MAIMON WORKING PAPERS No. 6 APRIL 2023

METHODS TO SUPPORT HEALTH EQUITY EVALUATIONS: ICER AND NON-EVALUABLE MODELLED CLAIMS VERSUS RASCH MODELLED NEEDS FULFILLMENT

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

Abstract

Assessment frameworks to evaluate society's mandated goals to improve health equity for designated disadvantaged groups must meet both the standards for normal science and fundamental measurement. All claims for interventions to improve health equity must be founded on the standards for credible value claims, empirical evaluation and replication, where the measure for value assessment must meet standards for a unidimensional single attribute, with linear, interval and invariant properties. It is not sufficient to seek ad hoc or hearsay evidence for meet disability targets with proposed interventions, but to put these in a meaningful quantitative framework for empirical evaluation. The recent report by Institute for Clinical and Economic Review (ICER) fails on these grounds. The ICER track record with its commitment to the construction of assumption driven simulated, modeled, cost-per-QALY, non-evaluable claims for cost-effectiveness fails the accepted stands for demarcation; there is no appeal to empirical evaluation. ICER evidence models and their cost-per-QALY threshold claims and the calculated equal value of life years gained (evLYG) fail for one simple reason: the multiattribute preference score that support QALYs and the notional life valuation of evLYG are ordinal. They fail to meet the required standards for an interval or reference measure, the latter being required to construct QALYs and associate evLYG claims. This is a weak framework to support the ICER action statements and the apparent target of improving health equity, equal access to health care resources, in the US. The first step must be to agree the framework for assessing interventions to improve health equity and a measure that captures, not goals set by bureaucrats and interest groups, but the question of whether the needs of patients and caregivers in target patient groups are being met? The key, therefore, must be needs fulfillment as a basis for equity considerations: what are the needs of patients and caregivers? Are these needs being met? Do therapy and other policy interventions support improved needs fulfillment?

INTRODUCTION

The recent release by the Institute for Clinical and Economic Review (ICER) of a paper on Advancing Health Technology Assessment: Methods to Support Health Equity is a disappointing attempt to provide recommendations to improve consideration of health equity within a health technology assessment (HTA) review [1]. With the long-standing commitment by ICER to its reference case modelling, the recommendations appear to provide a defense for imaginary claims with the commitment to engage with diverse groups of patients to capture their views on the potential implications of an intervention under ICER review for health equity. This is an understandable position given the ICER business model resting on assumption driven, modeled simulations, producing non-evaluable cost-effectiveness claims [2]. ICER modeling lacks a meaningful framework to evaluate the quantitative impact of the equity implications of a new therapy. In the report ICER is quite explicit that there should be a minimum threshold for adequate representation of racial and ethnic populations in clinical trials while at the same time advising against cost-effectiveness calculations for subpopulations defined by these characteristic as well as avoiding quantitative equity-informative economic evaluation as a substitute for a deliberative process that should integrate social values in policy decisions.

These are strange recommendations which put to one side any effort to provide quantitative insight into the impact of what may be described as discrimination against certain groups, deliberative or otherwise, and the impact on equity defined by ICER as equal access to healthcare. Of course, there can be little disagreement with the proposition that people should have equal access to healthcare, but the ICER recommendations provide nothing more than a justification for an extended talkfest. There is no proposal for the provision of a coherent and robust quantitative basis for evaluating the needs of patients or caregivers in nominally disadvantaged groups and assess the extent to which needs are not met.

The purpose of this brief note is to put the ICER paper aside and to demonstrate, first, the manifest failings of the ICER approach to HTA and, second, to propose a framework for assessing needs fulfillment, a holistic quality of life construct, to provide a quantitative reference point for policy discussions and equity considerations in health care.

REQUIRED SANDARDS

The standards of normal science and modern or Rasch measurement are not optional; they are required. The standards for normal science have been recognized since the scientific revolution of the 17th century: all value claims must be for single attributes which are credible, evaluable and replicable. The standards for measurement are also clear cut: all instruments that yield measurement must be for single attributes, unidimensional, linear, interval and invariant [3]. Where subjective responses are the focus of a value claim the resolution is more complex as the subjective observations, counts or numbers must be transformed by the application of Rasch rules to an interval measure [4]. Importantly, Rasch rules are unique: they are the necessary and sufficient condition for such a translation to interval measurement [5].

While the standards for interval measurement are a foundation for physics and the more mature social sciences such as education and economics, they are almost completely ignored in health technology assessment. Indeed, HTA takes a markedly different tack in putting emphasis on assumption driven imaginary claims. The ICER reference case is entirely assumption driven with the apparent belief that assumptions from past clinical trials, the literature and expert opinion are a robust basis for non-evaluable, let alone replicable, claims on the future. This overlooks Hume's problem of induction (raised in 1748): the fact that past futures have resembled past pasts does not mean that future futures will resemble future pasts [6] [7]. One set of assumptions is no more justifiable than another unless they support evaluable claims and the possibility of falsification.

