MAIMON WORKING PAPERS No. 12 JULY 2023

INTEGERS, LINEAR TRANSFORMATIONS, LOGISTIC TRANSFORMATION AND VALID CLAIMS FOR THERAPY RESPONSE

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. 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. Application of a logistic transformation yields a probability score foreach of the difficulty items. This score, which can be interpreted as an item latent trait weigh, 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.

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 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] .

OBSERVATIONS AND COUNTS

Measurement is deduced from a well-defined set of counts [3]. 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 sciences; a measure which can support arithmetical and statistical operations.

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 scale with unidimensional, linear, interval and invariant application properties is achieved. 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].

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 [4]. Integer ratios, the proportion of positive responses, have no clinical significance (except in the mind of the developer) to indicate an ordinal range of integer sums that have particular clinical implications for diagnosis. This caveat applies, not only to instruments that have been developed *de novo* with no attention given to the imperatives or Rasch measurement, but also to attempts to apply integer response counts for instruments that have been developed applying Rasch standards. Unless we have the confidence of an instrument that embodies the Rasch model, we have no basis for therapy response value claims. The gold standard is to develop a Rasch model and apply the logit continuum for assessing latent trait status and response [5] [6] .

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 function of the difficulty of the item and the ability of the respondent. The Rasch model proposes that our focus must be on relative differences between measures of ability and measures of difficulty. The probability of a respondent providing a successful response to an item is a logistic function of the difference between the person's ability and the difficulty of the item. The process for developing a logit measure that is common to the distribution of abilities and distribution of abilities is iterative.

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, such as an entity such as quality of life manifested as needs fulfillment [7], defined as a series of statements or questions (items) that are the initially selected questionnaire item responses 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 [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.

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 convergence 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 [8]. 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. This is the only basis for meaningful PRO therapy response claims. Rasch pre-empts all other techniques or claims for fundamental measurement.

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. One way is the apply a logistic transformation and estimate the probability that supports the odds ratio and then the transformation by application of the natural logarithm ($p = 1/(1 + e^{-\text{logit}})$ mapping the logits back to probabilities ensuring the transformed probabilities are 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 illustrates a selection of scale numbers set aside probabilities for +/- 4.0 and +/- 4.5, together with +/- 3.5.

TABLE 1

LOGISTIC PROBABILITIES AND LINEAR TRANSFORMATION MAPPING

LOGIT VALUES	LOGISTIC PROBABILITIES	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 probabilities or to each other. More importantly, the logistic transformation retains the properties of the logit because of maintaining the relationship between the manifested latent trait and the probabilities of success. While logits are on a log scale and probabilities are on a linear scale, the transformation retains the order and proportional relationship between logits and probabilities ensuring the required interval relationship on the probability scale. This also retains the meaningful and interpretable measurement of the manifest latent trait.

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 probabilities 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

If we are to retain the properties of the Rasch model in terms of creating a unidimensional, linear, interval and invariant measure, then the only way to evaluate therapy response is to transform, as the first step, logits to probabilities. But what is the interpretation to place on probabilities as indicators of likely success? If we consider the question of the possession of the manifested latent trait, as noted above, our measure of changes in that degree of possession must be in terms of item 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. Given the relationship between logits and the probability transformation and the retention of measurement standards, the probabilities can be interpreted as a measure of latent trait possession; as weights for the respective items with the higher probabilities contributing the greater amount to possession if successfully responded.

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 [9]. 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 case for 10 items and 10 respondents. The next step, given the probability weights or possession metric, 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 by 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 which range from 0.115 to 0.403 (Table 2) and 0.202 to 1.0 (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%.

CONCLUSIONS

If we are to provide measures of response to therapy, the Rasch model is our only option. Unless we can demonstrate that a questionnaire has been developed applying Rasch standards, it must be rejected. This applies to both dichotomous and polytomous instruments, as well as generic composite measures that fail at the first hurdle once the required Rasch standards are invoked. 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 responded to each contribute 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.

Response to therapy is just 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 that 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 now 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 case study presented, there are only ten items and ten respondents, which still yields a statistically significant claim for therapy response.

Once the Rasch model has been applied to provide a measure of the items relevant to capturing the manifestation of the latent trait with the required properties and expressing these on a latent trait continuum which provides a common measure, in logits, of respondent ability and item difficulty, the next steps are essentially trivial. All we are required to do is to transform logits to probabilities, treat these as manifested latent construct possession weights and estimate the possession of the manifest measured latent trait from the item responses successfully realized by the respondent. These ratios or proportions of manifest latent trait possession support basic statistical analysis of these distributions to provide claims for therapy responses defined by the change in mean values of latent trait possession. Claims for therapy response are in the Rasch tradition of item response in estimates of mathematics ability in educational assessment, as first applied in the 1950s, or in terms of the impact of a new therapy enabling a higher possession ratio of a single manifested and measured attribute.

TABLE 2

EVALUATING RASCH LATENT TRAIT POSSESSION: PRIOR RESPONSE DISTRIBUTION

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
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: The mean and standard deviation values may vary somewhat with software package used.

TABLE 3

EVALUATING RASCH LATENT TRAIT POSSESSION: POST RESPONSE DISTRIBUTION

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: The mean and standard deviation values may vary somewhat with software package used.

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] 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

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

[5] 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

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

[7] 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

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

[9] Langley P. Therapy Response, Rasch Modeling and Latent Constructs [Revised] Maimon Working Papers No. 11 July 2023 https://maimonresearch.com/therapy-response-rasch-modelling-and-latent-constructs