Evidentiary Standards for Patient-Centered Core Impact (PC-CIS) Value Claims

Paul C Langley, PhD
Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis MN

Abstract

Proposals for a patient centered core impact set (PC-CIS) are of little relevance to formulary and health system decisions, let alone patients and providers, unless the elements included in the data set meet the standards of normal science and fundamental measurement. Adhering to these standards will have the effect of focusing on the adequacy of proposed core impact measures, with a filter in place to accept only those that meet the standards not only of the physical sciences but also mainstream economics. and health economics. Fortunately, we are well aware of what the criteria for acceptance and rejection of the core impacts within disease states should be in terms of their required attributes and their relevance for supporting evaluable value claims, notably for patient reported outcomes, Rasch or modern measurement theory. Care must be taken to delineate the core impact elements: separately identifying those that are purely clinical from core patient centric impacts, which is turn should be separated from impacts defined in terms of drug utilization and resource utilization. The purpose of this brief commentary is to set out the required standards for core impact patient-centric value claims and the framework for evaluating those claims. The critical issue for patient-centered core impacts is to recognize the constraints imposed by the standards of fundamental measurement for target patient populations within disease areas; unless these constraints are recognized we will fail. The leads to the role of Rasch or modern measurement theory calibration as the framework for patient centric measures of latent traits or attributes. From these perspectives PC-CIS is premature; until we have agreed standards for measurement for the impact or outcomes for clinical, patient-centric and resource utilization as a core set of disease specific instruments, it seems pointless to push forward to a wider scope when the present evidentiary foundation is so weak.

Keywords: PC-CIS, core instrument standards, Rasch measurement, interval scales

INTRODUCTION

Value claims for pharmaceutical products and devices can only be understood if we are clear about the standards that must be applied in the development, application and evaluation of instruments to capture response to therapy within disease and therapy areas [1]. Certainly, within this potential profile of value claims, patient-centric value claims can play a key role in informing decision makers; but these must not be seen in isolation from the purely clinical claims for a product or device and impact claims in terms of drug utilization and other elements of resource utilization. Value claims will only make sense if they are relevant to decision makers and are captured in formulary submission guidelines.

We must avoid blanket claims for the Patient Centered Core Impact (PC-CIS) assumed need for patient centered core impact claims that comprise *a patient derived and patient prioritized list of impacts a disease and/or its treatment have on patients (and /or their family and caregivers)* [2]. We must be more specific and avoid a list of 'broad and inclusive' short-term and long-term so-called patient-centric health outcomes and other related implications that fail to meet the required interval or ratio measurement standards. If this is the intent nothing will be accomplished. The Institute for Strategy and Competitiveness (ICHOM) has, over the past decade and more, developed outcome recommended sets across many disease states [3]. This is a laudable achievement, but must be tempered by the caveat that the choice of instrument to capture these constructs must each meet accepted rules for fundamental measurement. PC-CIS is not, therefore alone, indeed it is prefigured in the ICHOM dictionary; but the same caveat applies to PC-CIS as we must respect and adhere to rules for measurement that meet the standards of the physical sciences. The argument here is that these rules are captured by the standards of Rasch measurement theory (RMT) [4 5].

If the focus is on value claims and their evaluation to support pricing and access it is preferable to focus on 'core' claims that are manageable and consistent with the standards of normal science and RMT for credibility, empirical evaluation and replication. Or, more broadly, as Dawkins, summarizes the requirement:

> *…. the selective forces that scrutinize scientific ideas are not arbitrary or capricious. They are exacting well-honed rules and they do not favor self-serving behavior. They favor all the virtues laid out in textbooks of standard methodology: testability, evidential support, precision, quantification, consistency, intersubjectivity, repeatability, progressiveness, independence of cultural milieu and so on* [6].

