Conditional Maximum Likelihood Rasch Model in Data Harmonization

Yan Wang* Heidi M. Crane† Honghu Liu‡

Abstract

Data from different studies often have large variability and data collected with various instruments usually have low comparability, even if they are in attempt to measure the same concept or construct. Pooling individual data is scientifically and technically very challenging. It requires the generation of harmonized datasets across studies. Data harmonization aims to promote common measure for the key indicators that can permit certain degrees of comparability over time and across studies. This common measure will be used to combine the datasets and therefore to increase the sample size and to allow for adjustment of confounding factors. We will review the statistical methods that will accommodate these differences to create the common latent trait to harmonize the measures. The conditional maximum likelihood estimation of Rasch Model has been identified to create the latent trait measure of multiple items of self-reported adherence. Finally, this method will used on a real data in practice to create the harmonized measures across different studies.

Key Words: Harmonization, Response Conversion, Self-reported, Item Response Theory, Rasch Model

1. Introduction

Harmonization is a process composed of a series of complementary steps which must be applied with rigorous procedures and decision-making in order to ensure validity and reproducibility of harmonization outputs (Griffith et al., 2013). Harmonization of data is the usual way to improve the comparability between studies. If harmonization cannot be done, then the options are restricted to either making unwarranted assumptions about the data, or not doing any comparison at all (Van Buuren et al., 2005). Data harmonization is limited as a method for achieving comparability between studies. Pooling data together using appropriate methods is able to generate rich and available data for new hypothesis. For example, in the Comparison of Longitudinal European Studies (CLESA), the cross-national longitudinal data on health and functioning among older people are available after harmonized the data from six studies (Minicuci et al., 2003). However, different methodologies and sampling techniques used in the different studies need a series of decisional strategies for the preparation of unbiased comparison across studies (Griffith et al., 2013).

Several methods of data harmonization have been proposed in the literature. They can be categorized into three classes (Ma, Raina, Griffith, 2013). The first class of methods are relative simple and straight forward. It creates a common metric for combining constructs measure using different scales. It creates a common metric for combining constructs measure using different scales. It directly operated on the original measures or items. By certain conversion formulas of monotone functions, into a comparable measure of the original variables. For example, re-categorization, Z-score transformation, and percentile conversion are all belong to this class of method (Ma, Raina, Griffith, 2013). The second class of methods involves using multiple imputation techniques. The imputed value or the estimated

*Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, CA 90095

†Center for AIDS and STD Research, University of Washington, Harborview Medical Center, Seattle, WA 98104

‡Division of Public Health and Community Dentistry, UCLA School of Dentistry, David Geffen School of Medicine, Fielding School of Public Health, Los Angeles, CA 90095

value of the missing item are used to replace the missing item. The method is efficient for up to 50% missing data, for certain group of sample (Ma, Raina, Griffith, 2013). The third class of methods assumes that there is a latent factor. The measures from the different data sets are measuring the same latent variable. The term latent means the true value for the item is unknown, but can be observed through some responses of the items. Ma, Raina and Griffith use a simulation study to compare the performance of these harmonization methods (Ma et al., 2013). Item response theory (IRT) models have been used in health status measurement and evaluation of Patient-Report Outcomes (PROs). In report of the health status, the polytomous items are common. Polytomous Rasch model is commonly used in psychometric to find the latent common measure.

Adherence refers to taking the medication exactly as prescribed. Adherence is crucial for the success of HIV viral load suppression and maintain the viral load level undetectable. There are many types of measures to evaluate the adherence level. Each has both advantage and disadvantage on accuracy, attainability, cost and effectiveness. Among all types of measures, self-reported measure is the most convenient way to collect patient's adherence of medication. However, due to recall intervals and the report measures issues, the self-reported adherence has a lot different way to collect, for example, one day recall, two days recall, up to one week recall.

