

MAIMON WORKING PAPERS No. 14 MAY 2022

MEASUREMENT AND VALUE CLAIMS FOR THERAPY RESPONSE: MAVACAMTEN AND SYMPTOMATIC HYPERTROPHIC CARDIOMYOPATHY (SHCM)

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

Abstract

The failure in the human sciences to apply the rigorous standards of modern measurement theory has been noted since the 1960s. Fundamental measurement is the basis for a rational assessment of, in the case of health technology assessment, patient reported outcomes (PROs) as response to therapy and the submission of credible and evaluable value claims to formulary committees and other health system decision makers. Certainly, it is important to emphasize the importance of ordinal as opposed to interval and ratio scales, but an understanding of conjoint simultaneous measurement and the contribution of Rasch or modern measurement theory (RMT/MMT) is critical. Failure to appreciate the role of RMT/MMT has led thousands of researchers to simply apply numerals to events, applying the techniques of classical statistical analysis, with the result that all that is produced are ordinal PRO scores. Instead, we should be aiming for interval and even ratio scores based on, with PROs, a comprehensible latent trait and the application of the Rasch model. Unfortunately, all too many instrument developers and analysts seem unaware of the need, if we are to meet the measurement standards of the physical science (and the more sophisticated social sciences) to construct fundamental measures before applying the tools of statistical analysis, including heart disease. The purpose of this brief commentary is to review the measurement properties of PRO value claims for Mavacamten in symptomatic hypertrophic cardiomyopathy. Utilizing the NEW START formulary submission application for value claims, the question of the validity of PRO claims, including claims proposed by the recent Mavacamten report by the Institute for Clinical and Economic Review (ICER) as well as the status of the instruments used to support both modeled and standalone PRO claims from both generic multiattribute instruments as well as disease specific ones. The review concludes that apart from purely clinical claims based on the various clinical trials, there are no PRO claims for Mavacamten that meet the required standards of Rasch or modern measurement theory. Mavacamten is not alone; a similar analysis applied to other PRO measures in chronic disease states would, for the most part, reach the same conclusion. What is overlooked, or just ignored, is the role of fundamental evidence to support credible value claims; we cannot continue to rely on valueless intellectual tools.

INTRODUCTION

Meaningful value claims are critical to the review and acceptance of therapy interventions. Unfortunately, in health technology assessment, we face a major problem: none of the generic multiattribute measures that are commonly applied to support value claims, such as the EQ-5D-3L/5L and virtually none of the disease specific patient reported outcome (PRO) claims meet the required

standards for fundamental measurement; they are valueless ¹. The uncritical acceptance of these various PRO instruments over the past 30 or more years illustrates, unequivocally, the analytical dead end that has been the centerpiece of the health technology assessment belief system. This has been well documented but, to all intents and purposes, ignored; the belief system (or meme) continues to hold as witnessed by the endorsement of the latest defense of this belief system, the recently released Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) guide for submitting imaginary health economic evaluations ². Although endorsed by some 15 journals, CHEERS 2022 is an analytical dead, notably in respect of its failure to endorse the standards of normal science and fundamental measurement ³.

Mavacamten for symptomatic hypertrophic cardiomyopathy presents an instructive case study of the failure to recognize the constraints imposed by Rasch or modern measurement theory (RMT/MMT) on value claims. Of course, Mavacamten is not an isolated case; it is representative of literally thousands of clinical trial protocols and modelled claims for cost-effectiveness, let alone single value claims, that, to use a well-worn metaphor, fail to cut the mustard. The purpose of this brief commentary is to set out the limitations and, indeed, inadvisability of value claims that fail the required measurement standards, for PROs and latent traits where the subject communicates directly with the researcher; the researcher captures the patient voice as the essence of MMT/RMT.

Measurement and Filters

If we are to capture a latent trait, with quality of life (QoL) as the prime example, then we must argue that the latent trait is actually measurable and that the resulting measure meets standards for statistical analysis. In other words, measurement and the claim for a relevant measure must precede statistical evaluation. This is typically overlooked; Mavacamten is no exception. It is an error to apply numerals to events and assume that these standards apply; absent fundamental measurement this leaves us in the realm of pseudoscience (and metaphysics).

