MAIMON WORKING PAPER No. 19 OCTOBER 2023

INTEGERS, LINEAR TRANSFORMATIONS, LOGISTIC TRANSFORMATION AND VALID CLAIMS FOR THERAPY RESPONSE VERSION 2 [EXTENDED VERSION OF MAIMON WORKING PAPER No. 12]

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

ABSTRACT

Calibrating subjective responses to capture response to therapy has long eluded practitioners in health technology assessment. Rather than recognizing that, based on Rasch or fundamental measurement, valid claims for therapy response must be unidimensional, linear, interval and invariant we have a plethora of measures that fail to meet these standards. What is overlooked, or not recognized, is that Rasch measurement for transforming ordinal observations or counts to interval measures is the only analytical framework that guarantees such an outcome. Rasch is unique in providing the necessary and sufficient means for such a transformation, setting the stage for therapy response claims based on liner and interval measures. There has been no challenge to this for over 60 years. Rasch is the only basis for interval measurement from counts and observations. The purpose of this brief note is to demonstrate that claims based on integer summation or linear transformations are completely unacceptable as measures. The only basis for evaluating therapy response is to create a Rasch logit continuum where item difficulty and respondent ability are iteratively mapped to a common measure. The logit continuum, a measure for the manifest of interest from a latent construct, can be assessed as a single attribute measure where each item in a questionnaire is assigned a logit score on a linear and interval scale. The purpose of this commentary is to give an example to illustrate how the interval logit scale can be transformed to a bounded ratio measure with the same required properties. This is a useful extension which allows application of a logistic transformation to yield a probability score which can be interpreted as an item latent trait weight, can then be used to provide a Rasch-consistent measure of therapy response in terms of the difference in possession of the manifest item scores. This is the only option if we are to measure patient centric response to therapy. The commentary concludes with a review of comments received on an early draft of this commentary to provide a window for views of Rasch measurement which include and how it might be circumvented in favor of instrument scores and claims which are essentially false.

INTRODUCTION

Awareness of the need to meet the standards of fundamental measure to evaluate value claims for therapy response has never been a priority in health technology assessment (HTA); the focus has been on modelling simulated claims rather than recognizing the standards of normal science and fundamental measurement [1]. Indeed, for the majority of those who have developed instruments to capture therapy response, it has not even been an issue; let alone an issue of which they have even been aware. This is unfortunate as the case to be presented here, which is one that could have been made decades ago before the majority of these failed measures were developed, is that the only

measure of response to therapy that is consistent with the standards of fundamental measurement, is one that meets Rasch standards [2] . There is no alternative. If we are to report accurately on therapy response then the measure of response must be unidimensional, linear, interval and invariant. The only way this can be achieved is by the application of a logistic function to the logits of the Rasch measurement scale to produce probabilities. This must be the first step in any value claim for patient or caregiver response to therapy; each questionnaire item must be accompanied by a Rasch probability weight. For those familiar with Rasch modeling, the starting point is the logit continuum represented by the Wright map, which represents the relationship between the distribution of person and item measures along a vertical logit scales [2] .

This requirement places a premium on recognizing that in therapy response, if it is to have any credibility, the only avenue is to apply Rasch measurement. There is no way to work around this requirement. Application of ersatz scales, such as integer summation or composite preference or utility scores, produce results which are false but non-falsifiable. Yet there are many advocates for the maintenance of the status quo to produce false (or imaginary) claims to support pricing therapy choice and resource allocation in health care systems globally, including reference case guidelines for imaginary claims issued by single paper health systems such as the National Institute for Health and Care Excellence (NICE) in the UK [3] [4] [5]. This commitment to pseudoscience; to fail the standards for demarcation is what distinguishes HTA. This is not an intention to deceive, or the commitment to fraud but a belief system which is more pernicious. Pseudoscience stands out as bullshit; a greater enemy of knowledge than science fraud [6]. This differs from lies which are designed to mislead about the truth while bullshit is not concerned with the truth at all [7].