This paper will focus on the PRO measures, self-reported adherence measure, which is the simplest way to collect the adherence directly from the report of the patients. The recall intervals have great impact on the accuracy and the validity of the measures. This paper will incorporate a statistical methods to calibrate the different types of self-reported measures into a latent trait measure of self-reported adherence. The harmonized measure will increase the comparability among different types of self-reported measures. This latent measure can be used for further analysis to predict the viral load and to compare the effectiveness of intervention program that focused on adherence improvement.

In section 2, we will talk about the details of polytomous Rasch model and the conditional maximum likelihood estimation developed by Christensen in 2013. The response conversion for the final latent trait of self-reported adherence will be reported. In section 3, the methods are applied to the self-reported medication adherence data. The conclusion of the calibrate measure using IRT theory to harmonize the measures of self-reported medication adherence is in section 4.

2. Methods

2.1 Polytomous Rasch Model

The Rasch model was named after Georg Rasch (Rasch, 1960). The dichotomous version (Rasch, 1960) and Polytomous version (Andrich, 1978) are both available for Rasch model. The Polytomous Rasch Model (PRM) using conditional maximum likelihood estimation (CMLE) is well developed (Christensen, 2013). Consider the number of items in is I, and item i has the response categories from $0, 1, \cdots, m_i$, where $i = 1, \cdots, I$. In this case, item i has $m_i + 1$ categories and $x_i \in \{0, 1, \cdots, m_i\}$. We will use θ to denote the latent variable - true adherence. We will assume $\theta \in [0, 1]$. The Polytomous Rasch Model is given by,

$$p(X_i = x_i|\theta) = \frac{\exp(x_i\theta + \eta_{ix_i})}{\sum_{x_i=0}^{m_i} \exp(x_i\theta + \eta_{ix_i})} \tag{1}$$

Where η_{ix_i} is the parameter in the model. Here we always let $\eta_{i0} = 0$.

An alternative but equivalent way to parameterizing the model into a dichotomous

Rasch (Christensen, 2013) model is to calculate the conditional probability,

$$p(X_i = k|X_i \in \{k-1,k\},\theta) = \frac{\exp(\theta - \beta_{ik})}{1+\exp(\theta-\beta_{ik})} \tag{2}$$

Here β_{ik} is easy to interpret as the location on the latent continuum scale, where the probability for item i is the same for choosing category k and choosing category $k-1$ (Christensen, 2013). Usually β_{ik} is called the threshold parameter and $\beta_{ik} = -(\eta_{ik} - \eta_{ik-1})$.

2.2 Conditional Maximum Likelihood Estimation

To derive the Conditional Maximum Likelihood Estimation (CMLE) is based on conditional on the summation score of the response among all the items space. The joint log likelihood function for a sample of subjects $v = 1,\cdots,N$ is given by (Christensen, 2013),

$$l(\eta_1,\cdots,\eta_I;\theta_1,\cdots,\theta_N) = \sum_{v=1}^{N}\sum_{i=1}^{I}(\theta_v X_{vi} + \eta_{ix_{nui}}) - \sum_{v=1}^{N}\log K \tag{3}$$

Where

$$K = \prod_{v=1}^{N}\prod_{i=1}^{I}\left(\sum_{x_{vi}=0}^{m_i}\exp(x_{vi}\theta_v + \eta_{ix_{vi}})\right)$$

Here X is the vector form of $X_1,\cdots,X_I$. The number of estimators will be increased as the sample size increased. Therefore, estimation of all the parameters is not consistent. The main focus is to estimate the item parameter for the person. Note from the above equation, the total score from all items is sufficient to estimate the location θ_v for each person.