Judged by the standards of normal science and measurement, the ICER reference case clearly fails the demarcation test between science and non-science [8]. The assumption driven simulated cost-per-QALY non-evaluable cost-effectiveness claims, allied with cost-per-QALY thresholds for pricing and access recommendations for pharmaceutical interventions is not only multiattribute, lacking by design any attempt to create single attribute, unidimensional interval measurement, but in failing to even consider rules to create an interval measure of response to therapy, it yields only ordinal numbers or counts. The result is a claim for a QALY that is mathematically impossible; to create a QALY you need a ratio preference score; a true zero with well-defined origin and a calibrated measuring system created to meet standards for linearity, interval calibration and invariance in its application.

ICER is perfectly well aware of these shortcomings, but defends its multiattribute health related quality of life ordinal scale by sharing with us ICER's confidence that all health economists recognize, by some mysterious transubstantiation, that the preference score has ratio properties. Although, as Socrates made clear, belief is not knowledge, the ICER position, which is no different from that of the majority of practitioners in health technology assessment, where, as witnessed by the leading textbook in the field and the recent CHEERS 2022 guidance for submitting imaginary model claims to leading journals, clearly rests on a sustained belief in an existential ratio scale despite not being based on evidence [9] [3]. Perhaps, as Dawkins argues, the strength of this confidence is in spite of not being based on evidence; quoting Tertullian, *Certum est quia impossible est* (it is certain because it is impossible) [10].

In any event, ICER's commitment to multiattribute-based HRQoL ordinal preference scores, has unfortunate implications for not only ICER's health status free construct, the equal value of life years gained (evLYG), but other proposed constructs such as the Generalized Risk-Adjusted Cost-Effectiveness (GRACE) framework, the aggregate distributional cost-effectiveness analysis and the proposed EQ-Health and Wellbeing (EQ-HWB) instrument [11] [12] [13]. The first three constructs fail once the required standards are considered. The evLYG the framework rests on an assumption driven multiattribute simulation that yields only ordinal scores together with an impossible adjustment to preference scores to yield life year gains at the full value of a healthy life. This is valued at 0.851 of the value of a healthy life (perfect health) based on age and gender adjusted utility of the US population; this is impossible as ordinal scores cannot be manipulated by any arithmetic adjustment. The result is that the evLYG is an impossible metric. The GRACE framework fails because, once again, it is based on generalizing existing assumption driven simulated modeled cost-effectiveness claims to incorporate diminishing returns to health improvements as disease severity increases. Again, an impossible exercise as the input is a multiattribute ordinal score. The aggregate distributional cost-effectiveness of health technologies as the basis for assessing the inequality impact of technologies also fails because health benefits captured in terms of QALYs are mathematically impossible. Finally, the EQ-HWB is of questionable value because it is not only multiattribute but yields only ordinal scores; ignoring Rasch measurement. As such it cannot capture response to therapy.

THE RASCH IMPERATIVE

The ICER reference case, and the application of similar models in a number of single payer health systems, is unquestionably a failure. It is not, however, simply a question of failing to apply the standards of normal science and measurement theory, but of a more egregious failure to appreciate the role of Rasch measurement in the construct of patient-centric instrumentation. It has been recognized for over 60 years that meaningful measurement in the social sciences must be based on the arithmetical properties of interval scales where the Rasch model is the only rules-based framework for accomplishing this transition from subjective ordinal to interval and, in some instances, ratio scales. To focus on HRQoL or clinician determined characterization of quality of life in terms of bundles of health state descriptions and then achieving a best fit preference algorithm is the fundamental mistake. For some 30 years before these generic multiattribute direct and indirect techniques were proposed, Rasch rules had been developed and applied in instrument development. Rasch focuses on developing from a latent construct a measurable manifestation of that construct that is of interest to the investigator. The steps are quite straightforward, starting with intensive subject interviews, item identification and selection of items to meet Rasch standards for instrument item fit.

But it is more than just item selection. The Rasch solution, which applies across any science, rests upon two requirements. First, irrespective of the context in which a measure is applied, it must be invariant, retaining its quantitative status. Importantly, this implies that not only must invariance hold irrespective of what it is measuring but each item that comprises the instrument interval scale must maintain its level of difficulty irrespective of who is responding, with the competence maintained by the respondent irrespective of the item encountered. These requirements point to the role of capturing and providing a calibrated interval measure for measuring single well-defined attributes as manifestations of a credible and robust latent construct. But that is not all. Rasch realized that the interaction between respondent ability and item difficulty must involve an unpredictable component: the more able is a respondent the more likely are they to respond successfully to an item. The probability of successfully responding to an item is a function of the distance between respondent ability and item difficulty.

