Performance of Five Equating Methods in Assessing Cognitive Impairment in Delirium Research

Jamey R. Guess, MS * Edward R. Marcantonio, MD, SM* †

Long H. Ngo, PhD* †

Abstract

The Mini-Mental State Examination (MMSE) is well accepted for cognitive assessment, while the newer publicly available Montreal Cognitive Assessment (MoCA) is growing in popularity. We examined several methods to obtain equivalence scores for these 2 assessments, both scored 0-30, 30 best. Mean, linear, equipercentile, circle-arc, and item response theory (IRT) equating procedures were used to derive MMSE-equivalent scores from MOCA scores in a study of 199 hospitalized older adults. All methods gave closed-form equating equations for point estimation. Standard errors for the circle-arc method were obtained via bootstrapping.

The estimated equivalence scores for all 5 methods were similar. The equipercentile method provided the best agreement with observed MMSE scores. In terms of precision, although the distributions of standard errors varied widely across the five methods, in the range of scores that most patients have (from 15 to 30), the standard errors were similar and small (within 1 point). Thus, for our data and from this evaluation of the 5 methods, equipercentile equating was the method of choice.

Key Words: Equating, MoCA, MMSE, Equipercentile, Item Response Theory

1. Introduction

Cognitive assessment is an important part of delirium diagnosis and many tools exist to determine a patient's cognitive impairment. A widely used global cognitive assessment tool is the Mini-Mental State Examination (MMSE). First introduced in 1975 and copyrighted in 2001, it has a total score ranging from 0 to 30 points. Scoring a 27 or higher indicates normal cognition and scoring a 26 or below indicates some impairment. Another tool for assessing cognitive impairment is the Montreal Cognitive Assessment (MoCA). Like the MMSE, scores range from 0 to 30 and has the cutpoint between normal cognition and some impairment at 27.

These assessments each have their strengths and with the growing popularity and open availability of the MoCA, it has become useful to be able to move from a MoCA score to an MMSE score without having to administer both. Many equating methods exist to create a crosswalk between tests. A crosswalk takes a score from one assessment and returns an equivalent score on the other assessment. However, it is unclear in our case which equating method is preferred. Therefore, our aim is to compare the accuracy and precision of five equating methods used to create a crosswalk between the MMSE and MoCA.

2. Methods

The data comes from a sample of 201 patients aged 75 and older, two of which did not receive the MMSE, giving a total sample size of 199. Each participant

*Department of General Medicine, Beth Israel Deaconess Medical Center

†Harvard Medical School

was given the MMSE followed by the MoCA. Mean, linear, equipercentile, circle-arc and observed score item response theory (IRT) equating methods were used to create the crosswalks. Standard errors were used to assess the precision of the estimates and a regression analysis of observed and equated MMSE scores with the intercept forced through the origin was used to measure the agreement between observed MMSE scores and MoCA-derived MMSE scores. Scores were equated with the mean, linear, equipercentile and circle-arc methods using the `equate (v.1.2-0)` package. The `kequate` package was used to implement observed score IRT equating. These equating methods are described briefly below and with more detail in Albano (2013), Kolen (2004) and Livingston (2009).

2.1 Mean Equating

The mean equating method considers the distributions of form X and form Y to differ only in their means. It takes a score from form X and adds a constant to obtain an equated form Y score. The constant is calculated as the difference between the mean score of form X and form Y. Therefore, given a score x_i from form X, the form Y equivalent is given as follows:

$$m_Y(x_i) = x_i - \mu_X + \mu_Y \tag{1}$$

The variance of this estimate with equal number observations for form X and form Y is:

$$var[\hat{m}_Y(x_i)] = \frac{\sigma_X^2 + \sigma_Y^2}{N} \tag{2}$$

2.2 Linear Equating

Similar to mean equating, the linear equating method also adjusts for X and Y distributions with unequal means. However, this method uses standardized scores, taking variances in to account as well. Using the linear equating method, a form Y score equivalent of a form X score of x_i is given as:

$$l_Y(x_i) = \sigma_Y \left[\frac{x_i - \mu_x}{\sigma_X} + \mu_Y \right] \tag{3}$$