The joint likelihood can also be written as (Christensen, 2013),

$$\begin{aligned} l(\eta_1,\cdots,\eta_I;\theta_1,\cdots,\theta_N) &= \sum_{v=1}^{N}\sum_{i=1}^{I}\left(\theta_v X_{vi} + \sum_{h=1}^{m_i} 1_{(x_{vi}=h)}\right) - \sum_{v=1}^{N}\log K \\ &= \sum_{v=1}^{N}\theta_v\sum_{i=1}^{I}X_{vi} + \sum_{v=1}^{N}\sum_{h=1}^{m_i}\sum_{i=1}^{I}X_{vi}1_{(X_{vi}=h)} - \sum_{v=1}^{N}\log K \end{aligned} \tag{4}$$

As the conditional maximum likelihood (CML) inference is based on conditioning on the summation score on the response vector for each person. For vector $X_v = (X_{v1},\cdots,X_{vk})$, where $k = 1,\cdots,m_i$, from the Rasch model, the distribution of the summation score $R_v = \sum_i X_{vi}$ is given by (Christensen, 2013),

$$P(R_v|\theta) = \frac{e^{r\theta}}{\prod K_i(\eta_i,\theta)} \times \gamma_r \tag{5}$$

Here the set $X^{(r)} = \{X|\sum_i X_i = r\}$ is the over all the response space that has $R_v = r$. The notation γ_r is,

$$\gamma_r = \sum_{X_v \in X^{(r)}}\exp(\sum_{i=1}^{k}\eta_{ix_i})$$

Therefore the conditional likelihood given the summation of the score is r can be estimated consistently. For each individual, the probability for the item score is,

$$P(X_v|R_v = r,\theta_v) = \frac{\exp(\sum_{i=1}^{k}\eta_{ix_i})}{\gamma_r} \tag{6}$$

Therefore the conditional likelihood can be written as,

$$\begin{aligned} L_C(\eta_1,\cdots,\eta_k) &= \prod_{v=1}^{N} \frac{\exp(\sum_{i=1}^{k}\eta_{ix_{vi}})}{\gamma_{r_v}} \\ &= \frac{\prod_x \exp\left(n(x)\sum_{i=1}^{k}\eta_{ix_i}\right)}{\prod_r \gamma_r^{n(r)}} \\ &= \frac{\prod_{v=1}^{N}\exp\left(\sum_{i=1}^{k}\sum_{h=1}^{m_i} X_{vih}\eta_{ih}\right)}{\prod_r \gamma_r^{n(r)}} \end{aligned} \tag{7}$$

The response vector is denoted as $x=(x_1,\cdots,x_k)$ and the number of people with response as x is denoted as $n(x)$. The number of people with summation score r is denoted as $n(r)$. The conditional log likelihood function for η is,

$$l_C(\eta_1,\cdots,\eta_k) = \sum_{i=1}^{k}\sum_{h=1}^{m_i}\sum_{v=1}^{N} X_{vih}\eta_{ih} - \sum_{r=0}^{km} n(r)\log(\gamma_r) \tag{8}$$

The summation $\sum_{v=1}^{N} X_{vih}$ is sufficient statistics (called item margin) to estimate the parameter (Christensen, 2013). This can be estimated by the number of persons in category h for item i. The expected value of the summation of items has the form,

$$\begin{aligned} E(\sum_{v=1}^{N} X_{vih}|R_v) &= \sum_{v=1}^{N} P(X_{vi}=h, R_v=r_v) \\ &= \exp(\eta_{ih})\sum_{r=0}^{km_i} n(r)\frac{\gamma_{r-h}^{(i)}}{\gamma_r} \end{aligned} \tag{9}$$

Here expected value is calculated by conditioning on the summation of all the items. This can be estimated by generalized linear model.

2.3 Response Conversion

We are more interested in estimating the location of the latent variable θ. This variable is being derived as the continuous scale. There are two types of methods to estimate the location. The variable will be rescaled to be within $[0,1]$ to represent the relative self-reported adherence location. The likelihood estimation equation for θ is given by,

$$E(R\theta) = \frac{\partial}{\partial\theta}\left(\log(\sum_r \exp(r\theta)\gamma_r\right) \tag{10}$$

We can use Newton-Raphson algorithm to estimate the MLE of θ. The score is a monotone increasing function of θ. In the MLE method, the probability of attain the maximum and minimum of the score is set to be $-\infty$ and $+\infty$. Another estimation method is the Bayesian model estimator. A special case of the model estimation of Bayesian models is Weighted likelihood estimation (wle). By choosing the appropriate prior for θ, we can estimate the parameter by maximizing the posterior density. The posterior density of θ is given by,

$$P(\theta|X) = \frac{p(X|\theta)p(\theta)}{\sum_\theta p(X|\theta)p(\theta)} \tag{11}$$

The construction of the conversion key can be derived to create the common measure based on the items.