OBSERVATIONS AND COUNTS

Measurement is deduced from a well-defined set of counts [8]. The most frequently found set of counts in health technology assessment is the presence of an event or response defined in binary terms (1,0) where 1 is the presence of the event. Counts can support a rating scale where integer sums indicate ‘more’ rather than ‘less’. But these are still subjective observations, counts on an ordinal scale where distances between categories are unknown. Certainly, we can apply non-parametric statistical assessment to these data, but whatever labels we attach to the observations, we still end up with an integer progression (0, 1, 2, 3 etc.). This is not measurement as understood in the physical and the social sciences (excluding HTA); a measure which can support arithmetical and statistical operations. It is a scale; one that fails that meets the requirements of fundamental measurement and meaningful claims for therapy response.

To categorize observations or counts as measures we have to apply them to a developed calibrated measuring system with a well-defined origin and a workable unit of manipulation [3]. If not, then we have to assume either that all items are equally difficult for the respondent (which admits to a degree of redundancy in the number of items) or that items are of differing difficulties which means we then face the problem of assigning difficulty weights. This raises the further issue of the abilities of the respondents; are all of equal ability or of differing abilities? If we are to develop a measuring system then we have to demonstrate how the transformation from observations and counts, ordinal scales, to a true measure with unidimensional, linear, interval and invariant application properties is achieved. There is no alternative; this has been made clear for over a hundred years. In subjective responses, patient reported outcomes (PROs) in health technology assessment (HTA) we have one, and only one model from such a transformation: Rasch Measurement Theory (RMT). A unique mathematical model which provides the necessary and sufficient means to transform ordinal

observations to an interval measure; a transformation which ensures our measure defined in logits is unidimensional, linear, interval and invariant in its applications to evaluate consistently measured attributes [3]. Invariance is a critical property as it requires readings from a unidimensional instrument to remain unchanged across all suitable applications. If an instrument meets Rasch model standards for item fit then it meets the invariance requirement.

Once the unique imperative of Rasch measurement to support the transformation from observations or counts to a linear and interval measure is recognized, as it was in the first applications in the 1950s, we can put to one side any integer-based value claim which attempts to summarize and make an interpretation of the possession of item responses scored for either dichotomous or polytomous instruments [9]. Integer ratios, the proportion of positive responses, have no clinical significance (except in the mind of the developer) to indicate response to therapy; all we can say, with confidence, is that your score has gone from A to B with no indicated of what this means in quantitative terms. This caveat applies, not only to the hundreds of integer summation instruments that have been developed *de novo* with no attention given to the imperatives or Rasch measurement, but also to attempts to apply re-interpret integer response counts for instruments that have been developed applying Rasch standards. Unless we have confidence in an instrument that embodies the Rasch model, we have no basis for therapy response or product value claims [10] [11] . The puzzle is that this failure of integer summations was well known before the majority of these instruments were developed in the 1980s and 1990s [12]

THE RASCH IMPERATIVE

The questions raised above as to item difficulty and respondent ability go to the essence of Rasch measurement. In probabilistic terms, the Rasch model looks to the likelihood of a successful response to an instrument item as a logistic function of the difference between the person's ability and the difficulty of the item.

The starting point for the development of a Rasch model, following subjective respondent interviews, is to develop a manifestation of a latent construct of interest; an entity such as quality of life manifested as needs fulfillment [13]. This is defined as a series of statements or questions (items) that are the initially selected to capture ability and difficulty. The objective is to fit the items to the Rasch model for a maximum likelihood measure which is for a single attribute such as needs fulfillment, unidimensional, linear, interval and invariant in its application. This item fitting involves application of Rasch standards; the model estimates how well a person fits the data and how well an item fits the data. It is important for ensuring that the items in a test are valid and can be compared as both respondents and items are on a common measurement scale.