THE ICER MISJUDGEMENT

If we are to address issues of health equity, seeking a defensible assessment of the quantitative relationship between an intervention, which could be a new therapy or

a proposed re-assessment of health care delivery for a target patient population, then the Rasch rules are the only game in town. Unfortunately, ICER takes a different tack as it wants to defend its baseline business model of an assumption driven imaginary modelled simulation which defies both the standards of normal science and Rasch measurement. If respondents believe that their needs are met, how much further should we to go invest in therapy responses and the delivery of access to health care to build on this response.

As a starting point we can reject the creation of simulated cost-effectiveness outcomes with ordinal preferences; this is a non-evaluable imaginary framework which rests on the non-defensible multiattribute construct of HRQoL. Reflecting on clinical decisions defining quality of life with community preferences to weigh patient response to support an algorithm seems an odd starting point; particularly where it is hoped that the resulting scores will fall in the range of zero to unity. While this, in practice, is a forlorn hope as the weighted health state scores produce states worse than death (and negative QALYs), there is a complete lack of consideration of the imperative of a unidimensional linear, interval score that captures a probabilistic interaction between item difficulty and respondent ability. Instead, the focus is on the necessity of a single metric where cost-per-QALY estimates can support health care decisions with, to support model building, ready access to the Tufts Medical Center library of ordinal HRQoL preferences; a singular waste for over 40 years of time and resources [14] [15].

If a modeling commitment focusing on a generic multiattribute instrument has been a focus, as with the ICER reference case, then we must avoid the attraction of trying to ‘bolt’ on additional dimensions of health status to extend its life, and retain our business case. In the ICER report the focus is on retaining the existing imaginary simulation but supplement this with an ‘equity’ profile for each disease state or target group. A profile which admits imaginary cost per QALY and cost be evLYG claims as integral to the formulary process.

If the goal is to improve health equity for racial, ethnic and other socially disadvantaged groups, as the primary objective of the ICER paper, then the first step must be to free ourselves from unnecessary and irrelevant baggage in terms of the ICER reference case. This will set the stage for the development of instruments that meet Rasch standards for the quantitative evaluation of equity. The next step is to consider how equity might be characterized as a latent construct from the perspective of the patient or respondent in a defined target population group. Since the early

1990s the question of a more holistic yet single attribute latent construct 'needs fulfillment' has been proposed and developed with over more than 30 Rasch modelled disease specific instruments. Disease states that have been addressed include: Pulmonary Hypertension. Alzheimer's Disease, Atopic Dermatitis, Psoriasis, Growth Hormone Deficiency, Plexiform Neurofibromas, Herpes, Migraine, Multiple Sclerosis, Depression and Asthma [16].

If we accept that it is the patient (or caregiver) who is the ultimate beneficiary of interventions then quantitative estimates of therapy response should be in terms of respondent ability and item difficulty [3]. Item content would be derived directly from a representative sample of the target patient population within disease states. This may be narrowly defined in terms of ethnicity, socio-economic status, education, comorbidities, polypharmacy and so forth; or a more broadly based representative sample may be selected where sub-population groups might be identified.

Again, if we accept the proposition that we are motivated by our needs and satisfied when they are met, then the value of a person's life is determined by the extent to which needs are fulfilled. It is reasonable to assume that, at least in chronic disease, the presence of disease and its treatment are primary contributors. This captures HRQoL, but this restricts our assessment of factors that directly result from health interventions. There are clearly other factors so that an HRQoL focus leaves out the possibility of non-clinical factors social support, disposable income, nutrition, social interaction, education, technology and the environment. However, rather than trying to tie these factors or bolt them on to an HRQoL cluster of dimensions and response a more reasonable approach to assess the needs of patients (and caregivers) directly through qualitative interviews to evaluate how the respondent's life has been impacted by disease and how other limitations (e.g., on function) also impact needs. Responses would then be assessed in needs fulfillment terms, interpreted as a manifestation of the latent construct quality of life, and the final item set for a questionnaire prepared, following Rasch rules, as a direct measure of patient (and caregiver) need, defined in terms of a target patient population in a disease area.

NEEDS FULFILLMENT IMPLEMENTATION

Focusing on the needs of patients and caregivers provides a viable framework for assessing and responding to disparity [17]. The focus on needs fulfillment is not new; it can be traced back to the Nottingham Health Profile in the 1980s and the development of Rasch-based needs fulfillment instrument development from the 1990s to the present [18]. The process for instrument development is touched on in the

ICER report with the role of the patient or caregiver, but then gets lost in the application and selection of techniques. Rasch measurement is never mentioned (or even understood) although the elements raised have been part of the Rasch instrument development model for the past 70 years.