3. Results

MACH 14 study is Multi-site Adherence Collaboration in HIV among 14 sites in United States. MACH14 studies has the following features. Eligible studies were required to have: (1) a longitudinal study design with at least 3 repeated measurements; (2) MEMS adherence data; (3) VL and clinical outcomes; and (4) psychosocial and behavioral measures. The details of the study is described elsewhere (Liu et al., 2013).

Self-reported adherence measures have been most commonly used among HIV patients to assess their level of adherence. They are valued for their convenience and practicality. However, scale items and methods of summarizing and analyzing data have varied considerably from one study to another. Although self-reported adherence usually correlates with virologic outcomes in the expected direction, it is generally found to overestimate adherence.

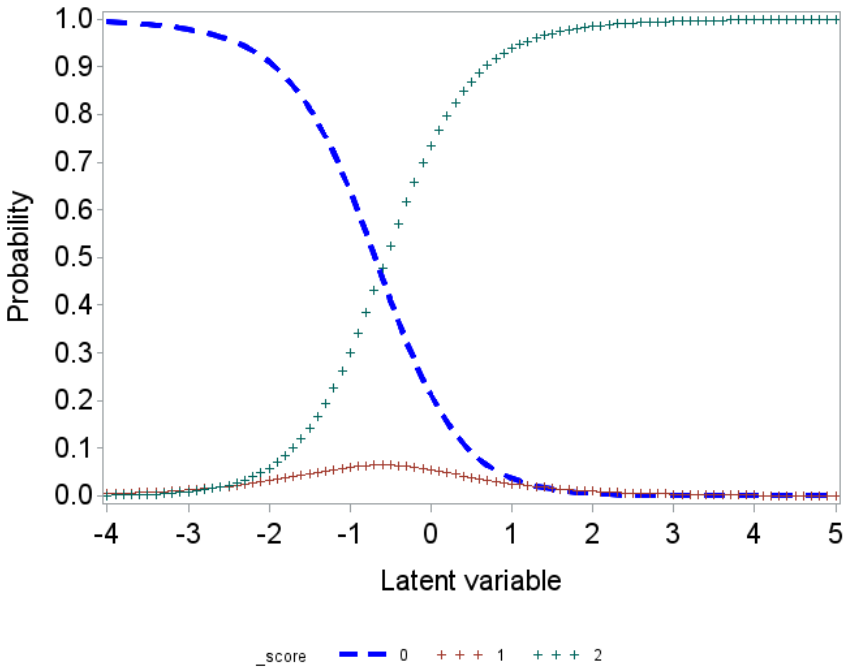
In MACH14 study, we have collected rich data to measure adherence. The recall interval on 1 day, 2 days, 3 days, 4 days, and one week has been reported from different studies. Although not all the studies have collected all the five different recall intervals of the self-reported adherence, we have these measure available for the first time to compare and calibrate the self-reported adherence measures. We categorize the continuous self-reported adherence in the data into three categories,

- 0 if the self-reported adherence $\leq 50\%$
- 1 if the self-reported adherence $(50\%, 85\%)$
- 2 if the self-reported adherence $\geq 85\%$