Corresponding author: Paul C. Langley, PhD
Adjunct Professor, College of Pharmacy
University of Minnesota, Minneapolis, MN
Director, Maimon Research LLC; Tucson, AZ
Email: langley@maimonresearch.com
Website: www.maimonresearch.com

DOI: https://doi.org/10.24926/iip.v13i3.5016

Standards for PRO claim have been proposed which are consistent with the standards for normal science and Rasch measurement theory (RMT) for ensuring claims have interval properties [7]. In the last resort, as discussed below, this requires recognition of the need for application of Rasch measurement assessments.

It is more manageable to insist on the application of standards for instrument design for patient-centric measures that reflect patient (and caregiver) experience and needs. Certainly, clinical consideration and experience are elements in patient experience, but if we are to capture these it should be through specific clinical attributes or outcomes, while recognizing that we need holistic measures of patient outcomes, in particular for quality of life. This does not mean that we should consider multiattribute HRQoL instruments such as the EQ-5D-3L/5L. Apart from their limited application in the symptoms covered and response levels for disease states, they are only ordinal composite scores and cannot, therefore, support response claims or quality adjusted life year (QALY) claims [8 9 10]. A proposed core value claim set of instruments should not include generic, multiattribute instruments. The same caveat applies to disease specific multiattribute measures; our focus must be on disease or target patient specific measures that capture single, unidimensional attributes not on composite measures [4].

The current thinking for the PC-CIS is to focus on disease or target patient population specific 'impacts', including health outcomes, *but also all other meaningful concepts patients might report*. Indeed, the wish list is considerable including *a variety of life impacts, such as symptoms, function, survival, biomarkers, out of pocket costs, family stresses and much more;* expanding upon the existence of core value sets in many disease states. This is a tall order, and probably overblown (and prefigured, as noted, by ICHOM), but must be premised on a commitment to the standards of normal science and fundamental measurement for each PC-CIS attribute or latent trait and its value claim; otherwise, PC-CIS will be a wasted effort.

The purpose of this brief commentary is to detail: (i) the required standards for core outcome or impact claims, whether they are patient centric or not; (ii) to emphasize the critical role played by measurement theory in developing patient-centric measure to support core claims; and (iii) to make clear the distinction between core clinical measures and core patient-centric measures; which are all too often collapsed into a single measure or where the clinical parameters dominate the development of measures such as multiattribute health related quality of life (HRQoL). The take home message is that we have few patient-centric or patient reported disease specific outcome (PRO) measures that meet the required standards for normal science and fundamental measurement [11 12 13 14]. After all, if the chosen or recommended core instrument sets are poorly specified then this failure in measurement will limit any value claims for response to therapy.

NORMAL SCIENCE AND FUNDAMENTAL MEASUREMENT

For those coming from a background and commitment to what is defined as pharmacoeconomics, the fact that application of assumption driven simulations to create imaginary non-evaluable claims for cost-effectiveness, which has lasted for some 30 years, is in fact an analytical dead end may come as a surprise [1]. It is not the role of core impact measures to support lifetime modelled claims. This is not mainstream economics, neither is it science [1]. Rather, it fails the demarcation test, the appeal to superior evidence, between science and non-science [15 16]. The failure of this belief system or meme has been made clear. We can only understand the importance of PC-CIS core value sets if we make clear that such a value set must, to meet the standards of normal science, recognize criteria for credibility of the claim or construct, the potential for empirical evaluation and a commitment to replication of the measure across target patient populations in the PC-CIS focused disease state. This is the first gateway that must be passed if PC-CIS is to be developed, accepted and applied given the intended audience for the various PC-CIS measures. Indeed, the audience must be defined; is it to track patients in a registry, is it to identify unmet medical and social need or is it to support value claims in formulary submission? Irrespective of the audience, these standards must apply. This holds, for example, for the FDA CDER Pilot Grant Program for Standard Core Clinical Outcome Assessments (COAs) and related endpoints program and, as noted, the ICHOM disease specific outcome sets.