As far as we are aware, this commitment to fundamental measurement to capture the patient voice is only met with simultaneous conjoint measurement; a measurement perspective introduced in the 1960s that recognizes and supports a commitment in measurement. It must be accepted, that PRO value claims can only be considered valid if it can be demonstrated that they have ratio or invariant interval properties. This follows from the classification of levels of measurement where classical measurement theory recognizes four scales: nominal, ordinal, interval and ratio. Each scale has one or more of the following properties: identity where each value has a unique meaning (nominal); magnitude where values on a scale have an ordered relationship with each other but the distance between each is unknown (nominal); invariance of comparison where scale units are equal in an ordered relationship with an arbitrary zero (interval scale); and a true zero (or a universal constant) where no value on the scale can take negative values (ratio); the ratio scale has the interval property. To these should be added the major contribution of conjoint simultaneous measurement introduced in the 1960s. This is a critical because conjoint measurement it is the basis for RMT/MMT as the only framework to date for creating PRO invariant interval scores; the seminal contribution to ensure that

measurement in the social science matches the standards of the physical sciences establishing the foundation for RMT ^{4 5}. Applying RMT for PRO latent constructs or attributes creates, if feasible, an instrument or measure that combines the difficulty of an item with the ability or likelihood of a respondent completing that item ¹. Under certain circumstances this invariant interval scale can be transformed to a bounded ratio scale while retaining the interval invariance ⁶.

An appreciation of the standards of fundamental measurement sets the stage for the two premises that support the NEW START formulary submission guidelines ⁷. These are:

- All value claims for a product or therapeutic intervention must refer to a single attribute that meets the demarcation standards for normal science: all value claims must be credible, evaluable and replicable
- All value claims must be consistent with the limitations imposed by the axioms of fundamental measurement: they must be unidimensional and meet interval or ratio measurement standards

These premises apply to value claims that are disease or target patient population specific, where every claim is supported by a reporting and assessment protocol.

These claims may be for individual clinical attributes, for quality of life or other PRO measure, and for drug and resource utilization. The formulary committee is in the box seat to determine the relevance of claims for a target patient population and the process for factoring these into pricing and access recommendations. The key point is that claims assessment is an ongoing process where each claim is judged by its credibility, ability to be empirically evaluated and replicated across different treating environments.

If we accept the standards for normal science where value claims must be credible, evaluable and replicable subject to the requirements of fundamental and conjoint simultaneous measurement, then we have the intellectual basis for the only acceptable PRO value claims in health technology assessment. The generic multiattribute instruments have to be rejected; this means the rejection of any composite health related quality of life (HRQoL) generic measure ⁸. The quality adjusted life year (QALY) is an impossible mathematical construct because multiattribute preference scales are ordinal ⁹. The most popular multiattribute generic instruments, the EQ-5D-3L and EQ-5D-5L cannot support any arithmetic operation, only non-parametric statistical analysis; multiplication to discount time in a disease state to create a perfect health equivalent is just impossible unless the 'preference scale' has bounded ratio properties with a true zero and capped at unity.

Importantly, any notion that the ordinal multiattribute preference scores are ratio measures in disguise (a view held by ICER) must also be rejected; if for no other reason that there is no true zero as the various multiattribute algorithms, except the SF-6D, produce health states with negative scores. The EQ-5D-5L, for example, with US value weights, has an ordinal range from -0.573 to 1; this means that of the 3,125 health states defined by the symptoms and response levels (5⁵) 625 or 20% have negative values ¹⁰. As detailed below for Mavacamten, attempts to create value claims from the EQ-5D-5L instrument fail because the scale is ordinal ¹¹. What is overlooked is that if a PRO invariant interval scale

is to be created you require Rasch conjoint simultaneous modeling for a coherent latent construct or attribute such as needs fulfillment quality of life (QoL). This measure will then support standard statistical techniques as an invariant interval scale not the ordinal scale that the EQ-5D-5L produces which cannot support therapy response claims ¹.

