers. The role of HHS in this transition to a patient centric framework is unknown, but there must be some responsibility for creating a vacuum.

## CONCLUSIONS

The Section 504 amendments to the Rehabilitation Act are a game changer in health technology assessment. They have overthrown what we can now see as 30 wasted years of neglecting the standards of normal science and fundamental measurement; a unique achievement in the social sciences. The translation from what might be described as a technology assessment belief meme that supports imaginary modelled claims to a technology assessment paradigm that is patient centric in meeting the standards for fundamental measurement.

To assist in this translation reference is made to the recently released University of Wyoming Certificate Program: A New Start in Health Technology Assessment which sets out the standards required, the techniques for creating patient-centric measures and the questions that formulary committees and others should address in establishing the required evaluation criteria. This on-line program comprises 14 modules with extensive notes (83,000 words) with audio-video links and short examinations. It is credited for 20.5 hours by the Accreditation Council for Pharmacy Education. For those who do not require ACPE the university will provide a certificate. The program provides sufficient information for those wishing to implement graduate courses with proposed working groups to provide in-person support. The link to the program is:

<https://www.uwyo.edu/pharmacy/resources/certificate-program-a-new-start-in-health-technology-assessment.html>

## REFERENCES

---

<sup>1</sup> 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>

<sup>2</sup> Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes (4<sup>th</sup> Ed). New York: Oxford University Press, 2015

<sup>3</sup> 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>

<sup>4</sup> Bond T, Yan Z, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4<sup>th</sup> Ed.). New York: Routledge, 2021

<sup>5</sup> Stevens S. On the Theory of Scales of Measurement. *Science (New Series)*. 1946;103: 2684

---

<sup>6</sup> Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Danish Institute for Educational Research, 1960; expanded edition with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press, 1980.

<sup>7</sup> Luce R, Tukey J. Simultaneous Conjoint Measurement: A new type of fundamental measurement. *J Math Psychol*;1(1):1-27

<sup>8</sup> McKenna S, Heaney A. Composite Outcome measurement in clinical research: The triumph of illusion over reality? *J Med Econ*. 2020;23(10): 1196-1204

<sup>9</sup> Wright B, Linacre J. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil*. 1989; 70(12):857-60

<sup>10</sup> Bernfort L, Gerdle B, Husberg M et al. People in states worse than death according to the EQ-5D UK value set: would they rather be dead? *Qual Life Res*, 2018;27(7):1827-33

<sup>11</sup> Brazier J, Peasgood T, Mukuria O et al. The EQ-HWB: Overview of the Development of a Measure of Health and Wellbeing. *ValueHealth*. 2022;25(4):482-91

<sup>12</sup> 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>)

<sup>13</sup> Husereau D, Drummond M, Augustovski F et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 22) Statement: Updated reporting guidance for health economic evaluations. *ValueHealth*. 2022;25(1):3-9

<sup>14</sup> Mulhern B, Feng Yan, Shah K et al. Comparing the UK EQ-5D-3L and English EG-5D-5L Value Sets. *PharmacoEcon*. 2018;36:699-723

<sup>15</sup> Schneider The QALY is ableist: on the unethical implications of health states worse than dead. *Qual Life Res*.2022;31:1545-52

<sup>16</sup> Langley P. Abandoning Eugenics and the QALY. *Innov Pharm*. 2021;12(3): Article 20

<sup>17</sup> Pigliucci M. Nonsense on Stilts: How to Tell Science from Bunk. Chicago: Chicago University Press, 2010

<sup>18</sup> Popper K. Objective Knowledge: An Evolutionary Approach (Rev Ed). Oxford: Clarendon Press, 1979.

<sup>19</sup> Andrich D, Marais I. A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences. Singapore: Springer, 2019

<sup>20</sup> 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

<sup>21</sup> Galen Research, Manchester UK. Measures Database. [www.galen-research.com/measures-database/](http://www.galen-research.com/measures-database/)

<sup>22</sup> Langley P, McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InnovPharm*. 2021;12(2):Article 20