If experience to date with disease specific patient-centric or PRO measures is any guide, there will be considerable wasted effort in creating ordinal composite scores. A recent example of wasted effort is the ICHOM international consensus on outcome measures for child and youth anxiety, depression, obsessive compulsive disorder and post-traumatic stress disorder [17]. This commits the expected *faux pas* of relying on classical statistical analysis to conform the various instruments but overlooks the constraints imposed by fundamental measurement and the importance of RMT or modern measurement, where the criteria for Rasch measurement need to be applied before applying these tools [4]. As an example of the misplaced choice of instruments, is the self-reported (ages 8 – 18) Revised Children's Anxiety and Depression Scale (RCADS-P) which comprises 47 items (questions) and 12 subscales with each item response on a 4-point Likert scale (never, sometimes, often and always) with the responses scored from 0 to 3 [18]. The item integer values are aggregated for an overall score and various subscale scores. As noted below, the limitations of Likert-based instrumentation, unless they meet Rasch criteria as an approximate interval score, render the RCADS-P redundant; a wasted effort if the objective is to assign a status category and evaluate response to therapy as the scale and subscales are ordinal and lack construct validity [4].

The second gateway is that the PC-CIS instrument must meet the standards for fundamental measurement, notably the

DOI: https://doi.org/10.24926/iip.v13i3.5016

application of simultaneous conjoint measurement; the basis of RMT. This requirement follows from Stevens' classification of scales of measurement: 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 where the scale supports claims for both absolute and relative differences. To these should be added the major contribution, as a new type of fundamental measurement, of simultaneous conjoint measurement to ensure that measurement in the social science for patient reported outcomes (PROs) matches the standards of the physical sciences [19] [20]. Applying RMT to PRO latent constricts or attributes creates, if feasible, a measure where for each item the probability of success is dependent on the difference between the difficulty of an item and the ability of the respondent. This yields, in RMT, an interval scale. Under certain circumstances, this interval scale may be transformed to a bounded ratio scale, the ideal measure for PRO value claims in therapy response [9]. The failure of the QALY rests on the failure to appreciate that utility or preference scores are ordinal and cannot support multiplication; that requires a ratio scale with a true zero[7].

An interval scale allows the application of classical statistical techniques and meaningful claims for response to therapy. This means we have to recognize the systemic error in instrument development that has characterized disease specific patient reported outcome (PRO) measures over the past 30 years: the misapplication of Likert scale integer responses (see below) to support instrument scores and claims for therapy response. As these are ordinal scores with no reference to Rasch measurement standards they lack scientific merit; they fail the demarcation test between science and non-science unless it can be demonstrated that they approximate a unidimensional interval scale.

Unfortunately, many would object to this insistence on RMT as the foundation for measurement for non-physical or latent attributes; some would go as far as to advocate almost an open season of measurement, a reflection of the continued impact of relativism; defined by Feyerabend as 'anything goes' [21] . Remember, as Stevens emphasized almost 80 years ago: *measurement is defined as the assignment of numbers to objects according to a rule* [4] [22] [23] . The question addressed by simultaneous conjoint measurement and RMT in its practical application is to go the further step and define the rules to identify measurement structures for nonphysical attributes in probabilistic terms.

PREMISES FOR PC-CIS VALUE CLAIMS

Two premises are proposed for all value claims in disease and therapy areas: clinical claims, PRO claims, drug utilization claims and resource utilization claims. These premises are critical not only for PRO response claims, but also as a necessary basis for clinical and resource utilization value claims:

- 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. Unfortunately, few PRO value claims meet these standards. Note, however, that the difficulties associated with PRO claims are not shared with other value claims which can support claims for clinical benefit as ratio measures. The formulary committee or health system 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 for a disease or target patient population. If not, then that value claim should be rejected.

THE LIKERT FALLACY

As noted in the case of the RCADS-P, a common feature in disease specific outcome measures is the development of instruments that are built around items for patient (or caregiver) response with responses presented as Likert scales capturing a number of response items (e.g., attitude to a statement). While these are popular, comprising the overwhelming majority of disease specific PROs, their application is fundamentally flawed. Each Likert scale presents responses as an ordinal scale where the distance between integer values assigned to the response level are unknown; we might as well apply letters instead of numbers. Assigning numbers to the various thresholds in the scale and then aggregating these across the various Likert response items is disallowed because each Likert response is ordinal. The aggregate score (and subscale scores) is just a composite ordinal scale based on the number of Likert items. It can tell us nothing about response to therapy or support any arithmetic operation. At best we can apply non-parametric statistics and focus on medians and quartiles as measures of the dispersion of the ranked ordinal scores.