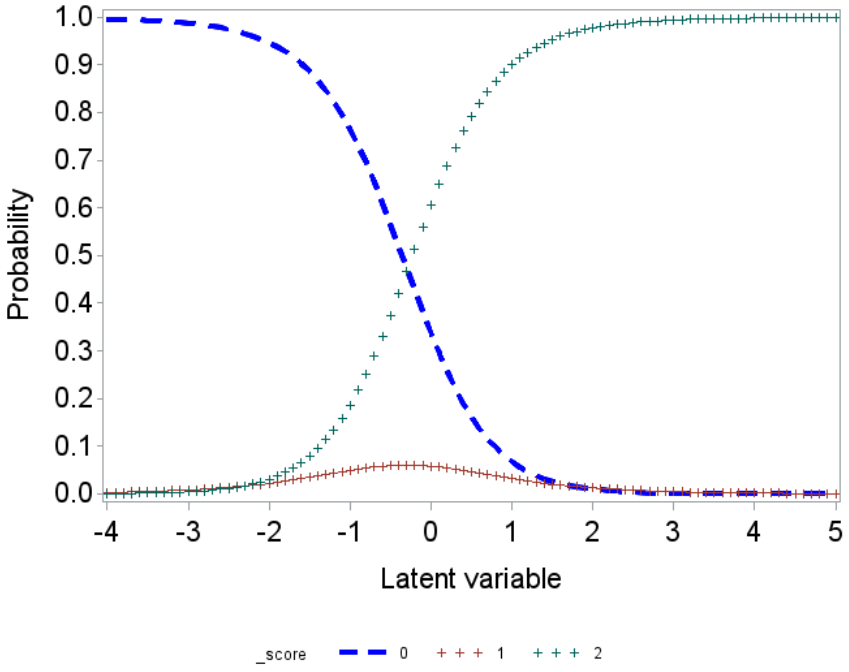
The Item Characteristic Curves (ICCs) for different recall intervals of self-reported adherence are shown in the below Figures respectively. In each figure, three curves are shown for those whose answer in corresponding categories. For the item with score equals 0, the probability for a person with lower latent trait of self-reported adherence with answer in this category for each item is very high. In the contrast, if the latent trait of the patient is very high, the probability for the person whose item response is 2 is very high. The probability for the person whose answer for the item is 1 increase when the latent trait of self-reported adherence is low, and decrease when the latent trait of self-reported adherence is high. The patterns of all the five different response intervals are very similar.

Compare the five figures together, the cross of response 0 and response 2 starts to shift to the right from the below three days and above four days of the recall intervals. This indicates that the higher latent trait of self-reported adherence is less sensitive for those who have less than or equal to three-day recall intervals. This is consistent with the literature that there is a cut-off below and above three days (Simoni et al., 2006).

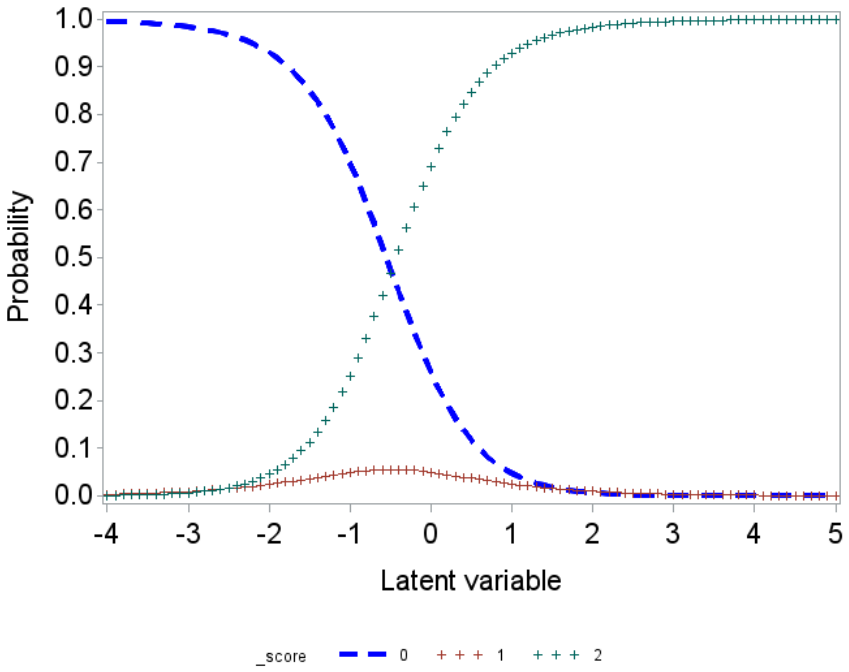
CML (cml). Category Probability Curves: catSR1d