The variance of this estimate requires the skewness and kurtosis of the two distributions. With N_X and N_Y observations from form X and form Y respectively, the variance is:

$$\begin{aligned} var[\hat{l}_Y(x_i)] = \sigma_Y^2 \Bigg\{ \frac{1}{N_X} + \frac{1}{N_Y} + \left[\frac{sk(X)}{N_X} + \frac{sk(Y)}{N_Y} \right] \left[\frac{x_i - \mu_X}{\sigma_X} \right] \\ + \left[\frac{ku(X)-1}{4N_X} + \frac{ku(Y)-1}{4N_Y} \right] \left[\frac{x_i - \mu_x}{\sigma_X} \right]^2 \Bigg\} \end{aligned} \tag{4}$$

2.3 Equipercentile Equating

Equipercentile equating makes no assumptions about the distribution of form X and form Y scores. This method works by matching on the percentile ranks of each distribution. Since these are discrete scores, log-linear smoothing is first applied to the distributions before the equating is done.

If we let $P_X(x_i)$ refer to the percentile rank of score x_i on form X and $P_Y^{-1}(p)$ refer to the inverse of the percentile rank function for form Y of percentile p, then the form Y equivalent of form X score x_i is:

$$e_y(x_i) = P_Y^{-1}(P_X(x_i)) \tag{5}$$

The variance of this estimate is omitted here but can be found on page 248 of Kolen (2004).

2.4 Circle-Arc Equating

The circle-arc method equates test scores by fitting an arc of a circle with radius r through two endpoints and a midpoint. The endpoints are set as the highest and lowest meaningful points, usually the points $(0, 0)$ and $(max(\mathbf{X}), max(\mathbf{Y}))$. Various methods exist for determining the midpoint but it is typically set to the point $(\bar{x}, \bar{y})$. With these three points, the radius and center point, (x_c, y_c), of the circle can be determined and a score of x_i equates to a form Y score as follows:

$$c_Y(x_i) = y_c \pm \sqrt{r^2 - (x_i - x_c)^2} \tag{6}$$

The location of the center point, either above or below the line connecting the endpoints, determines whether the quantity under the square root is added or subtracted. If it lies above the line, then the quantity is added. If it is below, then the quantity is subtracted.

The circle-arc method does not have a closed form variance formula. Therefore, bootstrapping was used to estimate standard errors for estimates from this method.

2.5 Observed Score Item Response Theory Equating

Observed score IRT equating first fits an IRT model to each sample of form X and form Y scores. Here, we used only the 2 parameter logistic (2PL) model as we assumed there was no guessing. This method uses person ability (θ_i), item discrimination (a_j) and item difficulty (b_j) to model the probability of person i answering item j correctly. With a sample size of N and an assessment with K items, the 2PL model is:

$$p_{ij}(\theta_i, a_j, b_j) = \frac{exp[1.7a_j(\theta_i - b_j)]}{1 + exp[1.7a_j(\theta_i - b_j)]}, i = \{1, ..., N\}, j = \{1, ..., K\} \tag{7}$$

Estimated distributions of form X and form Y scores were then created using these IRT models. The distributions were then equated using equipercentile equating.

3. Results

There were 199 subjects who completed both the MMSE and the MoCA assessments. The average MMSE score was 24.1 with a standard deviation of 5.7 and a median score of 26. The minimum observed score was 2 and the maximum was 30. The MoCA had an average score of 19.3 with a standard deviation of 6.5 and a median of 20. Minimum and maximum observed scores were 0 and 30. MMSE scores were greater than MoCA scores for 96% of patients and as seen in Figure 1,

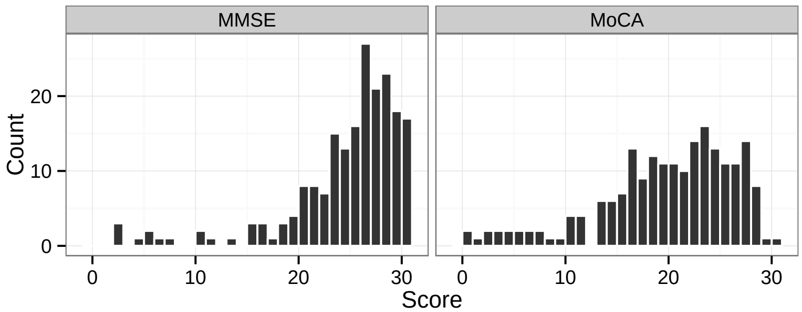

Figure 1: Histograms of observed MMSE and MoCA scores

the distribution of MMSE scores is more skewed left than the distribution of MoCA scores. This suggests that the MoCA is more dificult than the MMSE.