DOI: https://doi.org/10.24926/iip.v13i3.5016

The RCADS-P is a classic example of these fundamental errors. If we are to place reliance on the aggregate Likert score as a measure of response to therapy, then four conditions have to be me: (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 for each instrument, then the '*add em'up*' procedure for the integer values yields only a multiattribute ordinal scale.

Failing to meet these conditions ensures that Likert-based multiattribute PRO instruments with a single overall integer-based response score are clearly meaningless, and misleading, as the basis for therapy response claims. The RCADS-P clearly fails on each of the first three criteria: the four-level Likert responses are ordinal as the distance between them is unknown, there is no attempt to assess item difficulty for the patient, where patients are likely to differ in the ability to respond and there is no appreciation of the need to conceptualize single latent constructs or attributes.

If the proposed PC-CIS core value set is to have traction then it must recognize the importance of distinguishing, in the application of the standards of Rash measurement, between dichotomous and polytomous instruments. In the former we refer to a single latent construct (e.g., needs fulfillment quality of life) with binary responses to each item in the questionnaire; in the latter we have Likert-type responses to each item with instruments often invalidating the single attribute requirement with scoring over subscales. While dichotomous instrument, with application of Rasch measurement standards, can be transformed to interval scales the same does not apply to polytomous instruments unless we can demonstrate that they approximate an interval scale by applying Rasch rules.

The question, therefore, for polytomous instruments is: to what extent is a Likert-type instrument useful? Is there any way of making a claim for interval standard response to therapy? Can item reduction achieve this? Where there are multiple scales in an instrument, this question must apply to each subscale if considered to apply to a single attribute. Rasch measurement theory has a solution (first proposed in the late 1970s) in the application of the Rasch Rating Scale Model (RRSM) for Likert scale Items of equal length and the Partial Credit Rasch Model (PCRM) first proposed in the 1980s, for instruments with differing numbers of response levels for items in the same instrument [4 13 24].

It is not possible to transform a polytomous Likert-type instrument ordinal score to an interval score; what the Rasch Rating Scale Model provides is a check list of required standards that can be applied to an existing instrument or one being developed. These include overall instrument and item function (fit to the Rasch model), construct unidimensionality, item independence, category and threshold functioning and differential item functioning [4 20]. These evaluations provide a holistic framework for assessing the utility of the instrument. It is not the intention (in instrument development or possible modification) to achieve a perfect fit, but to come to a reasonable conclusion that the instrument is useful as an analog of an interval scale. This has to continually reassessed for different target patient groups within a disease area; a task that is made more tractable by the availability of Rasch statistical packages (RUMM, WINSTEPS, R) for both the RRSM and the PCRM.

BEST PRACTICES

The claim is made that there is a lack of standardized approaches to construct PC-CIS; the plea is for best practices for *guiding and maintaining PC-CIS with a primary focus on patient centricity, and flexibility for innovation and evolution of the set(s)* [2]. This is patently untrue; we have the standards and techniques that are required, once we accept the critical role played by Rasch rules or modern measurement theory which only a few authors have recognized in health technology assessment, to create acceptable dichotomous and polytomous instruments [25 26] . The primary focus of Rasch measurement rules, is to develop units of measurement which at first may be arbitrary (ordinal) but can be iterated along a scale of interest (interval) so the unit values remain the same (relative difference), with the ultimate (and difficult) objective of the Holy Grail of a ratio measure [9].