The item difficulties represent the level of change or complexity in the items being measured. They provide information on their discriminatory power and are well suited to differentiate individuals with differing levels of ability. Individual items can be evaluated for their effectiveness and removed if they do not meet Rasch standards; again, items are fitted to the Rasch model which stands in marked contrast to the classical approach of fitting the model to the data (e.g., item response theory) [2]. This ability to select and de-select items enhances the flexibility of the final item selection to evaluating response to interventions together with the reliability and validity of the instrument. Items don't change their position (their logit score); the focus becomes on how adept respondents are to successfully answering them and the impact on overall responses of new therapy interventions.

Indicative of the power of the Rasch framework is the fact that, as interpreted in HTA, the quality adjusted life year (QALY) is an analytical dead end; a construct that is mathematically impossible. The argument is straightforward: if we are to discount time spent in a modeled disease state by a preference or utility score which is designed to capture quality of life, then we need a single-attribute, linear, interval and invariant bounded ratio measures in the range 0 to 1. This measure, following the Rasch framework, is a measure of the manifestation of the construct quality of life. It could be, for example, needs fulfillment. Once this bounded ratio scale has been developed it can be applied to time (another ratio measure) to produce a Rasch-based QALY estimate. Unfortunately, the HTA version of preferences or utility scores produced by multiattribute algorithms, produce nothing but composite ordinal scores. The resulting QALY claim is meaningless as ordinal scores cannot support the basic arithmetic calculations, including multiplication. As a composite score it cannot support non-parametric statistics. We have to reject assumption driven simulation models where algorithms to create composite ordinal preferences from instruments such as the EQ05D-3L/5L are meaningless.

THE RASCH INTERPRETATION

As noted, a useful way of categorizing, in Rasch terms, the impact of a therapy or other intervention achieves is to shift the distribution of responses in logits to capture an increase in ability or the increased probability of successfully responding to the distribution of items. This makes clear that in evaluating response to therapy with the unique application of the Rasch model, the starting point is the creation of the logit or real number scale as an item-person map; all value claims for PROs must start with the Rasch logit scale. The crucial step is the iteration to convergence applying probabilistic conjoint measurement that continually adjusts item difficulties, measured in logits, to have a mean of zero to ensure that the measurement scale is anchored appropriately and centered around the average difficulty of the items. This centering simplifies the interpretation of the scale and allows for direct comparison between person abilities and item difficulties on the same scale.

The final logit scale measures the manifestation of the latent trait or construct where the latent trait in the Rasch model is a non-observable entity; what Rasch achieves is to quantify the manifestation of the attribute of interest (e.g., needs fulfilment). The logit scale, to re-emphasize the key point, is this manifestation as a single or unidimensional attribute with linear, interval and invariant measurement where equal distances on the scale are of equal size. We are, in effect, replicating the measurement standards of the physical sciences with the unique Rasch transformation from ordinal observations to interval measures for subjective responses. This is the only basis for meaningful PRO therapy response claims. Rasch pre-empts all other techniques or claims for fundamental measurement for PROs.

POSSESSION OF THE MANIFESTED LATENT TRAIT

Once the logit scale has been established for application in therapy assessment, the question we have to address is to consider the presence of negative values as the average logit is, by construct, zero. There are two ways of accomplishing this; one acceptable the other non-acceptable. Both start with the logit values for each item in the questionnaire. If there are 10 items then we have ten points on an interval scale. We could count the number of successful items directly and report therapy response in terms of the count of items before and after an intervention (possible expressed in percentage terms).