The failure to appreciate the necessary role of Rasch or modern measurement theory in the evaluation of PRO instrument claims and other single attribute value claims means that prior to acceptance of a measure, which may be a pivotal phase 3 randomized clinical trial (RCT) selected for systematic review, then both primary and secondary outcomes need to be evaluated for their concordance with the standards of fundamental measurement ^{12 13}. In the case of Mavacamten the pivotal clinical trials include, apart from clinical end point measures for Normal Mixed Venous Oxygen Tension (pVO₂) and Left Ventricular Outflow Tract Time Integral (LVOT), both of which are single attribute and have unidimensional interval properties, PRO outcomes that are unacceptable ordinal scores that cannot capture response to therapy. The EXPLORER-HCM trial (NCT03470545) includes four instruments as secondary endpoints which fail the required standards: the New York Heart Association (NYHA) questionnaire; the Kansas City Cardiomyopathy (KCCQ) questionnaire; the Hypertrophic Cardiomyopathy Symptoms Questionnaire (HCMSQ); and the EQ-5D-5L instrument. The VALOR-HCM trial (NCT04349072) includes as secondary endpoints the NYHA, KCCQ and HCMSQ, while the currently recruiting NCT05174416 trial in Chinese adults includes the NYHA and KCCQ questionnaires.

ASSUMPTION DRIVEN IMAGINARY CLAIMS

The manifest deficiencies of the reference modeling assumption driven simulations as the stock in trade of the Institute for Clinical and Economic Review (ICER) are well established ¹⁴. As these simulations are imaginary constructs producing imaginary, by definition non-evaluable, value claims based on the mathematically impossible QALY they hardly merit attention. They fail quite clearly the standards of normal science as well as those of fundamental measurement. The ICER report for Mavacamten, with the model created by an academic consulting group is clearly of no relevance to formulary decision making or pricing; it is a red herring ¹⁵. The model indicates quite clearly a lack of understanding by the academic consultants of the standards of normal science, in place since the scientific revolution of the 17th century for credible, evaluable and replicable claims but also the axioms of fundamental measurement.

Emblematic of this lack of awareness of fundamental measurement is the place of the QALY in ICER type models, a belief that is difficult to abandon as shown by the current efforts to endorse them through the CHEERS 2022 guidance for creating imaginary assumption driven simulations. The QALY is, without any doubt, a mathematically impossible construct as the preference or utility scales are ordinal. Yet ICER, academic groups and manufacturers persevere in believing the preferences have not just interval but ratio properties ¹⁶. This false belief supports imaginary lifetime modeled claims for incremental cost per QALY and their application to cost-per-QALY thresholds to support pricing and formulary access recommendations. In the case of Mavacamten, this fantasy world yields, in comparison to the standard of care for SHCM, a lifetime QALY improvement of 0.97 (from 12.54 to 13.51 QALYs) with cost-per-QALY

fictional values of \$893,000 versus standard of care and \$1,100.00 versus disopyramide. Set against cost-per-QALY thresholds ranging from \$100,00 to \$150,000 per QALY gained, the ICER Health Benefit Price Benchmark, which represents discounts or price premiums for a price range that for ICER aligns well with the treatment's imaginary benefits to patients over their lifetimes, resulted for Mavacamten in a recommended five-fold price reduction. This recommendation and the supporting analytical framework are clearly absurd. A conclusion, as demonstrated below, that will hold whenever the basis for value claims relies upon generic multiattribute instruments such as the EQ-5D-5L.

THE LIKERT FALLACY

A common feature of disease specific PRO value claims is the application of instrument scores that result from aggregating over the integer points of Likert scales. It has been recognized since the 1970s, at least among measurement theorists, that this simple process of aggregation is only valid if four conditions or assumptions are met: (i) that the Likert items and the proposed scale refer to a coherent and meaningful single attribute or latent construct; (ii) that all of the Likert items (or statements) are, from the prospective respondents perspective, of equal difficulty; (iii) that the thresholds between integer steps for each Likert item are of equal value or equal distance and (iv) that each Likert item has the same number of integer responses or thresholds ¹. If these assumptions cannot be demonstrated then the 'add em'up' procedure for the integer values yields only a multiattribute ordinal scale. With the requirement of a single unidimensional attribute, Likert-based multiattribute instruments with a single overall integer-based response score are clearly meaningless as the basis for therapy response claims.

