MAIMON WORKING PAPER No. 8 JUNE 2024

THE TEN COMMANDMENTS FOR PATIENT REPORTED OUTCOME VALUE CLAIMS

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

Abstract

With the demise of value claims for therapy impact based on multiattribute generic instruments preference scores and the QALY, the focus has shifted to patient centric value claims. Rejecting generic multiattribute instruments also means rejecting attempts to discriminate between patients on the basis community preferences for health states. These are abandoned not only because they fail standards for fundamental measurement but because of their eugenic implications in the opportunity to deny access to care with application of mathematically impossible QALYs and cost-per-QALY thresholds. The focus must be in the concerns and objectives of patients. Unfortunately, the demise of the application of multiattribute scores means that a vacuum has been created for patient centric or disease specific instruments as virtually none of those in current use meet required fundamental measurement standards. The only exceptions are those instruments that have been developed following the application of Rasch rules for transforming observations to measurement. To achieve this it is important to follow the Rasch model requirements. The Ten Commandments presented here achieve that objective and must be followed if a manufacturer is to justify patient centric value claims.

INTRODUCTION

In the United States, the amendments to Section 504 of the Rehabilitation Act of 1973 have effectively outlawed the acceptance of value claims for health interventions based upon multiattribute preferences and quality adjusted life years (QALYs) [1]. Generic value claims driven by assessments of health states by the community are considered not only discriminatory but have unfortunate eugenic implications illustrated by the category of health states worse than death [2]. At the same time, this is an entirely wasted effort as these multiattribute value claims fail the standards for fundamental measurement; the QALY is a failed mathematical construct [3].

The purpose of this brief note is twofold: first, to detail why these generic multiattribute preference scores and (QALYs are a failure; and second to propose ten criteria or 'commandments' that must be met if a patient-centric, disease specific instrument to support value claims is proposed. These criteria, which formalize the required patient reported outcomes (PRO) submission standards for manufacturers as well as those that a formulary should acknowledge, follow from an earlier study that focused on the presence, across the board, of PRO instruments which inform little about the impact of interventions [4]. A checklist was provided that identified issues that should be addressed in selecting a PRO instrument and interpreting the data they produced. As the focus, as here, is on Rasch measurement, the majority of PRO instruments would fail to meet checklist requirements; a conclusion that escapes those undertaking systematic reviews or formulary committees. The study concludes with a review of the development of the Chron's Life Impact Questionnaire

(CLIQ) to meet Rasch standards [5] . The purpose in this further consideration of Rasch standards in PRO development is to emphasize the link between the fundamentals of Rasch modeling and the standards that should be met in any patient centric value claim submission.

Establishing these ten criteria is important if we are to meet the challenge of meaningful value claims. The overwhelming majority of disease specific measures that are available are redundant; they also fail to meet the standards of fundamental measurement . To overcome the vacuum created by the Section 504 prohibitions, considerable effort will have to be put into creating a new set of disease specific instruments which meet the Rasch standard for value claims [6 7]. Rasch standards are critical as they are the necessary and sufficient condition for translating observations to measurement [8].

THE EUGENIC MYTH

There is, quite obviously, a downside when a community sample is asked value health states of which few if any have any concept of what the bundle of health states translates into disease states. Even though the respondents may be asked to value the health state, with an instrument such as time-trade off (TTO), trying to force respondents to a 0 to 1 valuation scale, can still present problems when the valuation is negative [9]. In the development and application of multiattribute preference scores the result is that, despite tweaking of the observations, a substantial proportion of the health states result in negative scores. These scores, euphemistically described as 'states worse than death', with no implied religious status, create problems. In the case of the EQ-5D-3L, the most widely used ordinal multiattribute score has 243 health states defined by 5 health dimensions and 3 problem levels. Of these, with a UK valuation, 34.6%. of the resulting ordinal scores are negative (range 1 to – 0.594). Clearly this is nonsense unless the community wishes to withdraw health system support, leaving them to fend for themselves.