The approach proposed here takes us one step further in applying a transformation of the logit vales to their equivalent proportions (percentages); this retains the Rasch properties of the measure but gives more flexibility in representing the scores in a range 0 – 1 as an approximation to a ratio scaler. One way is the apply a logistic transformation and estimate the proportion that supports the odds ratio with a logistic transformation ($p = 1/(1 + e^{-logit})$ mapping the logits back to proportions in the range 0 – 1. The other way is to transform by applying a linear transformation to transform logits to scale numbers in a range of 0 – 1. These scale numbers are not probabilities. Where the logit range is +/- 3.5, the transformation is scale number = (logit + 3.5)/7. Unfortunately, the scale number transformation is dependent upon the logit range. Table 1, as an example, for 7 items illustrates for a symmetrical range of logit vales the logistic transformation to proportions (Col 1) and in columns 2 to 4 corresponding proportions for a selection of scale numbers for +/- 4.0, +/- 4.5 and +/- 3.5.

TABLE 1

LOGISTIC PROPORTIONS AND LINEAR TRANSFORMATION MAPPING

LOGIT VALUES	LOGISTIC PROPORTIONS	LINEAR MAPPING +/- 3.5 LOGITS	LINEAR MAPPING +/- 4.0 LOGITS	LINEAR MAPPING +/- 4.5 LOGITS
2.75	0.940	0.893	0.843	0.806
1.75	0.852	0.750	0.719	0.684
0.65	0.657	0.593	0.581	0.572
0.0	0.500	0.500	0.500	0.500
-0.65	0.343	0.407	0.419	0.428
-1.75	0.153	0.249	0.281	0.306
-2.75	0.059	0.107	0.156	0.194

Note: Transformations have a common midpoint of 0.500

None of the linear transformation bear any resemblance either to the logistic transformed proportions or to each other; each is determined by the end points chosen for the logit range. This means that if the transformation to proportions depends on the arbitrary selection of logit end points, resulting claims for therapy response will vary. This is not the case, however, for the logistic transformation where each item logit value yields only one proportion. The transformation retains the order and proportional relationship between logits and proportions ensuring the required interval relationship on the proportion measure. This also retains the meaningful and interpretable measurement of the manifest latent trait, possession of which gives the assessment of therapy response to baseline.

With a linear transformation of logits, the underlying relationship between logits and probabilities of the Rasch model is no longer retained. The transformed values, as noted, have no meaningful interpretation in terms of probabilities of success, thus failing to preserve the properties of the Rasch model. The linear transformation destroys the linear relationship assumed in the Rasch model between logits and the manifested latent trait which allows interval level measurement. In other words, the transformed scale (or the choice of scale) may not represent equal intervals of the manifested latent trait latent trait while the ordering of items may not reflect the true order of the manifested latent trait. The linear transformation may lead to a distortion of the original logits introducing bias or skewness compromising estimates of person abilities and item difficulties.

The result is clear cut: linear transformations of logits are not to be attempted. We have to apply a logistic transformation to provide proportions because this is the only transformation that retains the properties of the Rasch logit measure. This retains our commitment to the application of the Rasch model as the only acceptable framework for evaluating therapy response which is truly patient or respondent centric.

RASCH THERAPY RESPONSE

The interpretation to place on the proportions from the logistic transformation is that each is a possession weight. By possession we mean the proportion of the overall latent construct manifested in the instrument items that respondents have successfully responded to. Given the distribution of respondent ability, with increasingly more difficult items determined by the Rasch analysis and item fitting, a new therapy may claim that it improves the overall possession of a latent construct as defined by items that meet Rasch standards. As items become increasing difficult (the probability of a successful response is a function of the difference between respondent ability and item difficulty) the proportions are weights that capture item difficulty.

The more difficult the item *ceteris paribus* the greater the weight that is attached to capture a successful response. When a new therapy is introduced, the argument is that in terms of the latent construct the number and value of the items will indicate the extent to which possession is enhanced. Given that the items are ranked by their degree of difficulty, success with the more difficult items will ensure a greater contribution to possession than success with the least difficult items; or, as noted, we will likely observe a shift in the distribution of abilities reflecting an increased likelihood of successful response, possibly across the board for all respondents. This yields a new possession distribution.