It is instructive to contrast the Rasch framework for constructing fundamental measures to that typically found in instrument development where data have primacy with the results describing those data; the resulting instrument or model is exploratory of all data elements that are observed and describe the data, a central feature of Item Response Theory (IRT) [19 4]. This stands in contrast to the confirmatory Rasch model where the data elements are selected to fit a predictive model. To achieve the standards of the Rasch model to achieve fundamental evidence rules have to be applied to select the required data elements to ensure they fit the model. This applies the principles of conjoint simultaneous measurement to focus on the size and structure of residuals for fitting. If this is achieved, and there is no guarantee, then it can be claimed that the results can be applied as a measurement scale for the attribute or latent construct that has invariant interval properties. This does not deny the application of statistical analysis, just that Rasch measurement precedes it in establishing a claim for interval measurement. The fundamental question, therefore, that has to be applied to have confidence in best practice for instrument development is whether or not it yields interval level properties for evaluating, among other applications, response to therapy.

DOI: https://doi.org/10.24926/iip.v13i3.5016

PC-CIS NEEDS-FULFILLMENT

The application of Rasch rules is only the starting point. If the focus is on therapy interventions for target patient groups, then we need to report on the value claims and their impact on outcomes defined in both clinical and patient-centric terms over the lifetime of the product and the patient; a basis for ongoing disease area and therapeutic class reviews. In this framework the patient (and caregiver) should have potentially equal billing in therapy assessment. This can be captured with the long-standing concept of disease or target patient population specific needs-fulfillment with Rasch measurement modelling; the subjectively assessed needs of the patient and/or caregiver and to the extent to which we might infer that their needs being met. Purely clinical parameters (e.g., functional status) may only go part way, if at all, to meeting needs.

This can be considered in quality of life terms, where life takes its value from needs being fulfilled. The more needs that are fulfilled, the greater the quality of life. The important point is that over the past 25 years needs fulfillment measures across some 30 disease states have been developed that meet both the standards of normal science and RMT [27] . For the most recent example we have the Alzheimer's Patient Partners Life Impact (APPLIQue) questionnaire [28] [29]. This sets the minimum standards bar for the potential identification and development for other PC-CIS patient centric measures as inferences of latent traits and as complements to purely clinical measures.

The challenge for PC-CIS is to recommend for specific target populations patient-centric measures that meet the required measurement standards. This is not easily accomplished as it requires not only subjective assessments based on patient (and caregiver) interviews, possibly to meet the somewhat ambitious (and probably unrealizable) objective of *gathering information from patients, carers and families about what is important, and deriving from them what is most important to them* [27], but ensuring that the latent constructs or attributes are credible and measurable. This is why the needs-fulfillment Rasch framework is not only critical but essential. As noted above, a transformation algorithm has been developed to translate the Rasch interval measure to a bounded ratio scale; that is a scale defined in the range 0 and capped at unity [9] . This allows a direct measure of the extent to which needs are fulfilled in the target population and also the possibility of assessing the factors most closely associated with the level of needs fulfillment. If the RACDS-P is to be rejected, then the answer is to develop a needs-fulfillment quality of life instrument, following the Rasch model, for each of the sub-domains identified as potentially measurable latent-constructs.

CONCLUSIONS

If the proposed PC-CIS core and supplemental disease specific tool is to meet its objectives in patient-centric outcome assessment then attention has to be given to the standards of normal science and fundamental measurement. Notably, in the latter case, the application of simultaneous conjoint measurement captured by the Rasch or modern measurement framework in instrument development. The objective must be to accept only those measures that yield interval or ratio measurement. Ordinal data are not measurement if our objective is to provide a basis for response to therapy claims; they are just irrelevant. Fortunately, with the wide application of RMT over the past 60 years, we have the techniques for ensuring that we achieve measures of latent traits, commonplace in patient-centric investigations, that allow meaningful measures of response to therapy.

At the same time a review is required to assess core instruments that have already been proposed to support health technology assessment in specific disease states. Many will, no doubt, fail to meet the required measurement standards. These need to be weeded out. Rather than accepting a core list of required instruments to capture, accurately or otherwise, the impact or outcomes of interventions in disease states, we should start with the required instrumentation standards; notably in respect of patient-centric or patient reported impacts, to agree on a minimum set that are meaningful that conform as interval scales, at least, to Rasch measurement requirements. Only then is it appropriate to cast a wider net to capture ancillary or complementary measures, but ones that must meet the required standards. In this framework (and its agenda), PC-CIS is premature. Until a sound measurement base is agreed, it seems pointless to speculate and develop additional measures that lack the required measurement properties.