The focus on a target patient specific needs fulfillment instrument presents few if any problems of development. There are numerous primers together with software packages available for over 40 years to ensure that the claim for an interval measure can be made that the instrument for assessing needs fulfillment meets Rasch properties [19]. There is no need for assumption driven simulations. The first step is instrument development to assess the needs of respondents in a target disease area. Clearly, there must be for that disease, a systematic review of patient characteristics to establish the claim for a representative sample of interviewees to support a subjective assessment of needs. This is where the needs fulfillment modelling parts company with the ICER descriptive approach. The process of instrument development, whether the selected items support a dichotomous or polytomous response format, is well established following Rasch rules. The objective is to produce an instrument that yields an integer measure that is unidimensional, linear, interval and invariant. Under certain circumstances, this interval measure can be transformed to a bounded ratio scale with calibration is the range 0 to 1 [20]. The equivalent of the ordinal preferences created from multiattribute generic instrument, but with facility of creating meaningful need fulfillment N-QALYs.

The application of the Rasch instrument to evaluate needs and what may be described as ‘equity shortfalls’ where needs are unfulfilled, is again well established. The instrument can be administered to a target patient group with supporting socio-demographic, income, work status, location and health status questions. This provides a framework for evaluating the extent to which needs are met and the extend to which the supplementary patient characteristics impact need fulfillment (even to the extent of item assessment as these each define a unique need with the items ranked by respondent assessed difficulty). This package can be implemented outside of any particular therapy assessment or as part of a therapy specific clinical trial or observational study.

CONCLUSIONS

If we are concerned to address equity issues in access to health care and the evaluation of competing therapies then the first question to ask is whether, within disease areas and for target patient and caregiver groups, are their needs being

fulfilled? Taken as a holistic capture of quality of life, manifested as needs fulfillment, this is a key input to policy formulation and the quantitative impact of policy and therapy interventions. We must not presume what these needs are in purely qualitative terms; we need to determine what these needs are within different disease states and target groups, assess their relative importance and the extent to which they are met. We can put to one side the ICER commitment in HTA to creating imaginary assumption driven claims driven by generic QALYs and ersatz cost-effectiveness claims, including imaginary evLYG, and the attempt to perpetuate this framework by adding discussions on presumed equity considerations. It is not clear as to why, with assumption driven modelled imaginary claims, ICER can argue that this provides a robust basis for equity considerations. Rather, if a new therapy is to be considered for assessment there is the opportunity to require pivotal protocols to address needs fulfillment and provide a quantitative claim for quality of life defined in needs fulfillment terms. This can support further reviews of the equity implications of a new intervention while providing a quantitative starting point for discussions. This is not, however, a QALY surrogate to be factored into a simulation model. Indeed, this would be moot with the anticipated prohibited use of QALYs and similar measures in coverage and payment determination under Federal health care programs. The N-QOL is not a QALY! This is made clear in the proposed new start for health technology assessment and the University of Wyoming, School of Pharmacy, Certificate Program that focuses on meeting the standards of normal science and fundamental measurement in HTA [21].

REFERENCES

[1] Agboola F, Whittington M, Pearson S. Advancing Health Technology Assessment Methods that Support Health Equity. Institute for Clinical and Economic Review, March 15, 2023

[2] 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

[3] 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)

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

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

[6] Russell B. The Problems of Philosophy. London. 1912

[7] Magee B. Popper. London: Fontana. 1974

[8] Pigliucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010

[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] Dawkins R. A Devil's Chaplain. New York: Houghton Miflin, 2003

[11] Lakdawalla D, Phelps C. Health technology assessment with diminishing returns to health: The Generalized Risk-Adjusted Cost-Effectiveness (GRACE) approach. *ValueHealth*. 2021;24(2):242-49

[12] Love-Koh J, Cppkson R, Gutacker N et al. Aggregate distributional cost-effectiveness analysis of health technologies. *ValueHealth*. 2019;22(5):518-26

[13] Brazier J, Peasgood T, Mukuria C et al. The EQ-HWB: Overview of the development of a measure of health and wellbeing and key results. *ValueHealth*. 2022;25(4):482-91

[14] 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)

[15] Tufts University, Tufts Medical Center. Center for the Evaluation of Value and Risk in Health. Cost-Effectiveness (CEA) Registry https://cevr.tuftsmedicalcenter.org/contact

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

[17]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

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

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

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

[21] Langley P, Sogol E. After the QALY: Training for a New Start Paradigm in Health Technology Assessment. *InnovPharm*. 2023; 14(1): No. 10