MAIMON WORKING PAPER No. 15 AUGUST 2023

MEDICATION ADHERENCE, RASCH MEASUREMENT AND THE MORISKY MMAS-8 INSTRUMENT

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

ABSTRACT

A continuing feature of health technology assessment (HTA), to include the development of patient reported outcome (PRO) measures is the neglect or lack of awareness of modern or Rasch measurement theory. Where value claims for products are based on subjective responses there is no conception, with few exceptions, of the need to apply the Rasch model to transform the subjective responses to a questionnaire to an interval or ratio scale. The Rasch model, widely accepted over the past 60 or more years in education and psychology, is unique as the necessary and sufficient mathematical model to transform observations or ordinal responses to single attribute or unidimensional linear, interval and invariant measures. Integer based instruments designed to capture medication adherence, including the Morisky-8 instrument, fail to meet the standards for Rasch measurement; there is only the exception of the ProMAS instrument, developed in 2015 which applies Rasch techniques. At the same time concerns with the foundation analysis and presentation of results in the 2008 Morisky paper, have led to its retraction (23 August 2023). The argument presented here is that the basis for this retraction misses the major point: the integer score and cut-off for the Moriosky-8 fails the required standards for measurement. It is just an ordinal score and not, as required for statistical assessment, a single attribute linear, interval and invariant scale. In other words, failure to meet Rasch standards should have been applied as a basis for rejection almost 15 years ago.

INTRODUCTION

The Morisky MMAS-8 instrument is widely used in the evaluation of medication adherence and as a predictor of patient behavior [1]. It comprises eight statements that are scored in a binary response format where a score of >2 indicates low adherence, 1 or 2 is medium adherence and 0 = high adherence. It is an extension of a previous 4 item instrument the MMAS-4 and, in its present form, was release in 2008. Although widely used (332 links on PubMed and 57 studies listed on ClinicalTrials.gov [26 August 2023]), it has come in for sustained criticism with some 20 retractions of papers and most recently the original article removed, after sustained effort and an editorial glacial response, from the *Journal of Clinical Hypertension* [18 August 2023] [2]. In part the retraction statement said:

> Following publication, concerns were raised by a third party regarding the statistical analysis presented in the article. The Journal conducted an independent statistical review of the article and concluded that the results were misleading due to issues regarding the sensitivity and specificity of the medical adherence scale used. The authors responded to the Journal's request to address the findings of the independent statistical review, but were unable to adequately address the concerns. As a result, the Journal no longer has confidence in the reported conclusions and is issuing this retraction.

The purpose of this brief note is to make the case that had the reviewers of the MMAS-8 (and MMAS-4) understood the importance of applying the Rasch mathematical model to the development and application of this instrument; it should never have been proposed in the first place. If the Rasch model for creating a single attribute or unidimensional, linear, interval and invariant scale had been recognized in medication adherence, then the extended debate over retraction of papers incorporating the MMAS-8 scale, over some 15 years, would have been not only unnecessary but irrelevant [3]. The MMAS-8 item scale would have been a totally different instrument. In order to create an instrument to capture medication adherence we have no option but to apply, to the respondent counts from the questionnaire, a transformation to move from ordinal observations to a calibrated measure with interval or ratio properties [4]. The object should have been to apply the necessary and sufficient requirements of the Rasch model to transform ordinal counts to a linear measure before proceeding to a further analysis. This has been known for almost 70 years with an extensive global following of Rasch in academic institutions. Instead, we have hundreds of PRO instruments applied in clinical trials and HTA, including medication adherence claims, which are nothing more than integer counts that support only non-parametric statistics. If we want to apply classical parametric statistical techniques to assess medication adherence, then we need an interval or ratio measure; a calibrated measuring system with a well-defined origin and unit of measurement. The MMAS-8 fails to provide such a measure which must cast doubt on claims from manufacturers that competing products yield greater medication adherence or, indeed, for the prevalence of non-adherence in disease and target patient groups.