This is an issue of particular relevance to those with disabilities. This is seen in the health state bundles designed to allocate preference scores, both positive and negative. The issue is the relevance of this attempt to capture the Holy Grail of a single interval metric to guide the allocation of health care resources through the application of cost-per-QALY thresholds. A search that is denied from the outset once Rasch measurement is recognized.

FAILING FUNDAMENTAL MEASUREMENT

The search for a Holy Grail becomes even more absurd once the standards of fundamental measurement are invoked to judge the merits or otherwise of the preference scores created by multiattribute community weighted algorithms. The answer is clear cut: multiattribute preference scores fail to meet the standards of fundamental measurement. The preference scores are ordinal; no account was taken of the need to apply Rasch rules to transform single attribute ordinal observations to interval The first step in measurement, a requirement that has been dictated for over 60 years, is to apply Rasch rules as the necessary and sufficient requirement. By fitting observations to the Rasch model we can justify the creation of a single attribute, linear, interval and invariant measure that can be empirically assessed across target patient populations in a disease state. Composite scores fall at the first hurdle .

Although the existence of Rasch procedures, formalized in readily accessible software packages, had been known well before the focus on multiattribute scores and instruments such as the EQ-5D-3L, this pursuit of the unattainable measure dominated health technology assessment and has continued for the past 30 years. The emphasis has been on creating cost-per-QALY claims as outcomes from assumption driven modelled simulation. This pursuit of the unattainable cost-effectiveness score, not a measure, continues with the recent CHEERS 2022 guidance for assumption driven simulations driving imaginary non-evaluable modelled claims; a guidance intended to convince journal editors to accept these models with imaginary and deliberate false claims in all too many cases [10] [11].

THE LATENT CONSTRUCT

If we are to understand the unique role of Rasch measurement then we have to make a distinction between latent constructs and the properties or manifestations of that latent construct. A trait such as quality of life or temperature is not measured directly but indirectly through its manifestation. As Andrich makes clear: in order to measure a trait we require a controlled assessment procedure to manifest the property of interest [7] . We have, as Wright and Linacre make clear, to transform observations to measurement [8]. The challenge that Rasch overcomes is to provide the necessary and sufficient rules for creating an assessment instrument that allows us to measure the manifestation of a latent trait and evaluate claims in terms of that instrument. As Andrich continues: *The Rasch model represents the structure that responses from assessments should have before they can provide measurement and how they can be transformed to provide measurement* [7].

For those that might find this daunting there are inexpensive software packages that support the assessment, the transformation from a preliminary item or statement set to one that has selected those items that fit the requirements of the Rasch model. The fact that these have been available for some 40 or more years casts further doubt on the knowledge and skill set of those involved with multiattribute PRO development and their continued promotion of QALYs as a viable policy tool [3].

NEEDS FULFILLMENT QUALITY OF LIFE

The failure of the reference case assumption driven simulations that produce imaginary or deliberately false claims, does not mean that we have to reject the latent construct we call quality of life. Rather, we have to consider an attribute or manifestation of quality of life that can yield a meaningful measure that meets the required Rasch standards to evaluate therapy impact. This is not difficult; an acceptable attribute, needs fulfillment, has been proposed and applied to create PRO instruments in a number of disease states since the early 1990s [12].

Individuals, as proposed by the needs model, are motivated by their needs; they are satisfied when they are met [9]. The individual values his or her life by the extent to which needs are met. Disease and its treatment are considered the major influences on the ability to fulfill needs. If interventions are effective then more needs will be fulfilled. These interventions need not be clinically focused; improving a handful of clinical symptoms may have little impact on meeting the needs of patients, even in terms of functional status. Multiattribute ordinal scores may increase but the quality of life of the patient, even if it is health related in the case of chronic conditions, may not.

Rather than judging quality of life by a bundle of clinical symptoms and response levels, the needs model generates the items that comprise the disease or target patient population from patient or caregiver interviews. Respondents are asked how the patient's life has been affected by the disease and probe how limitations of function adversely affect them. Statements provided by the interviewees form the basis for item generation and feedback to capture an item or statement set (usually 25 – 30 items) that are the input to the Rasch model.

WHY RASCH?