Interpreting the probabilities as item weights gives a straightforward approach to manifested latent trait possession as our measure of therapy response. A case study to assess the extent to which possession can be estimated was presented in our previous Maimon Working Paper [14]. For our present purpose, a more simplified process is presented which gives a more accurate representation of the possession distribution and the assessment of the significance of therapy response. Tables 2 and 3 give an overview. The first step is to create for the respondent sample a matrix of item responses. In this example for 10 items and 10 respondents the prior distribution of successful responses is given in Table 2 and the post-intervention distribution in Table 3. The second step, given the proportion weights or possession metric, is to estimate the weighted sum of items that were successfully responded to for each respondent. For respondent 1 this is 0.616 (Table 2). Third, take the ratio of the count of overall possible item responses or the sum of the probability weights (5.342) divided into the sum of weights for successful responses and apply this for each respondent. In the case of respondent 1 (Table 2) this yields a possession proportion of 0.115 (0.616/5.342). This retains the properties of the logistic transformation from logits to probabilities as we are dividing the latter by a constant. Finally, estimate the mean and standard deviation of the 10 possession proportions with mean values 0.248 (Table 2) to 0.464 (Table 3) and standard deviations of 0.093 (Table 2) and 0.252 (Table 3)

Response to therapy can be judged by the difference between the mean values, the 95% confidence interval and p-statistic, reported for the item distributions in Tables 2 and 3. In this case the respective means and standard deviations are 0.248/0.093 for the pre-intervention baseline and 0.461/0.252 for the post intervention outcome in its impact on possession of the manifest latent trait. This yields a 95% confidence interval of 0.0345 to 0.3915 and $p = 0.0220$ (significant at the

5% level). The effect size is substantial with Cohen's d = 1.121. Note that these possession ratios include the impact of omitted item responses with the average possession increasing by 0.113 or 45.6%.

Response to therapy, therefore, is the extent to which the average latent trait possession for the respondents' changes; reflecting the distribution of abilities for the respondents and the impact of a new therapy on the ability of each respondent possibly to more successfully respond to items than they were unable to successfully respond to previously. Remember, however, the Rasch model is probabilistic; we observe the distribution of item responses which implies some respondents may, as a result of the intervention, now successfully respond to an item but others may still be unsuccessful. As the distributions of item possession meets fundamental measurement standards, we can apply basic statistics to provide an estimate of the significance of a change in possession employing only means and standard deviations. This assumes, of course, that the possession distribution is approximately normal. It is also worth noting that our estimate of the significance of a change is a function of the number of respondents and our choice of the number and distribution of items on the latent trait continuum. In the example presented, there are only ten items and ten respondents, which still yields a statistically significant claim for therapy response.

DISCUSSION

It has been noted on a number of occasions, including commentaries posted to *Innovations in Pharmacy*, that the current belief system HTA offers a unique perspective in assessing the merits of competing therapies and other interventions by the rejection not only of the application of fundamental measurement but also any commitment to the standards for normal science. HTA stands out in the physical and other social sciences in training students, both in pharmacy and other health related occupation, to create and justify false claims for therapy impact. The example that is central to HTA is the assumption driven Markov simulations which create claims for cost-effectiveness which are entirely spurious; it is an analytical dead end.

Comments received on an earlier version of this commentary suggest two possible scenarios for this embrace of false claims [15]. The first scenario is that while Rasch is recognized and understood, it is brushed aside because in the prevailing belief system it is seen as a potentially disruptive actor. Once HTA practitioners accept the standards for credible, evaluable and replicable therapy or value claims, allied by definition with fundamental measurement the focus on creating modelled imaginary claims is seen as nothing more than a charade. Certainly, when a product is launched evidence is limited; but that does not justify creating evidence through modelling imaginary claims to support formulary submissions. Rasch is critical. Once the importance of aiming for single attribute or unidimensional, linear, interval and invariant measures to capture observations is accepted then all composite and the overwhelming majority of PROs in HTA become unacceptable. This must not be allowed to happen. Evidence for this attitude is found in one comment where it was proposed that Rasch measurement is only helpful for high stakes decision making (high stakes undefined); if there are low-stakes then Rasch can be ignored. Whether this reflects a belief that Rasch is not relevant for measuring response to therapy in low-stakes product and device contracting or that if it is a low-stakes, any claim will do. One wonders what decision makers will think of this and the way they see themselves as viewed.