THE RASCH MODEL

The Rasch model represents a probabilistic framework for assessing the likelihood of a positive response to a questionnaire item [3]. The likelihood of success is defined, for dichotomous responses, as a function of the difference between item difficulty and respondent ability. In medication adherence there is one Rasch model that has been proposed: the Probabilistic Medication Adherence Scale (ProMAS) [5] . The latent construct of interest is medication adherence disposition which was manifested as a single attribute with 18 dichotomous item statements. The development of the instrument followed, applying the WINSTEPS software package, item selection (18 items selected from an initial list of 37), with a sample response to identify item difficulty and respondent ability scores transformed to a logit scales. With the average logit set arbitrarily at zero, with positive logits indicating higher than average probabilities and negative logits lower than average. Starting with the first round of item estimates in logits produces a first round of person adherence estimates, these are iterated as conditional maximum likelihood estimation against each other to meet a preset convergence value for a common logit scale. These transformed logits are represented on a real number line with linear and interval properties with adherence or person free and item free calibrations to compare persons and items directly. In the ProMAS instance the logits for the real number line are for person adherence (item response) and item difficulty, with a difference between difficulty and adherence in mean logit values of 1.42 logits.

The adherence logits can be translated to interval logit scores to assign adherence categories for low <1.68 logits, medium-low -1.29 to -0.03 logits; medium-high 0.34 to 1.69 logits and high

>2.12 logits medication adherence categories. Respondent distribution was respectively 4.05%, 16.0%, 46.8% and 33.3%.

Presenting outcomes in terms of a logit scale with the appropriate measurement properties, is not an obvious metric for assessing medication adherence in distributional terms. A more intuitively acceptable approach, and one that is not emphasized in the Rasch literature, is to translate the logit values to their corresponding odds ratios and proportions. This is accomplished, recognizing that a logit is the natural logarithm of an odds ratio (p/1-p) by a logistic transformation (e.g., a logit of 1.3 is transformed to p = 0.79). This will allow us to operate on a linear scale to create estimates of the individual respondent and overall medication possession for the target group and appropriate measures of dispersion. We can also focus on response to medication adherence interventions to assess the significance of mean differences and effect size where the instrument is applied to assess the impact of interventions to improve medication adherence, measured on a possession scale calibrated as probabilities.

The procedure, detailed in a recent study [6] involves four steps: (i) transform the logit values for each item to a probability (or weight); (ii) with dichotomous responses, estimate for each respondent the sum of weights for each positive item response; (iii) take the ratio of the count of overall possible item responses fir each respondent or the sum of the probability weights divided by the sum of weights for successful responses and apply this for each respondent to create a medication adherence proportion; and (iv) utilize the distribution of medication proportions to estimate the mean medication adherence proportion for the respondent group, together with other descriptive statistics. If the mean values are estimated before and after an intervention to improve medication adherence, we can estimate the significance of the change and the effect size. This transformation is equally applicable to instruments with polytomous responses following application of either the Rasch Rating Scale Model or the Partial Credit Rasch Model [3].

The advantage of this approach to reporting medication adherence is that we have a measure which meets the required Rasch standards for a unidimensional, linear, interval and ratio scale, and one that allows a summary single metric for medication adherence as the proportion of the latent trait possessed by respondents. This provides a far more robust base for evaluating the extent of medication adherence as a continuous variable This allows us to reject cut off points such as the 6-point Morisky-8 criterion for low adherence (the basis for questionable sensitivity and specificity claims as noted in the retraction) as adding integer scores where they are only ordinal counts is meaningless; although the implicit assumption is that the 8-tem integer scale actually has ratio properties. If we are to assess the determinants of medication adherence or intervention impacts, the proposed Rasch logit transformation is not only a considerably more information rich basis for assessment but is one that meets the required measurement standards. Certainly, cutoff points can be applied to support matching to assess sensitivity and specificity, but this seems pointless when we can evaluate the individual respondent possession estimates against scales for other medication counts at the individual level.