All patient centric value claims to support claims for therapy response in target patient populations must meet the Rasch standards for instrument development [6 7] . This is not an option but a requirement. It must be emphasized that the Rasch model is probabilistic. Items for the instrument are designed to reflect the ability of the patient to respond to items of increasing difficulty. These items can be presented in dichotomous or polytomous terms and provisionally ranked in terms of the probability of the need being achieved given the distribution of abilities in the target patient population. The probability that a patient will successfully respond to an item is a function of the difference between the ability of the patient and the difficulty of the item. Following application of Rasch rules to fit items to the Rasch model, responses are recorded as integers as an interval measure or, by transformation, proportions as a ratio measure bounded by zero and unity (the N-QOL) [13]. The scales support traditional statistical analysis to assess, for example, effect size. By design, there are no negative values. Community preferences for health states play no role in Rasch instrumentation; Rasch model instruments are non-discriminatory. The notion of a health state worse than death is alien.

If patient-centric PRO instruments are to support meaningful interval or ratio measures and value claims for therapy response then the Rasch model, to emphasize its unique contribution, is the only option. It identifies not just item difficulty as a need but also the ability of the patient to achieve that need. In other words, patients will vary in their ability to respond to a specific therapy intervention. It is this functional relationship between ability and need which distinguishes the Rasch model from all other PRO instruments which produce only ordinal scales, failing to capture response to therapy in neglecting the requirement for an interval single attribute measure. Remember, in the Rasch model the items or statements are selected to fit the model; there is no attempt to fit a preconceived model to the data. If the items fit the requirements of the Rasch model then we can justify the claim that the questionnaire yields a measure which is unidimensional, linear, interval and invariant. This is the unique contribution of the Rasch model.

INVARIANCE

These 'ten commandments' detail the information that should accompany all patient centric value claims that are submitted to a formulary committee or other health system evaluation group. It is important to note that while a manufacturer may have developed a patient centric Rasch instrument and reported its result in a pivotal clinical trial (which is likely underpowered) the value claim is focused on replication of the protocol or reproduction of the claim with a less restrictive protocol. Hence the importance of invariance in instruments: *That for any device, the readings will remain constant across all suitable contexts, and for any one context, all suitably calibrated*

devices will yield invariant readings [6] . This emphasizes the importance of unidimensionality; the measurement of one single construct at a time. A requirement that is central to the concept of fundamental measurement. Unless a single attribute is the focus of measurement it is impossible to produce interval measures of the manifestation of a latent construct.

The construct validity standards means that composite instruments fail fundamental measurement [14]. This includes, by definition, all generic multiattribute instruments and attempts to crosswalk instruments. The requirement by NICE, for example, that modelled reference case imaginary claims must focus on the EQ-5D-3l with required crosswalking from the EQ-5D-5L are absurd as both instruments produce only ordinal scores. The same applies to crosswalking attempts from disease specific to multiattribute scores; both are wasted efforts.

THE TEN COMMANDMENTS

The Ten Commandments are:

1. **All PRO value claims must be patient centric for a defined specific disease or target patient group**
2. **All PRO value claims must be unidimensional or for single attributes with linear, interval and invariant properties**
3. **All PRO value claims must detail the latent construct that is of interest in justifying the development of the value claim together with the manifest of interest that is the focus of the analysis**
4. **All PRO value claims must meet the probabilistic standards of Rasch measurement where the success of response to an item is a function of the difference between item difficulty and respondent ability**
5. **All PRO value claims must be accompanied by a description of how the value claim instrument was developed following Rasch rules for translating observations to measurement**
6. **All PRO value claims must be accompanied by a description of the software used to apply the Rasch standards with a summary of the final analysis to establish the fit to the Rasch model**
7. **All PRO value claims must be accompanied by details of the instrument that is intended to be applied to support the value claims**
8. **All PRO value claims must provide the results of a previous application of the value claim instrument in the target patient population**
9. **All PRO value claims must detail the current baseline for the target patient population and the proposed therapy response following introduction of the therapy for the target patient population as interval integers and ratio proportions**
10. **All PRO value claims must be accompanied by a protocol detailing how the claim is to empirically evaluated and reported in a meaningful time frame**

CONCLUSIONS

Manufacturer's must ensure that PRO claims designed to establish the therapeutic value of their product meet Rach measurement standards. Failure to achieve this objective means that the value claim must be put to one side. Any value claim involving multiattribute modeling will, of course, be summarily rejected. The acceptance of Rasch standards will represent a sea change in health technology assessment (HTA).