The competing scenario is that, as a comment proposed, that there is little awareness of Rasch measurement as normal science standards and fundamental measurement, these are seldom part of

a medical or pharmacy degree programs. There seems no concept of the philosophy of science and the debates over method; presumably also a reflection of the lack of awareness by teachers. Rasch must therefore be brushed aside, not because there is any coherent argument against it, but because no one is sure or even aware of the implications of meeting fundamental measurement standards and the application of Rasch model in HTA. The leadership are perfectly happy with composite preference measures which are false and the use of integer summations which produce ordinal false claims to set the standards for professional development.

Further comment argued that we should judge measures by the rigor with which they have been developed not in fundamental measurement terms; on these grounds Rasch measurement is, and has been for 60 years, the most rigorous measure. Indeed, Rasch is the necessary and sufficient application that transforms observations to interval measures. Failing to recognize this and arguing that Rasch asks colleagues unreasonably to cast aside their hard work to develop ordinal instruments and scales that fail measurement standard is specious. Rasch has been readily accessible for 60 years. To be unaware of Rasch requirements is no defense.

A number of comments took exception to the draft being read as single-minded, unwavering and antithetical. This is what is intended; we can't try and defend a failure to recognize that unless the Rasch model is followed all we are producing are ordinal scales with false claims. If we are committed to discovery and objective knowledge with replication and reproducibility then there is no option. If we are not interested in progress and replication, as evidence by the thousands of published, assumption driven, non-evaluable one-off simulations, then we are supporting only marketing activities for the sponsors product. There is no concept of progress in understanding better therapy impacts Of course, if the interest is in publications and false claims, then brushing aside standards for normal science and measurement has a lot to offer. Although overlooked in comments, the failure to meet the standards for a ratio measure means that the QALY is an impossible mathematical construct; QALY based claims are, from a fundamental measurement standpoint, meaningless.

This failure is endemic, both in HTA and in the wider community perception of the QALY. The leading textbook on how to create imaginary cost-effectiveness claims makes no mention of Rasch let alone fundamental measurement; all that is offered is a confused account of the apparent properties of cost-per-QALY claims to justify their application in imaginary simulations [3] . In the wider community, notably from patient advocacy groups, the measurement properties of the QALY are a closed book; as is any understanding of Rasch measurement. A recent monograph on the limitations of the QALY has no concept of measurement standards [16] .The fact is that the debate over the QALY misses the target fails; commentators fail to make the simple point that it is an impossible construct. The same failure extends to the summation of integer values to assess response to therapy and to the creation of modeled imaginary cost-effectiveness claims.

But Rasch is deeper than the comments made indicate (or are aware of). Rasch is a probabilistic framework which considers the interaction between the difficulty of items and the ability of respondents as indicators on an interval scale of likelihood of therapy response. The comments also failed to appreciate the need to transform the Rasch logit interval scale to a bounded ratio scale to given a more nuanced assessment of therapy response and the impact of therapies on constructs such as needs fulfillment quality of life. This was made clear, with an example, in the draft.

Illustrative of the importance of the Rasch model in supporting both replication and reproduction of value claims in a therapy intervention is the failure to recognize the invariant nature of the Rasch developed instrument. Invariance means that for any instrument the readings will remain invariant across all suitable contexts (replication in different target populations) while all suitably calibrated devices will yield invariant readings. A key input here is unidimensionality. The instrument must be designed to measure one attribute at a time; the objective in Rasch instrument development. This is a critical property if we subscribe to falsification of claims rather than confirmation. Simply rejecting Rasch out of hand due to lack of understanding or making a claim that we can advance on a number of measurement fronts with non-Rasch instruments to collect evidence misses the point. With item response theory (IRT) models the results of the analysis yield to data primacy and the resulting model is descriptive of those data; Rasch is confirmatory and predictive where the data are required to fit the model through probabilistic conjoint measurement. Rasch is unique, yet necessary and sufficient to create a probabilistic framework for therapy response. While some may long for the simple integer summation, these instruments are sunk costs and should play no role in instrument development to map disease specific programs; replicating and improving our understanding of therapy response.