COMPETING MEDICATION ADHERENCE INSTRUMENTS

The ProMAS has not been used wisely in medication adherence; in fact, there are only a handful of references and one proposed application in a clinical trial NCT02488343 *Profile of Adherence to Therapy and Interventions to Promote Adherence in Multiple Sclerosis* [7] [8] [9] [10]. This is not unusual given the extent to which instruments such as the MMAS-8 have become embedded in the literature and the lack of interest and understanding of Rasch measurement; not only in the area of medication adherence but more widely in health technology assessment. This is unfortunate because if we are looking to a gold standard for measuring medication adherence, the Rasch standards are the unique tool for transforming counts or observations to linear and ratio scales; Rasch is the necessary and sufficient means for achieving this because it recognizes the importance that in any science or social science original observations are not yet measures [4]. It has been long recognized that measurement implies the *previous construction and maintenance of a calibrated measuring system with a well-defined origin and unit which has been shown to work well enough to be useful* [4].

A key concept in Rasch measurement is the standard for invariance: an instrument must not have only linear and interval properties but must be for a single attribute or dimension. Invariance in the context of Rasch means that item and person parameters must remain the same across all appropriate applications. In other words, measures attributable to variables by any measurement system, must be independent of the instrument used; calibrations must remain invariant. In any one context calibrated devices must yield the same reading and the readings must remain invariant across all contexts [3].

Invariance means unidimensionality; we can only measure one attribute or manifestation of a latent construct at a time. The instrument must be designed to measure that single attribute. It is this requirement that gives us linear and interval measures of an underlying latent construct. It was not, as far as can be ascertained, an objective in the development of the MMAS-4 and MMAS-8 instruments.

Item selection is the critical first step; item selection and successive reassessment of items must be in terms of the single attribute we are proposing to measure; we cannot measure a latent construct directly as it is an abstract concept such as quality of life; what we are attempting to do is to create a measure that is validated by conformity to Rasch standards, notably fitting items to the Rasch model. This is what the ProMAS has achieved. This is not achieved by attempting to fit Rasch criteria to an existing item list defining a medication adherence instrument; the starting place must be item selection. There is no reason why, in medication adherence, complementary or competing Rasch models could not be developed as long as the standards are articulated.

SYSTEMATIC REVIEWS

Many systematic reviews leave a lot to be desired in terms of their coverage and procedures in place to identify studies from predatory journals, many originating from paper mills, that may be considered misleading if not supporting false claims. Systematic reviews in medication adherence leave a lot to be desired. Two features stand out: first, a limited selection of instruments and their application and, second, a failure to recognize the unique status of Rasch measurement to support

instrument development and claims. A recent review of 59 medication adherence instruments evaluated the extent to which the items selected for the instrument could be classified as intentional as opposed to behavioral, which included the ProMAS instrument. The assessment failed to detail how the items were selected and the importance of Rasch as the only basis for item selection [11]. In the assessment of psychometric properties, no account was taken of the measurement properties of the instrument and whether the items selected made any contribution to the objective of creating a unidimensional, linear, interval or invariant measure. The fact that the ProMAS instrument was unique in applying Rasch standards where the items were selected to meet the Rasch model was not apparently of interest. This is important given the unique status of the ProMAS instrument (alone of the 59 instruments) in recognizing the important of fundamental measurement in evaluating medication adherence.

Another recent systematic review, this time of 121 medication adherence instruments from 214 studies also failed to address the key question: what are the measurement standards for medication adherence [12]. The ProMAS instrument was included but was not considered for a detailed assessment of its measurement properties. Rasch was not mentioned as a measurement standard; instead, the study determined there was no gold standard instrument or otherwise for evaluating the methodological quality of any medical adherence instrument. There was no mention of Rasch measurement standards as the accepted gold standard in fundamental measurement, including medication adherence.