The KCCQ, with some 381 hits on PubMed (7 May, 2022) is a classic example of a Likert-based multiattribute instrument where response claims are meaningless ¹⁷. The KCCQ comprises, in both its 23 and 12 item versions, Likert scales with both 5 and 6 integer response levels (4 and 5 thresholds respectively) for the Likert items; supposedly capturing 7 domains symptom frequency, symptom burden, symptom stability, physical limitations, social limitations, quality of life and self-sufficiency limitations. Each Likert item is scored from 0 – 100 (violating assumption [iii] above) with the overall integer score scaled from 0 to 100, and frequently summarized in 25-point ranges, where scores represent health status as follows: 0 to 24: very poor to poor; 25 to 49: poor to fair; 50 to 74: fair to good; and 75 to 100: good to excellent. Implicit assumption questions have never been raised or challenged, for a period now of almost 30 years; questions of fundamental measurement are completely absent even to the extent of implicitly claiming equal item difficulty across 7 disparate domains and the implicit assumption that this final score had bounded ratio properties.

The fact that the KCCQ clearly fails to meet the standards of fundamental measurement with MMT/RMT has not diminished its popularity, even in the case of the mavacamten EXPLORER-HCM trial ¹⁸. The KCCQ, in attempting to capture, presumably, seven latent attributes, means it cannot be modified by application of the Rasch rating scale model (RSM) ¹. This relies on a coherent single attribute polytomous Likert structure, where attributes are analyzed and reported on individually; the KCCQ as an ordinal composite score over 7 domains simply fails. If these domains are considered key endpoints

for a comparative evaluation of Mavacamten then they should be reported on separately through application of the Rasch RSM with an instrument designed explicitly to capture the patient voice for each domain (or as instruments defined in terms of dichotomous or binary responses).

There have been extensive favorable assessments of the KCCQ 23/12 psychometric properties. These are, unfortunately, beside the point. Certainly, we can apply the tools of classical statistical analysis; but what is overlooked is that these assessments require prior application of the Rasch model to measure the attribute of interest. RMT stands out in being the only technique where data items are selected to fit the model. This produces an invariant interval measure that is the required input to a range of statistical evaluations, not ordinal data. RMT translate ordinal level data to interval data to support statistical analyses and to support claims for therapy impact ¹. In the case of Mavacamten this requirement for an internal measure is overlooked with the reporting in terms of the overall multiattribute ordinal KCCQ score with 'improved' symptom scores of +9.1 comparing 5.5 to 9.1 for the comparator and Mavacamten respectively ¹⁹. As the KCCQ overall score is ordinal it is impossible to make any claims for comparative therapy impact on the health status of the target patient population. If a blanket claim for improved health status is required there are two options: (i) to focus on a latent construct such as the single attribute need fulfillment quality of life, develop an instrument with dichotomous responses and transform the Rasch interval score to a bounded ratio scale which meets all required fundamental measurement standards or (ii) create a single attribute Likert polytomous data model and apply the recognized Rasch RSM to create the required interval measure ¹. The dichotomous model is the preferred option with many examples of its application in chronic disease states to support needs fulfillment quality of life claims.

The Mavacamten EXPLORER-HCM trial also reports on therapy response defined as the proportion of patients who advance by at least one class in the NYHA. The NYHA is a symptom scale for the impact of systolic dysfunction based on four clinician assessed functional capacity classes: Class 1 – asymptomatic; Class 2 – symptoms with moderate exacerbations; Class 3 - symptoms with minimal exertion; and Class 4 – symptoms at rest. The NYHA is best considered as a Likert scale with fuzzy thresholds and an unknown psychometric distance between fuzzy classes; it is just a rule of thumb, subjective application of heart failure symptoms (with an emphasis on dyspnea). It has an adjunct role in clinical practice, but fails as a PRO scale with fundamental measurement to establish value claims for response to therapy. The NYHA has been subject considerable criticism over the past 39 years, not least because the classes are subjective with a failure to replicate claims across RCTs with similar protocols; indeed, physicians often find it difficult to assign patients to classes with many opting for multiple or bridging class assignment (e.g., class 2/3 or class 3/4).