While it is usual in the history of science, since the scientific revolution of the 17th century, to find long held beliefs being overturned through experimentation (e.g., luminiferous ether) in the commitment to science and criteria for demarcation [15]. Pharmacoeconomics (or pharmacoeugenics) and more appropriately entitled health technology assessment, stand apart. Thousands of practitioners have held to a non-science belief system for almost 40 years that resource allocation, denial of care, in health systems can be justified by multiattribute preferences and the QALY, together with the crude techniques for reporting disease specific outcomes. All fail the standards for fundamental measurement. While the implicit denial of the standards for normal science and fundamental measurement is unusual, if not unique, outside of beliefs systems such as intelligent design, the fact is that measurement standards, embodied in Rasch measurement, were widely accepted even while these multiattribute instruments were being developed. Add to this the fact that claims based on these multiattribute cost-effectiveness assumption driven simulations claims were not designed to be empirically evaluable.

While this denial of universally accepted standards to support the evolution of objective knowledge, is tragic, maintaining this belief has now achieved the status of farce. The effective abolition of preference scores and QALYS on grounds of discrimination with the Section 504 prohibitions should be seen as a prohibition for the wrong reasons. Whether this will resonate with agencies such as NICE in the UK, CDA (CADTH) in Canada, the PBAC in Australia and PHARMAC in New Zealand is an open question; the most likely response is to circle the wagons and defend the status quo. After all, with some 40 years of a wasted effort to promote QALYs, too many have too much to lose [16]. There is no doubt that the multiattribute reference scores and QALYS fail standards for fundamental measurement. Once we are prepared to accept the Rasch rules for PROs, then the discrimination argument is redundant. Discrimination can only be assessed if the metric in question has single attribute, linear, interval and invariant properties. It is a wasted effort to argue for observed discrimination if the metric is nonsensical; even if the intent is discriminatory. The focus must be on the illegitimate foundations.

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

REFERENCES

[1] Langley P. The United States outlaws multiattribute preference scores and the QALY. *Maimon Working Paper.* No 7 May 2024
https://maimonresearch.com/wp-content/uploads/2024/05/MAIMON-WORKING-PAPER-No.-7-MAY-2024-V21.pdf

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

[3] 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)

[4] McKenna S, Heaney A, Langley P. Fundamental Outcome Measurement: Selecting Patient Reported Outcome Instruments and Interpreting the Data they Produce. *InovPharm*. 2021; 12(2): No. 17

[5] Wilburn J, McKenna S, Twiss J et al. Assessing quality of life in Chron's Disease: development and validation of the Chron's Life Impact Questionnaire (CLIQ). Qual Life Res. 2015;24:2279-88

[6] Bond T, Yan Z, Heene M. Applying the Rasch model: Fundamental measurement in the human sciences (4th Ed.) New York: Routledge, 2021

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

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

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

[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] Langley P. Facilitating bias in cost-effectiveness analysis: CHEERS 2022 and the creation of assumption-driven imaginary value claims in health technology assessment [version 1; peer review: 3 approved]. *F1000Research* 2022, **11**:993 (https://doi.org/10.12688/f1000research.123709.1)

[12] 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

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

[14] McKenna S, Heaney A. Composite outcomes measurement in clinical research: the triumph of illusion over reality. *J Med Econ*. 2020;23(10)1196-1204

[15] Pigliucci M. Nonsense on Stilts: How to tell Science from Bunk. Chicago: University of Chicago Press, 2010

[16] Langley P, Mckenna S. Fundamental Measurement and Quality Adjusted Life Years. *Value Health*. Letter to the Editor. 2021;24(3): 461