A final review that is worth citing considered the validity of the MMAS 8-item instrument as a screening tool for medication adherence [13]. The problem with the analysis is that it failed to recognize the fact that integer values attached to question item yield only an ordinal scale. The assignment of a 6/8 cutoff, the MMAS-8 definition for low adherence as the basis for a binary assessment of specificity and sensitivity (adherent/not adherent) is allowed but overlooks a fundamental question: what is the item? Before applying a cutoff, we need to specify the item order (a foundation of Rasch analysis). Are all items equally difficult with a selection that is random? If this is the case why have eight, four or only one item? If the items are not equally difficult then the cutoff must be redefined to indicate which items comprise the cutoff. But this approach then raises questions about respondent characteristics? Are respondents equally able to respond to these items or, as in Rasch, will some be less likely to respond to some items that others when items are ordered by their difficulty as manifesting the latent trait of adherence behavior? Following Rasch, is the probability of a successful response a function of the difference between respondent adherence ability or behavior and the difficulty of the adherence item? If this is the case then a cutoff is denied until Rasch analysis has been applied to transform an ordinal scale, recognizing the need for item and respondent selection, to a single attribute, linear, interval and invariant measure. This raises once again the specter of retraction: if it is based on presumed concerns with the cutoff and claims for sensitivity and specificity, then it is focusing on the wrong target.

REJECTION PRECEDES RETRACTION

While the Morisky scales is under a cloud given the withdrawal by the journal in which it first appeared in 2008, due to concerns with the application of sensitivity and specificity of the claimed

responses and associated claims for accuracy, the various critiques and applications of the scale have overlooked a key issue: measurement properties

Despite foot dragging by authors, academic colleagues, heads of departments, Deans, university administrations, journal editors, publishers and review groups, the groundswell for increased and speedier restrictions is increasing – albeit from a relatively low base - with increased attention given to ringfencing the lucrative fields of paper mill production and predatory or just simply, dodgy, journals. The Morisky case is merely one example; after some 15 years the 'foundation' article has been retracted, although the process might be best described as getting blood out of a stone.

There is no doubt that the proportion of retractions is increasing, but this overlooks a key consideration regarding Rasch measurement: to what extent do reviewers take account of the measurement standards of a scale? The MMAS-8 scale, irrespective of the debate over the interpretation of dichotomized sensitivity and specificity estimates, still fails at a fundamental level: it is only an ordinal integer summation scale. It is not a measure in the sense of meeting Rasch standards for a single or unidimensional attribute with linear, interval and invariant properties. To achieve this requires the application of the Rasch model, starting with item selection.

Failure to do this, as noted above, means that the debate over the statistical properties of the MMAS-8 is irrelevant. In common with all medication adherence measures it is just an integer summation of item responses, not measurement; indeed, the items are not even ranked independently of the response to each question. If it was a ranked ordinal scale then we are limited to non-parametric statistics. It cannot support claims for medication adherence nor as a tool to evaluate the determinants of or changes in adherence following behavioral or other interventions. Applications of standard statistical tests are only possible if the measure of interest has single attribute, linear, interval and invariant properties [3]. Absent Rasch, there is no medication adherence measure that can support such assessments. Given the widespread application of the MMAS-8 instrument, irrespective of the reasons for its retraction it has in common with dozens of other instruments failed to rise above being only an ordinal scale.

The question that reviewers and journal editors have to address is how to handle value claims for medication adherence that fail to meet the required Rasch or modern measurement standards. The straightforward answer is that any research paper that is based on measures which fail the required standards should be rejected; supported by a retraction policy that flags all previously published studies. Rejection, as a standard based on the Rasch or modern measurement theory, must by an integral part of peer review. Once the conformity with Rasch transformation has been successfully demonstrated then the merits in terms of application and claims from the application of statistical techniques can be evaluated. The Rasch transformation, it must be emphasized, may fail to meet the required number line standards. This points to the importance to the need that if the intent is to manifest an attribute from a general latent construct such as 'compliance and adherence', then the development of the instrument to capture the manifest trait or property of interest must start with the patient and item selection that anticipates respondent adherence and item difficulty.