In what must be considered a somewhat bizarre application, the ICER modeling maps EQ-5D-3L utilities to NYHA classes to represent staging in heart failure response to Mavacamten. Attempting to map one ordinal scale to another is clearly an impossible exercise; one which ignores completely the requirements of fundamental measurement. This applies not only to the EQ-5D-3L/5L scales but also to mapping with the KCCQ ordinal scores. As an example, a recent paper assigned the KCCQ 80 plus (ordinal) score group (assumed to represent perfect health) to NYHA classes²⁰. The comparison found

that this KCCQ group comprised 24.4% of patients categorized as NYHA Class 1; 51.7% categorized as NYHA Class 2 and 18.9% categorized as NYHA Class 3. Given the subjective nature of any appraisal to reduce symptoms to a given class the obvious question is whether the thresholds between classes have any meaning – not only from the clinician’s perspective but, more importantly, the patient’s perspective in their subjective experience of heart failure; does a shift from class 2 to class 3 as assessed by the physician have any impact on the baseline needs fulfillment of the patient? Is it of interest to the patient? The EXPLORER-HCM trial claims that, as a response measure, that 34% more patients improved by at least one NYHA class compared to placebo. Unfortunately, as the psychometric distance between classes is unknown a blanket claim for moving between classes (or remaining in a class) has questionable validity. Given that EXPLORER-HCM covers the assignment of patients to both KCCQ and NYHA classes, it would be of interest to assess the extent to which claims for these matched each other for identical individuals, possibly qualifying overall separate PRO value claims for Mavacamten against placebo. It is also worth noting that the primary endpoint for the EXPLORER-HCM trial was a composite of a clinical endpoint, a 1.5ml/kg per min or greater increase in pVO_2 and at least one NYHA class reduction or a 3.0ml/kg per min or greater improvement in pVO_2 , and no worsening of NYHA class. As the NYHA classes are ordinal, lacking a common psychometric threshold distance, it is not clear how this would affect the overall primary value claim. The advice must be to focus only on the pVO_2 criterion.

If the question is one of combining a clinical with a PRO measure to establish a primary endpoint, then the PRO must have invariant interval or ratio measurement properties. This excludes the ordinal KCCF, NYHA, HCSM- SoB and EQ-5D-5L scales. The solution is to ignore the physician as an intermediary to assign patients to NYHA classes but to use a needs fulfillment quality of life measure which is completed by the heart failure patient. This, developed with the tools of RMT will create both an invariant interval and a transformed bounded ratio scale which can then be combined with the pVO_2 to support valid claims for Mavacamten primary outcome claims. There are many examples of RMT based needs fulfillment. One example is the 20-item Psoriatic Arthritis Quality of Life Questionnaire (PSORIQoL), first developed in 2004 with some 60+ language versions and most recently translated to Brazilian Portuguese^{21 22}. This is a binary-response interval cored RMT application, although it has yet to be transformed to a bounded ratio scale for needs-fulfillment QoL assessment.

The Mavacamten EXPLORER-HCM protocol also proposes value claims based on a subdomain of the Hypertrophic Cardiomyopathy Symptom Questionnaire (HCSM: SoB): selecting four items that cover dyspnea²³. The items cover shortness of breath in the previous 24 hours for a 7-day period. Each item score is averaged over the 7 days to create a total sub-domain score in the range 0 – 18, with the lower score indicating less dyspnea. Once again, there is the constraint that the item scores are ordinal. As such they cannot be averaged or an overall score created and assessed with standard statistical techniques. The HCSM and HCSM-SoB suffer from the same limitations as the KCCQ; they cannot support value claims if the standards of fundamental measurement and Rasch measurement theory are applied.

EQ-5D-5L ORDINAL SCORE

Assumption driven imaginary claims are the foundation for the current belief system in health technology assessment. While it may seem odd that assumption driven non-evaluable claims should be taken seriously by health system decision makers, there is no doubt as to their acceptance both by academic groups, who are consultants to ICER, and manufacturers. It has been argued that this reflects a lack of appreciation of the standards of normal science, where value claims should be credible, evaluable and replicable, as well as a lack of understanding of MMT/RMT. Nevertheless, the Mavacamten EXPLORER-HCM protocol puts these considerations to one side in promoting the EQ-5D-5L as the basis of value claims; but claims which have only ordinal measurement properties.