When equating MoCA scores to MMSE scores, the mean and linear methods gave estimates that exceeded the upper limit of the MMSE scale. Figure 2 shows the crosswalk for equating MoCA scores to MMSE scores for each method. The regression analyses in Figure 3 and Table 1 suggests that on average, the mean method underestimates the observed MMSE score by 1.8% and the linear method by 0.8%.

Table 1: Estimated slopes from regression analysis

Method	Slope
Mean	0.982
Linear	0.992
Equipercentile	0.994
Circle-Arc	1.024
IRT	1.065

The crosswalks in Figure 2 for the equipercentile, circle-arc and observed score IRT methods show that these methods are more flexible than the mean and linear methods. The lines are not restricted to be parallel as they are in the mean and linear equating methods. This is most clear with the equipercentile method which has the largest jumps in equated scores. The equipercentile and IRT methods slightly exceed the upper limit of the MMSE score because of the pre-smoothing performed before the scores are equated. The circle-arc method forces the endpoints of each scale to equate to each other so equated scores will always fall within the correct range. The regression analysis shows that the equipercentile method underestimates the observed MMSE scores by only 0.6% on average. The circle-arc and IRT methods, on average, overestimate the observed MMSE scores by 2.4% and 6.5% respectively.

Standard errors for equating a given MoCA score are plotted in Figure 4 by equating method. Standard errors for the mean method were constant throughout the scale as they only depend on the sample variances of the MoCA and MMSE distributions. The linear method gave estimates with some of the highest standard errors in the lower end of the scale, as did the equipercentile method. The IRT estimates had standard errors around 1 for much of the lower end of the scale. The circle-arc method gave some of the lowest standard errors throughout the entire scale of MoCA scores. The circle-arc method is also the only method where standard errors needed to be calculated with bootstrapping. As equating at scores of 0 and 30 are always fixed, these two points had no variability. The middle was the most variable because this is in the neighborhood of the only other estimated point, $(\bar{x}, \bar{y})$. Towards the higher end of the scale, all methods had comparably low standard errors.