CONCLUSIONS

Apparently, in the US, 30% to 50% of medications are not taken as prescribed [14]. This is claimed to be associated with 125,000 excess deaths, 10% of hospitalizations and $100 billion in annual health care costs. One of the most common methods of identifying patients at risk for poor adherence in randomized controlled trials for improved adherence is self-reporting. If this is the case; then we must invest in instruments that meet standards for Rasch or modern measurement that identify such patients and provide a measure of response to competing interventions. This is not achieved with the MMAS-8 instrument. Unfortunately, the MMAS-8 is not alone in this regard.

The MMAS 8-ltem instrument joins all other medication adherence instruments, with the exception of the Rasch ProMAS instrument ,in failing fundamental measurement standards. None of these instruments has created a scale that focuses on a single attribute (or is unidimensional) with linear, interval and invariant properties. They are ineligible for use in clinical trials as they cannot measure medication adherence or as a predictor of adherence behavior. Certainly, they have been applied in numerous clinical trials and observational studies but this, in retrospect, has been a wasted effort.

The eventual retraction of the MMAS-8 paper is, of course, a salutary lesson in misleading claims; although, to be honest, with the number of studies utilizing it over some 13 years, the horse has well and truly bolted. Whether the other item or integer scored instruments would experience the same fate is unknown. What is clear, however, is that the attention to medication adherence, the proliferation of studies, have failed the Rasch test for measurement. With the exception of the ProMAS we have no measure of item integer-based instrument claims that are acceptable. The puzzling feature is that while Rasch standards and the requirements for fundamental measurement have been accepted for almost 70 years, the leaders in medication adherence studies appear unaware of this fundamental measurement standard.

REFERENCES

[1] Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. *J Clin Hypertens (Greenwich).* 2008; 10(5):348–54 [Retracted].

[2] Retraction Watch. Paper that helped form basis of pricy research tool retracted. August 24, 2023. https://retractionwatch.com/2023/08/24/paper-that-helped-form-basis-of-pricy-research-tool-retracted/

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

[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] Kleppe M, Lacroix J, Ham J et al. The development of the ProMAS: A probabilistic medication adherence scale. *Patient Preference and Adherence*. 2015;9:355-367.

[6] Langley P. Integers, linear transformations, logistic transformation and valid claims for therapy response. *Maimon Working Papers* No 12 July 2023 https://maimonresearch.com/maimon-working-paper-no-12-august-2023

[7] Aluzaite K, Braund R, Seeley L et al. Adherence to inflammatory bowel disease medications in southern New Zealand. *Crohn's & Colitis 380*. 2021; 3,1-08

[8] Green J, Zerhouni O, Gauchet A. Contrasting implicit and explicit measures of attitudes to complementary and alternative medicines. *Complement Ther Clin Pract*. 2019;34:105-108

[9] Vluggen S, Hoving C, Schaper N et al. Psychological predictors of adherence to oral hypoglycaemic agents: an application of the ProMAS questionnaire. *Psychol Health*. 2020;35(4):387-404

[10] Neter E, Glass-Marmor L, Wolkowitz A et al. Beliefs about medication as predictors of medication adherence in a prospective cohort study among persons with multiple sclerosis. *BMC Neurology. 2021;21:136*

[11] Fahrni M, Saman, Alkhoshaiban A et al. Patient-reported outcome measures to detect intentional, mixed or unintentional non-adherence to medication: a systematic review. *BMJ Open*. 2022; 12:e057868 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9486224/pdf/bmjopen-2021-057868.pdf

[12] Kwan Y, Weng S, Loh D et al/ Measurement properties of existing patient-reported outcome measures on medication adherence: Systematic review. *Med Internet Res*. 2020;22:10 e 191679 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7584986/

[13] Moon SW, Lee W-Y, Hwang J et al. Accuracy of a screening tool for medication adherence: A systematic review and meta-analysis of the Morisky Medication Adherence Scale – 8. *PLoS One*. 2017; 12 (11) e0187139

[14] Kini V, Ho P. Interventions to improver medication adherence: A review. *JAMA*. 2018;320 (23):2461-2473