CONCLUSIONS

If we are to provide measures of response to therapy, the Rasch model is our only option. The focus must be on single attributes as a manifestation of a latent construct. Once we have estimated the Rasch common logit continuum for item difficulty and respondent ability the estimate of the manifest latent trait is straightforward. This estimation retains all the required properties of the continuum with a single latent trait which is unidimensional linear, interval and invariant. All we are required to do is to apply the logistic transformation to the logits and consider each as a measure of the extent to which items that are successfully answered. Each positive item response contributes to the proportion of the latent trait possessed by that individual. This is, quite simply, the total of the probabilities, which are independent of each other, as the maximum possible possession of the latent trait as a summation of possession weights; this is also the basis for the estimate of the summation of item responses.

The unique contribution of Rasch measurement to transform cardinal counts and observations for subjective responses to single-attribute, linear, interval and invariant measures is the divide that separates HTA as a pseudoscience from the potential of HTA as a science. It is this failure to appreciate the concept of demarcation that ensures the HTA belief system is best described as a meme rather than a paradigm; with the focus of the latter on progress within a discipline, a focus on objective knowledge, rather than creating one-off imaginary cost-effectiveness claims.

In the 60 or more years since the Rasch framework was unveiled there has been no sustainable critique that has challenged the unique contribution of the Rasch model as the necessary and sufficient means for transforming observations or counts to an instrument that has interval properties. As noted, IRT is not designed to create fundamental measures where it can be claimed that, if a successful fit to the Rasch model, the instrument has the required properties. The contribution of this paper has been to take the Rasch logit interval scale as the starting point to demonstrate how this can be transformed to a bounded ratio measure. This is a necessary step if the focus is on response to therapy and the impact of therapy interventions on the extent to which

target patient populations can improve their possession of a construct such as need fulfillment quality of life.

Whether those who presently subscribe to the HTA meme will be convinced, after 30 years, that the focus on composite and integer- based scores is not ideally suited to capture therapy response is an open question. With few exceptions, the comments received from practitioners indicate that, at best, they will reluctantly move to Rasch for special occasions (high-risk clients and products) while for the low-risk run-of-the mill clients and products they will stay with the existing false claims from instruments applying integer summations and the composite QALY to create evidence to support imaginary cost-effectiveness claims. This commitment to pseudoscience may be unshakeable.

TABLE 2

EVALUATING RASCH LATENT TRAIT POSSESSION: PRIOR ITEM DISTRIBUTION OF SUCCESSFUL RESPONSES

Items Increasing Difficulty	Item Logit	Item Proportion Weight	Respondents (1 – 10) Respondent Ability increasing ….									
			1	2	3	4	5	6	7	8	9	10
1	-2.484	0.078	1		1	1		1	1	1	1	1
2	-1.437	0.192	2	2	2	2	2		2	2		2
3	-0.636	0.346	3	3	3		3	3	3	3		3
4	-0.156	0.461			4	4	4	4	4	4	4	4
5	0.0	0.500					5	5	5	5	5	5
6	0.310	0.577									6	6
7	0.805	0.690										
8	1.203	0.769										
9	1.704	0.846										
10	2.041	0.884										
Sum Item Weights		5.343	0.616	0.538	1.077	0.731	1.499	1.385	1.577	1.577	1.616	2.154
Latent Trait Possession		Mean = 0.248 SD = 0.093	0.115	0.101	0.202	0.137	0.271	0.259	0.295	0.295	0.302	0.403