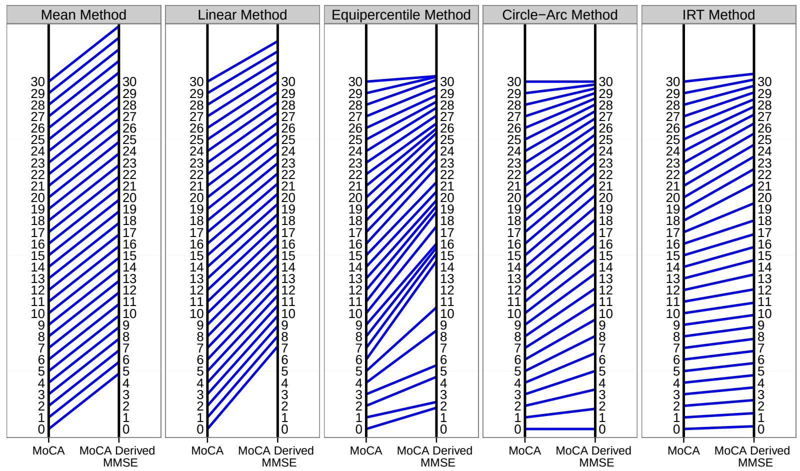

Figure 2: MoCA to MMSE crosswalks by equating method

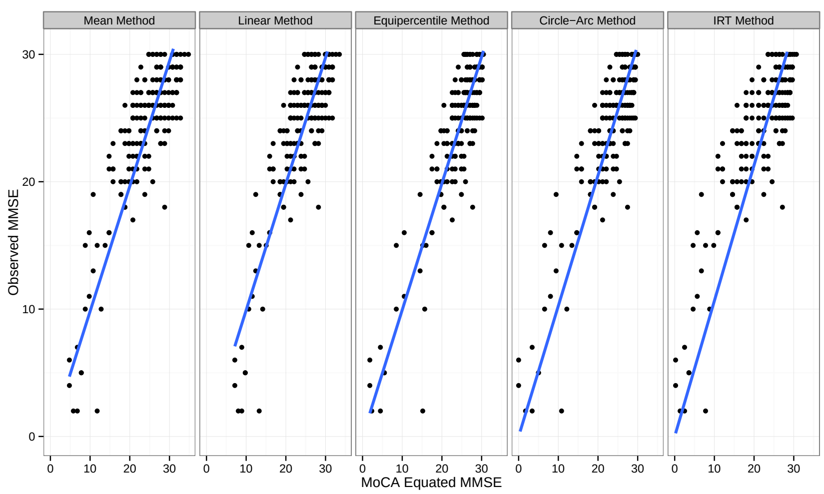

Figure 3: Scatterplots of observed MMSE vs. equated MMSE scores with regression line

4. Conclusion

The mean, linear and equipercentile equating methods performed similarly in mapping patients' MoCA scores to MMSE scores. However, the mean and linear methods are inadequate because they result in MoCA scores equating to MMSE scores that are beyond the upper limit of the test scale. Observed score IRT equating and the circle-arc method did not perform as well as the other methods in the slope

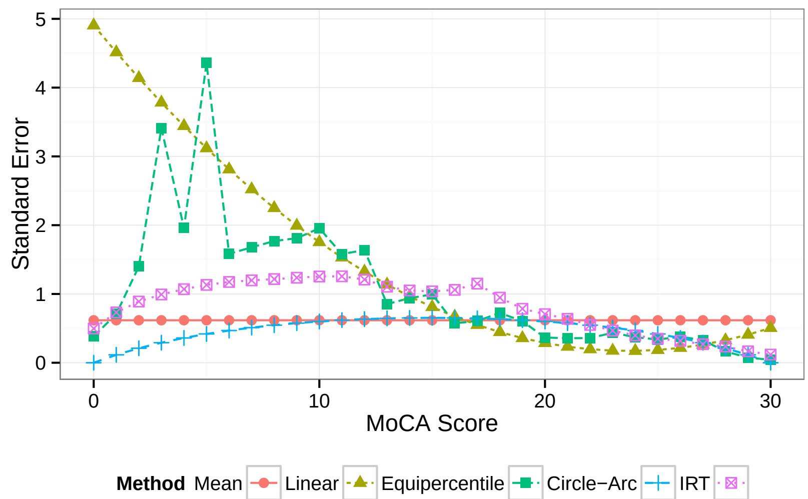

Figure 4: Standard errors of equating across the MoCA scale by equating method

analysis but had smaller standard errors in the lower end of the scale. In the middle and upper end of the scale, the choice of equating method did not matter much as all five methods performed comparably.

The results of this comparison suggest that equipercentile equating is the preferred method in this case. It gave the best agreement between observed and predicted MMSE scores with a slope closest to 1 and although it had some of the highest standard errors in the lower end of the scale, it gave comparable, and sometimes the lowest, standard errors in the middle and upper end of the scale where all methods performed similarly and where most patients scored.

REFERENCES

Albano, Anthony (2013). ***equate****: An R Package for Observed-Score Linking and Equating.* R package version 1.2-0. http://CRAN.R-project.org/package=equate.

Andersson B, Branberg K, Wiber M (2013). "Performing the Kernel Method of Test Equating with the Package kequate". *Journal of Statistical Software*, 55(6), 1-25. URL http://www.jstatsoft.org/v55/i06/.

Andersson B, Branberg K, Wiber M (2012). ***kequate****: The Kernel Method of Test Equating.* R package version 0.5.1, URL http://CRAN.R-project.org/package=kequate.

Folstein MF, Folstein SE, McHugh PR. "Mini-mental state" : A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research.* 1975;12(3):189-198.

Kolen, Michael J., and Robert L. Brennan. *Test Equating: Methods and Practices.* New York: Springer-Verlag, 2004. Print.

Livingston, Samuel A., and Sooyeon Kim. "The Circle-Arc Method for Equating in Small Samples." *Journal of Educational Measurement* 46.3 (2009): 330-43

Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society.* 2005;53(4):695-699.

Wickham, Hadley. ***ggplot2****: elegant graphics for data analysis.* Springer New York, 2009.