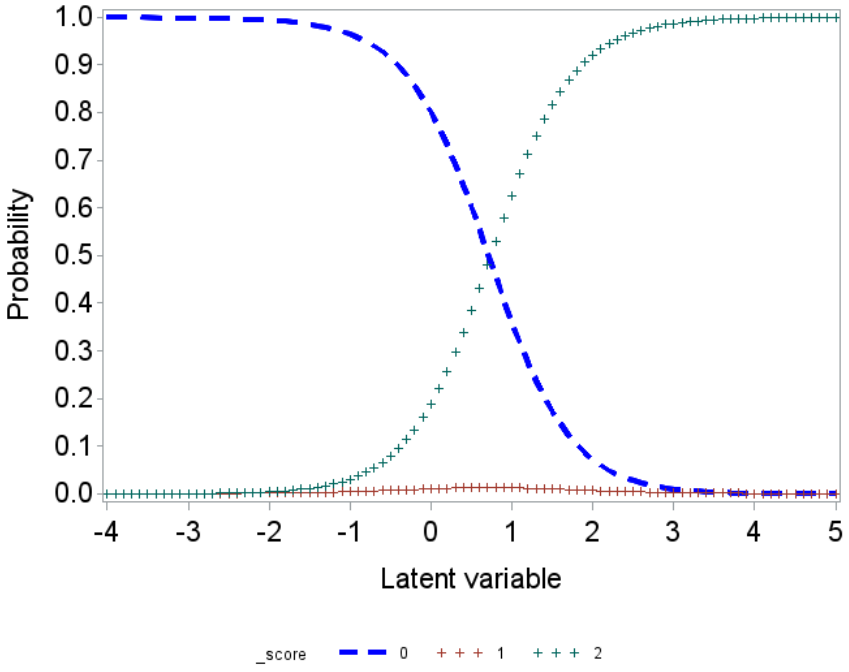
CML (cml). Category Probability Curves: catSR2d