Note: Latent trait possession is equal to sum of item weights for items successfully responded to divided by the overall sum of item weights (e.g., for respondent 1 this is 0.616/5.343 = 0.115)

TABLE 3

EVALUATING RASCH LATENT TRAIT POSSESSION: POST ITEM DISTRIBUTION OF SUCCESSFUL RESPONSES

Items Increasing Difficulty	Item Logit	Item Probability Weight	Respondents (1 – 10) Respondent Ability increasing ….									
			1	2	3	4	5	6	7	8	9	10
1	-2.484	0.078	1		1	1		1	1	1	1	1
2	-1.437	0.192	2	2	2	2	2		2	2		2
3	-0.636	0.346	3	3	3		3	3	3	3		3
4	-0.156	0.461	4	4	4	4	4	4	4	4	4	4
5	0.0	0.500		5	5	5	5	5	5	5	5	5
6	0.310	0.577				6	6	6	6	6	6	6
7	0.805	0.690							7	7	7	7
8	1.203	0.769								8	8	8
9	1.704	0.846									9	9
10	2.041	0.884										10
Sum Item Weights		Total = 5.343	1.077	1.499	1.577	1.808	2.076	1.962	2.844	3.613	3.921	5.343
Latent Trait Possession		Mean = 0.461 SD = 0.252	0.202	0.281	0.295	0.204	0.389	0.367	0.532	0.676	0.734	1.000

Note: Latent trait possession is equal to sum of item weights for items successfully responded to divided by the overall sum of item weights (e.g., for respondent 1 this is 1.077/5.343 = 0.202); respondent 10 has successfully responded to all items so the possession is 1.0

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

[3] 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.

[4] Husereau D, Drummond M, Augustovski F et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 22) Statement: Updated reporting guidance for health economic evaluations. *ValueHealth*. 2022;25(1):3-9 https://www.valueinhealthjournal.com/article/S1098-3015(21)03146-6/pdf

[5] National Institute for Health and Care Excellence. NICE health technology evaluations: the manual Process and methods [PMG36] Published: 31 January 2022 https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation

[6] Frankfurt H. On Bullshit. Princeton NJ: Princeton University Press, 2005

[7] Ladyman J. *Towards a demarcation of science from pseudoscience* in Pigliucci M, Boudry M (Eds). Philosophy of Pseudoscience: Reconsidering the demarcation problem. Chicago: University of Chicago Press, 2013

[8] Wright B, Linacre J. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989; 70(12):857-60 https://www.researchgate.net/publication/20338407_Observations_are_always_ordinal_measurements_however_must_be_interval/link/5563b02408ae9963a11ef326/download

[9] Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960.

[10] Andrich D. Understanding resistance to the data-model relationship in Rasch's paradigm: a reflection for the next generation. *J Appl Meas*. 2002;3(3):325-9

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

[12] Merbitz C, Morris J, Grip j. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil*. 1989;70:308-32

[13] McKenna S, Doward L. The needs-based approach to quality of life assessment. *ValueHealth*. 2004; 7(5): Supp 1. S1-S40 https://www.valueinhealthjournal.com/article/S1098-3015(10)60227-6/fulltext

[14] Langley P. Therapy Response, Rasch Modeling and Latent Constructs [Revised]. Maimon Working Papers No. 11 July 2023 https://maimonresearch.com/wp-content/uploads/2023/07/Maimon-Working-Paper-No.-11-July-2023-1.pdf

[15] Langley P. Integers, linear transformations, logistic transformation and valid claims for therapy response. *Maimon Working Papers* No 12 July 2023 https://maimonresearch.com/wp-content/uploads/2023/08/Maimon-Working-Paper-No-12-REV-V7-Aug-2023.pdf

[16] Smith W. Rationing Medicine: Threats from European Cost-effectiveness Models to America's Seniors and other vulnerable populations. Pioneer Publications, 2023,