While the ICER simulation modelling focused on the EQ-5D-3L ordinal preference score to support the mathematically impossible QALYs, the EXPLORER-HCM Mavacamten study has based comparative value claims on the EQ-5D-5L ordinal instrument scores. The difference between the two EQ-5D variants lies in the response levels (five) for the same symptom set; they also produce quite different scores for similar health states. Unfortunately, the 5L version suffers from the same fatal flaws as the 3L version of the EQ-5D family. First, the value claims expressed as preferences are not invariant single attribute interval measures only multiattribute ordinal scores. Second, the preferences are a composite bundle of symptoms, each of which is reported on an ordinal scale; for bundles to be 'aggregated' each has to have ratio measurement properties; this is not the case with the EQ-5D-5L scoring algorithm. Third, the EQ-5D-5L does not have ratio properties because it lacks a true zero; in a recent study of the US valuation of the EQ-5D-5L, 20% of health states had negative values (range -0.573 to 1). Lacking a true zero and invariance of comparisons means that the scale is ordinal. It cannot support measures of mean values, standard deviations or comparisons over time (in this case baseline to 30 weeks); the analysis presented is therefore a wasted effort. Fourth, the EQ-5D-5L scale can only support non-parametric statistics; it cannot support claims for response to therapy, only changes in rank orderings, or comparisons with other scales. Finally, considering the role of a single attribute latent construct, the EQ-5D-5L is only an ordinal health-related quality of life measure (HRQoL). It fails to capture the patient voice is evaluating, not clinician determined generic symptoms defined to represent quality of life, but a latent measure such as needs-fulfillment quality of life. capturing both needs difficulty and patient ability in a single attribute interval scale, and in some cases a bounded ratio measure²⁴. This is the only acceptable basis for PRO value claims. As it stands, the EQ-5D-5L claims are not only mathematically impossible but fail, in not asking basic questions on measurement in instrument development, to be anything other than a PRO curiosity. It is to be hoped that no thought will be given to creating QALYs for Mavacamten with this study as these are a mathematically impossible construct.

In summary, the application of the EQ-5D-5L to establish comparative value claims is a waste of time. A multiattribute HRQoL ordinal scale cannot be transformed to a single attribute, unidimensional scale with a true zero and ratio properties. In common with the visual analogue scale (VAS) it lacks invariance of comparisons; for a VAS scale representing distances between fractions or percentages, we need to capture the property of relative difference. The VAS must be transformed from an ordinal to an interval scale by a natural log odds application; a key step in Rasch measurement modeling¹.

CONCLUSIONS

Although Mavacamten is not alone in failing to appreciate the importance and constraints imposed by fundamental measurement, it is salutary to conclude that none of the various PRO measures proposed as value claims or as elements in a value claim have any merit as value claims for comparative therapy response. The NEW START formulary submission framework divides value claims into four categories: (i) clinical claims which require interval or ratio measurement; (ii) PRO claims that have the same interval or ratio options; and (iii) drug and (iv) resource utilization claims that have ratio properties. The various Mavacamten clinical trials fail to provide any surety for PRO claims: the NYHA functional classification, the KCCQ scale, the HCMSQ-SoB scale and the EQ-5D-5L. While the clinical claims meet the required standards (LVOT gradient and pV_{O_2}), combining the pV_{O_2} with an ordinal scale to support primary outcome claims is clearly problematic. It would have been preferable to focus only on the clinical endpoints. Add to this, of course, the ICER assumption driven simulation which is designed to create only imaginary (and non-evaluable) claims for QALYs, incremental cost-per -QALY, QALY thresholds and an overall and imaginary claim for pricing point likelihood of cost-effectiveness.

If a formulary submission to meet NEW START standards is required then Mavacamten has a long way to go to support PRO claims that meet MMT/RMT standards; the first step is to accept the contribution of RMT, recognizing the imperative of fundamental measurement. Few clinical trials are designed with PRO instruments that meet the required standards of fundamental measurement to support acceptable value claims; in this respect Mavacamten is no exception. If filters are applied, all generic and the majority of disease specific PROs would be eliminated. Fortunately, in the case of Mavacamten, the clinical endpoints are noteworthy in their support for purely clinical assessment; but that is as far as it goes. The patient voice is absent with no possibility of value claims for symptom relief or quality of life that meet NEW START requirements. A research program to discover new, yet provisional, facts cannot succeed if the intellectual tools applied are unfit for the task. On a positive note, however, is the fact that we have the tools (and have had for many decades) to apply the required measurement standards to support PRO quality of life needs fulfillment value claims in heart disease and across the board for chronic conditions.