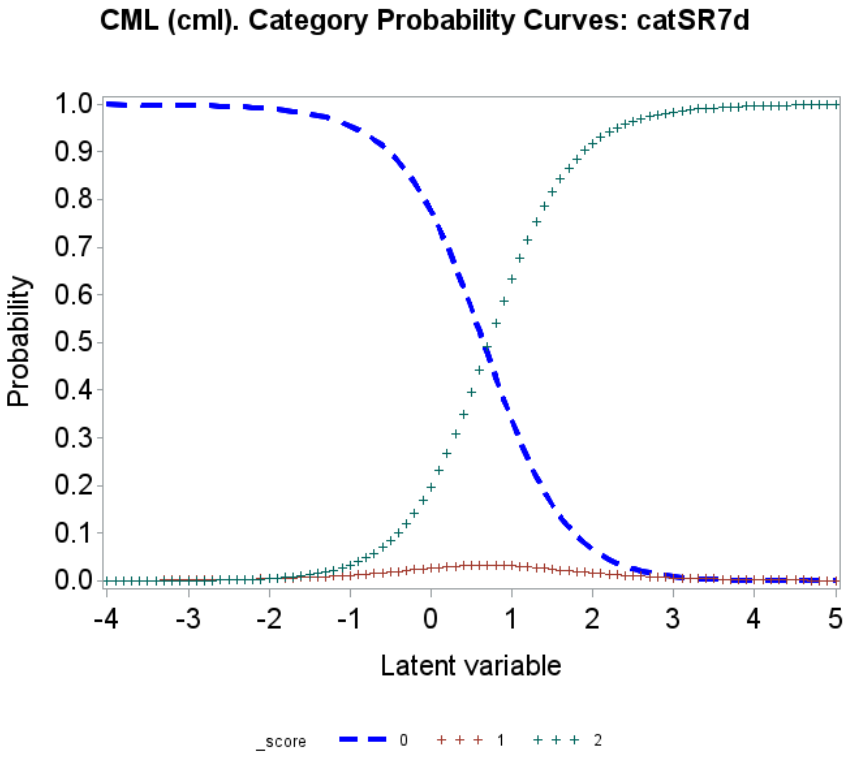


CML (cml). Category Probability Curves: catSR3d



CML (cml). Category Probability Curves: catSR4d





The conversion key for construct the latent trait θ can be estimated from Table 1, the score is the summation of the response from the five self-reported adherence measures in MACH14 data sets. The recall intervals are from 1 day, 2 days, 3 days, 4 days, and one week (7 days). The range for the summation of these five different recall periods is from 0 to 10. The higher summation score of all five items indicates that the larger chance for the latent trait on a higher scale. Those who answered all the five scales with 2 will have a summation score of 10 and will have 100% percent of adherence on a re-scaled latent trait as shown in Table 1.

The standard error is also reported in Table 1 for the converted latent trait θ and rescaled adherence level. It shows that the variance is relative large for those who anwered all the five items 0 and for those who answered all the items 2. This means those who have 0% adherence and those who report have 100% adherence have larger variability than those who have moderate adherence.

4. Discussion

Self-reported adherence is the most convenient way to obtain the adherence measure among HIV patients by directly asking the patients themselves. Most adherence measures are questioned for the validity and accuracy due to the different recall intervals and the response tasks. Lu et. al have concluded that the one-month recall period is optimal and accurate regarding to less overestimation (Lu et al., 2008). They did not only compare the recall periods of different self-reported adherence measures, but also concluded that items that ask respondents to rate their adherence on a six-point scale from very poor to excellent may be more accurate than those that ask about frequencies or percents. There are several studies have compared self-reported and MEMS adherence (Liu et al., 2001). Reynolds et al. compared the different recall intervals with MEMS adherence in MACH14 studies and concluded that the three-day recall interval is the most correlated with MEMS adherence in terms of percentage (Reynolds et al., 2013).

Table 1: Key construction for latent θ

score	MLE	se	WLE	se_WLE	Re-scaled θ	se_Re-scaled θ
0	$-\infty$	.	-1.58	0.82	0.00	0.26
1	-1.39	0.80	-1.10	0.60	0.15	0.19
2	-0.91	0.61	-0.79	0.53	0.25	0.17
3	-0.58	0.55	-0.52	0.50	0.33	0.16
4	-0.29	0.53	-0.27	0.50	0.41	0.16
5	-0.01	0.53	-0.02	0.51	0.49	0.16
6	0.28	0.54	0.25	0.52	0.57	0.16
7	0.58	0.57	0.53	0.53	0.66	0.17
8	0.93	0.62	0.81	0.55	0.75	0.17
9	1.41	0.80	1.14	0.61	0.85	0.19
10	$+\infty$	.	1.61	0.80	1.00	0.25

In this paper, we only calibrate the difference of varies of recall periods, from one day only, up to one week. We do not assess the different scenarios of ways asking the questions. Such as the frequencies missing of dosage or the actual percentage of self-reported of adherence. We use the average of medication adherence percentage among all the medications that the patient is currently taking. The one-day recall interval is usually more accurate but suffers from large variation for summarizing the overall adherence level. The one-month recall period for self-reported adherence is usually less accurate but with smaller variance for across different assessment over time.

We concluded from the category probability curve that there is a possible cut-off between the self-reported adherence with less than three days recall interval and more than three days recall. This is consistent with the literature that the self-reported adherence was associated with VL in 88% of recall periods that were greater than 3 days and in 64% of those that were 3 days or less (Simoni et al., 2006). The study found that the difference by the relation of self-reported adherence recall periods and viral load results.

