ENHANCING THE RASCH RESPONSE MODEL FOR VALUE CLAIMS: LATENT TRAIT POSSESSION AND FORMULARY EVALUATIONS

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, and then be used to provide a Rasch-consistent measure of therapy response in terms of the difference in possession of the manifest item measures.

Keywords: False HTA claims, Rasch model, Rasch possession, unique manifest measures

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 a measure of response to therapy that is consistent with the standards of fundamental measurement, is one that meets Rasch standards [2 3] . 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 [4]. This is met in the Rasch model through creating a number line calibrated in logits; where the logit = ln(p/1-p) where ln is the natural logarithm, p is a proportion and p/1-p the odds

ratio. This requirement has been recognized for over 60 years in measurement theory, but completely ignored in HTA. The result is that HTA supports dozens if not hundreds of ersatz instruments that are claimed, falsely, to capture response to therapy [5]. With the few exceptions where the Rasch modelling framework has been followed in instrument development, HTA is locked into a position that is unacceptable but which continues to be promoted under the guise of approximate information [6] [7] with, most recently, a guidance for submitting imaginary and false modeled claims as approximate information to journals [8].

The purpose of this commentary is to propose how the logit continuum can be transformed to a bunded ratio measure which captures the extent to which respondents possess a latent trait such as quality of life manifested as needs fulfilled. This transformed represents a significant step towards both a measures for a single attribute with linear, interval and invariant properties as well as a response to the challenge raised by *Innovations in Pharmacy* for measures of therapy response that meet the standards for fundamental measurement and emphasizes a new start in HTA and rejecting pseudoscientific claims for capturing response to therapy [9]. This rejection applies to ersatz scales, such as integer ordinal summation or composite ordinal preference scores and quality adjusted life years (QALYs) [10]. These are not only false measures but reflect an indifference towards truth in HTA, notably in the construction of evidence with assumption driven simulation models [11] [12].

THE RASCH IMPERATIVE

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, in the present example, 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, is 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.

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, measured in Respondent status is defined in terms of the distribution of successful responses to items. Therapy response is captured by impact on the distribution of successful item responses, given the fixed distribution of the sample of respondent abilities.

THE RASCH INTERPRETATION

Ensuring items meet the standards of Rasch modelling produces, in typical applications, a logit framework for evaluating change where on the common logit or real number scale we can present the distribution of respond abilities and the difficulty of the items selected for the instrument. Measured in logits the number line has 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 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);a measure that reflects the unique necessary and sufficient requirement to transform counts or observations from an ordinal to an interval scale. The logit scale, to re-emphasize the key point, is that this manifestation is 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 RASCH 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 (possibly 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 measure. One way is to apply a logistic transformation ($p = 1/(1 + e^{-logit})$ mapping the logits back to proportions in the range 0 – 1. Another 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 proportions. Where the logit range is +/- 3.5, the transformation is $p = (logit + 3.5)/7$. 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

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.

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 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.

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 proportions as item weights gives a straightforward approach to manifested latent trait possession as our measure of therapy response. For our present purpose a small sample example is presented to give a framework for possession distribution and the assessment of the significance of therapy response (Tables 2 and 3). 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 proportions as we are dividing the former by a constant. Finally, we 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.

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 proportions, 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 true 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 that admit ample opportunity for deception and false sponsored claims [14].

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, and 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 a false construct for therapy response is an open question. This commitment to pseudoscience may be unshakeable. Those subscribing to the HTA meme or belief system may show no motivation to arrive at the truth; no authentic motivation for knowledge

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

.

[4] 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

[5] McKenna S, Heaney A. Composite Outcome Measurement in Clinical Research: The triumph of illusion over reality? *J Med Econ*. 2020; 23(10):1106-1204

[6] 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

[7] Neumann PJ, Willke R, Garrison LP. A Health Economics Approach to US Value Assessment Frameworks – Introduction: An ISPOR Special Task Force Report. *Value Health*. 2018; 21:119-123 https://www.valueinhealthjournal.com/article/S1098-3015(17)33891-3/pdf

[8] 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

[9] Langley P, Sogol M. After The QALY: Training for a New Start Paradigm in Health Technology Assessment. *Innov Pharm*. 2023;14 (1) https://pubs.lib.umn.edu/index.php/innovations/article/view/5211

[10] Langley P ,McKenna S. Measurement, modeling and QALYs [version 1; peer review: 2 approved]. *F1000Research* 2020, **9**:1048 (https://doi.org/10.12688/f1000research.25039.1)

[11] Frankfurter H. On Bullshit. Princeton: Princeton University Press, 2005

[12] Langley P. Bullshit with Scientific Pretensions: Assumption driven simulated claims in health technology assessment. Maimon Working Paper No. 20 October 2023. https://maimonresearch.com/wp-content/uploads/2023/10/Maimon-WP-No-20-October-V3.pdf

[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] Ritchie S. Science Fictions. New York: Henry Holt, 2020