REFERENCES

¹ Bond T, Yan Z, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th Ed.). New York: Routledge; 2021

² Husereau D, Drummond M, Augustovski F et al. Consolidated health economic evaluation reporting standards 2022 (CHEERS 2022) explanation and elaboration: a report of the ISPOR CHEERS II good practices task force. *Value Health*. 2022; 25: 10-31

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

-
- ⁴ Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danmarks Paedagogiske, Institut, 1960
- ⁵ Luce R, Tukey J. Simultaneous Conjoint Measurement: A new type of fundamental measurement. *J Math Psychol.* 1964;1(1):1-27
- ⁶ Langley P, McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm.* 2021;12(2): No. 6
- ⁷ Langley P. The New Start Formulary Submission and Evaluation Guidelines for Value Claims with Pharmaceutical Products and Devices. Version 2.0 May 2022. <https://maimonresearch.com/category/formulary-submissions>
- ⁸ 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>)
- ⁹ Langley P. The Great I-QALY Disaster. *InovPharm.* 2020; 11(3): No 7
- ¹⁰ Pickard A, Law E, Jiang R et al. United States valuation of EQ-5D-5L health states using an international protocol. *ValueHealth,* 2019;22(8):93141
- ¹¹ Xie J, Wang Y, Xu Y et al. Assessing health-related quality of life in patients with symptomatic obstructive hypertrophic cardiomyopathy: EQ-5D-5L – based utilities in the EXPLORER-HCM trial. *J Med Econ.* 2022; 25(1):51-58
- ¹² 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
- ¹³ McKenna SP, Heaney A. Setting and maintaining standards for patient-reported outcome measures: Can we rely on the COSMIN checklists? *J Med Econ.* 2021;24(1):502–511
- ¹⁴ Langley P. Nonsense on Stilts – Part 1: The ICER 2020-2023 value assessment framework for constructing imaginary worlds. *Inov Pharm.* 2020;11(1):No. 12
- ¹⁵ Wasfy J, Walton S, Beinfeld M et al. Mavacamten for Hypertrophic Cardiomyopathy: Effectiveness and Value; Evidence Report. Institute for Clinical and Economic Review, October 7, 2021.
- ¹⁶ 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
- ¹⁷ Spertus J, Jones P, Sandhu A et al. Interpreting the Kansas City Cardiomyopathy Questionnaire in Clinical Trials and Clinical Care. *J Am College Cardiology.* 2020;76(2)
- ¹⁸ Spertus J, Fine J, Elliott et al. Mavacamten for treatment of symptomatic hypertrophic cardiomyopathy (EXPLORER – HCM): A health status analysis of a randomized, double-blind, placebo-controlled, phase 3 trial. *Lancet.* 2021;397(10293):2467-2475
- ¹⁹ Olivotto I, Oreziak A, Barriaes-Villa R et al. Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): A randomized, double-blind, placebo controlled, phase 3 trial. *Lancet;* 2020;396(10253):750-69
- ²⁰ Tran A, Cham P, Jones P et al. Comparison oof patient self-reported health status with clinician assigned New York Health Association classification. *JAMA Network Open.* 2020;3(8)

²¹ McKenna S, Doward L, Whalley D et al. Development of the PSORIQoL; A quality of life instrument specific to psoriatic arthritis. *Ann Rheum Dis*. 2004;63(2):162-0

²² Goncalves R, Heaney A, McKenna S et al. Psoriatic Arthritis Quality of Life Questionnaire: translation, cultural adaptation and validation into Brazilian Portuguese language. *Advances Rheumatology*. 2021;61{13

²³ Zaiser E, Sehnert A, Duenas A et al. Patient experiences with hypertrophic cardiomyopathy: a conceptual model of symptoms and impacts on quality of life. *J Patient Reported Outcomes*; 2020;; 4(102)

²⁴ Langley P, McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*.2021;12(2):No. 6