In this study, we do not consider the longitudinal feature of the self-reported adherence. The longitudinal version of Rasch model for repeated adherence measures can be further studied to generate the latent measure over time. Therefore, this measure can be used to study the association with clinical outcome and MEMS adherence.

5. Acknowledgments

We would like to express our great appreciation to MACH14 investigators for their contribution to the rich data set.

REFERENCES

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561573.

Booth, J., and Hobert, J. P. (1999), "Maximizing Generalized Linear Mixed Model Likelihoods With an Automated Monte Carlo EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 61, 265–285.

Christensen, K. B. (2013), "Conditional Maximum Likelihood Estimation in Polytomous Rasch Models Using SAS" *ISRN Computational Mathematics*, vol. 2013, Article ID 617475, 8 pages, 2013. doi:10.1155/2013/617475

Griffith, L., van den Heuvel, E., Fortier, I., Hofer, S., Raina, P., Sohel, N., . . . Belleville, S. (2013) "Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis," Rockville (MD).

Liu, H., Golin, C. E., Miller, L. G., Hays, R. D., Beck, C. K., Sanandaji, S., Wenger, N. S. (2001). A comparison study of multiple measures of adherence to HIV protease inhibitors. *Annals of Internal Medicine*, 134(10), 968977.

Liu, H., Wilson, I. B., Goggin, K., Reynolds, N., Simoni, J. M., Golin, C. E., . . . Bangsberg, D. R. (2013). "MACH14: a multi-site collaboration on ART adherence among 14 institutions." *AIDS Behav*, 17(1), 127-141.

Lu, M., Safren, S. A., Skolnik, P. R., Rogers, W. H., Coady, W., Hardy, H., Wilson, I. B. (2008). Optimal recall period and response task for self-reported HIV medication adherence. *AIDS and Behavior*, 12(1), 8694. doi:10.1007/s10461-007-9261-4

Ma, J., Raina, P., Griffith, L. (2013). "Comparison of Statistical Harmonization Methods in Individual Participant Data Meta-Analysis," *Paper presented at the Joint Statistical Meetings 2013*, Montreal, Quebec, Canada.

Minicuci, N., Noale, M., Bardage, C., Blumstein, T., Deeg, D. J., Gindin, J., . . . Group, C. W. (2003). Cross-national determinants of quality of life from six longitudinal studies on aging: the CLESA project. *Aging Clin Exp Res*, 15(3), 187-202.

Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. *Danmarks Paedagogiske Institut*.

Reynolds, N. R., Sun, J., Nagaraja, H. N., Gifford, A. L., Wu, A. W., Chesney, M. A. (2007). Optimizing measurement of self-reported adherence with the ACTG Adherence Questionnaire: a cross-protocol analysis. *Journal of Acquired Immune Deficiency Syndromes* (1999), 46(4), 402409.

Reynolds, N., Wang, Y., Golin, C., Simoni, J., Goggin, K., Erlen, J., Wilson, I., Liu, H. for the MACH14 Investigators. A MACH14 cross protocol comparison of self reported and electronic antiretroviral (ARV) adherence: It's not only the type of measure used that influences adherence estimates. Poster presentation(Abstract ID 244), 8th International Conference on HIV Treatment Adherence, June 2-4, 2013, Miami, FL.

Simoni, J. M., Kurth, A. E., Pearson, C. R., Pantalone, D. W., Merrill, J. O., Frick, P. A. (2006). Self-Report Measures of Antiretroviral Therapy Adherence: A Review with Recommendations for HIV Research and Clinical Management. *AIDS and Behavior*, 10(3), 227245. doi:10.1007/s10461-006-9078-6

Van Buuren, S., Eyres, S., Tennant, A., Hopman-Rock, M. (2005). Improving Comparability of Existing Data by Response Conversion. *Journal of Ofcial Statistics*, 21(1), 53-72.