Conflicts of Interest: PCL is an Advisory Board Member and Consultant to the Institute for Patient Access and Affordability, a program of Patients Rising

Note: The opinions contained in this paper are those of the author (PCL).

REFERENCES

[1] 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)

[2] Perfetto E, Oehtlein F, Love T et al. Patient-centered Core Impact Sets: What they are and why we need them. *The Patient – Patient-Centered Outcomes Research*. 2022.

[3] Institute for Strategy and Competitiveness https://www.ichom.org/patient-centered-outcome-measures/

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

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

[6] Dawkins R. A Devils Chaplain. New York: Houghton Miflin, 2004

[7] 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

[8] Langley P. The Great I-QALY Disaster. *InovPharm*. 2020; 11(3): No 7

[9] 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)

[10]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: 2 approved]. *F1000Research* 2022, 11:993 (https://doi.org/10.12688/f1000research.123709.1)

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

[12] McKenna S, 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.

[13] Mokkink L, Terwee C, Bouter et al. Reply to the concerns raised by McKenna and Heaney about COSMIN..*J Med Econ*. 2021;24(1):857-859.

[14] McKenna SP, Heaney A. COSMIN reviews: the need to consider measurement theory, modern measurement and a prospective rather than retrospective approach to evaluating patient-based measures. *J Med Econ*. 2021;24(1):860-861.

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

[16] Wootton D. The Invention of Science: A new history of the scientific revolution. New York: Harper Collins, 2015

[17] Krause K, Chung S, Adewuya A et al. International consensus on a standard set of outcome measures for child and youth anxiety, depression, obsessive compulsive disorder, and post-traumatic stress disorder. *Lancet Psychiatry*. 2021;8(1):76-86

[18] Chorpita B, Moffitt C, Gray J. Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behavioral Res Ther*, 2005;432:309-22.

[19] Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danmarks Paedagogiske, Institut, 1960

[20] Luce R, Tukey J. Simultaneous Conjoint Measurement: A new type of fundamental measurement. *J Math Psychol*. 1964;1(1):1-27

[21] Feyerabend P. Against Method. New York: Shocken, 1975

[22] Stevens S. On the theory of the scales of measurement. *Science*. 1946;103:677-80

[23] Tal, E, "Measurement in Science", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Zalta E (ed.) https://plato.stanford.edu/archives/fall2020/entries/measurement-science/>.

[24] Combrinck C. Is this a useful instrument? An introduction to Rasch measurement models. In Kramer S, Lahey A Flynn H et al Eds. *Online Readings in Research Methods*. Psychological Society of South Africa. Johannesburg 2020 https://www.psyssa.com/wp-content/uploads/2020/08/Chapter-6_Is-this-a-useful-instrument_An-Introduction-to-Rasch-measurement-models.pdf

[25] Tennant A, Conaghan P. The Rasch Measurement Model in Rheumatology. What is it and why use it? When should it be applied and what should one look for in a Rasch paper. *Arthritis & Rheumatism (Arthritis Care & Research*. 2007;57(8):1358-1362

[26] Gibbons C, Mills R, Thornton E et al. Rasch analysis of the Hospital Anxiety and Depression Scale (HADS) for use in motor neurone disease. *Health Qual Life Outcomes*. 2011;9:82

[27] Galen Research, Manchester UK. www.galen-research.com

[28] Hagell P, Rouse M, McKenna S. Measuring the impact of caring for a spouse with Alzheimer's disease: Validation of the Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQue). *J App Measure*. 2018: 19(3):271-82

[29] Mckenna S, Rouse M, Heaney A et al. International development of the Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQue). *Am J Alzheimers Dis Other Dem*. 2020;35