Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Brandon Schneider Date

Examining Effect of Age, Education, and Race on the Montreal Cognitive Assessment and the Mini-Mental State Exam

By

Brandon Schneider
Master of Science in Public Health.

Department of Biostatistics and Bioinformatics

_____

John Hanfelt, PhD
Committee Chair

_____

Felicia Goldstein, PhD
Committee Member

Examining Effect of Age, Education, and Race on the Montreal Cognitive Assessment and the Mini-Mental State Exam

By

Brandon Schneider
B.S
Muhlenberg College
2016

Thesis Committee Chair: John Hanfelt, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2018

Abstract

Examining Effect of Age, Education, and Race on the Montreal Cognitive Assessment and the Mini-Mental State Exam
By Brandon Schneider

Introduction: The Mini-Mental State Exam (MMSE) has begun to be replaced by the Montreal Cognitive Assessment (MoCA) to detect the presence of cognitive impairment. As such, there have been multiple attempts to equate the two tests, usually using equipercentile weighting. This method does not adequately control for confounding variables such as race, education, and age. Therefore, we used regression methods to control for these potential confounding variables.

Methods: We used a set of quantile regression models to evaluate potential nonlinear effects between each of the tests and our three variables of interest: age, education, and race. If there was no nonlinear effect present, then we used parametric methods to equate MoCA and MMSE. First, we proposed a linear model with splines to control for potential ceiling effects in the MMSE. This model does not restrict the predictions to be within scientific boundaries. If there were a large number of scientifically absurd predictions we used a non-linear link function.

Results: There were 927 subjects, split into training (n=648) and validation (n=279) datasets. Our population was largely white and highly educated. The quantile regression showed different effects at the tails, with larger effect sizes for the MoCA. However, these estimates were not determined to be different from the OLS (SPELL OUT) estimates, so parametric regression was adequate for equating the two tests. Linear Regression produced significant effects for age and race, as well as the two splines. Education was not significant at any level. Predictions resulted in only one scientifically implausible value, and the majority of predictions were within 2 points of the true values.

Discussion: We conclude that there is no substantial nonlinear effect as determined by the quantile regression, and parametric assumptions should be adequate. In addition, race and age are two previously uncontrolled for confounders that are significant. Limitations include low diversity of the subject population, no restrictions on validity of the predictions, and insufficient controlling for socioeconomic status and cerebrovascular disease.

Examining Effect of Age, Education, and Race on the Montreal Cognitive Assessment and the Mini-Mental State Exam

By

Brandon Schneider
B.S
Muhlenberg College
2016

Thesis Committee Chair: John Hanfelt, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2018

Contents

Introduction

The Montreal Cognitive Assessment (MoCA)[1] and the Mini-Mental State Exam (MMSE)[2] are two screening tools that are commonly used to differentiate between normal cognitive aging and cognitive impairment. Since the 1970s, the MMSE has been the clinical standard; however, this test is being phased out due to low sensitivity in detecting cases of mild cognitive impairment (MCI), a state between normal cognitive aging and dementia, and a fee that must be paid by clinicians if they use it.[1] On the other hand, the MoCA was designed to have better sensitivity in detecting MCI, and it is also free.

As a result, the National Alzheimer's Coordinating Center (NACC) recommended that the MMSE be replaced by the MoCA in their Uniform Data Set (UDS). The UDS consists of a large number of patients with information such as their demographics, health history, clinical diagnosis, and a number of different cognitive test results. One of the tests originally reported was the MMSE. In 2008, a clinical task force recommended replacing the MMSE with the MoCA. As a result, the new UDS contains information from both tests.

The MoCA and the MMSE tests had to be harmonized for data analyses that use the dataset over multiple years. Statisticians were tasked with finding a method to equate scores from the MoCA with scores from the MMSE. The results from this method are described in *Monsell et al.*[3] The Monsell study recruited 935 patients who were given both tests, and used a method called equipercentile weighting with log linear smoothing.[3] Equipercentile weighting consists of matching the scores using the standardized quantiles of each test, thereby creating a crosswalk between the MoCA and the MMSE. This study

was repeated a number of times in different populations and different diseases, all using the equipercentile weighting method.[4-7]

The problem with the above method, however, is that there is no adequate method of controlling for potentially important confounders such as race, age and education. One effective manner to properly control for more than one potential confounder is to use regression methods. Others papers, such as *Wong et al.*, have proposed regression methods to control for age and education.[8] However, these papers have not evaluated their parametric assumptions, and have not included an evaluation of the race variable.

This paper proposes a set of regression methods that will evaluate the parametric assumptions of the model, and control for important confounders. We will use the same dataset as the Monsell study and compare our results against theirs. Our goals are twofold. First, we will use quantile regression on each of the two tests with race, education and age covariates to evaluate parametric assumptions. If the parametric assumptions hold, we will use a parametric linear regression to equate the two scores. However, if we find that our assumptions are broken, we propose a non-parametric method for the equating. Either way, we will compare the accuracy of our predicted MoCA scores against the accuracy of the Monsell crosswalk, using a validation dataset. This will show the effect of controlling for these covariates on the accuracy of the crosswalk.

Background

The Mini-Mental State Exam was first proposed in a paper by Folstein, Folstein, and McHugh in 1974.[2] It was designed to provide a brief (5-10 minute) test to evaluate the cognitive aspects of mental functions. It consists of two sections: one focuses on orientation, memory, and attention with a possible total of 21 points, while the other focuses on following commands both verbal and written with a total score of 9 points. Folstein, Folstein, and McHugh found a clear decreasing trend in the average scores in degrees of severity, with normal patients scoring 27.6 points on average, and patients with dementia scoring 9.7 points on average. The MMSE was also found to have a strong correlation to similar tests from that period.[2] Thus, the test provides a short, relatively accurate diagnostic tool to evaluate cognitive impairment.

The Montreal Cognitive Assessment was developed to provide clinicians with a method to detect Mild Cognitive Impairment, a condition between normal aging and dementia, an area of cognitive difficulty where the MMSE has low sensitivity. The MoCA includes items examining orientation, memory, executive functioning, attention, language, and visuospatial abilities. Similar to the MMSE, scores range from 0-30 points, with higher scores denoting better performance. *Nasreddine et al.* proposed this test in 2005, and performed a validation study to access the sensitivity and specificity of the MoCA in MCI and mild Alzheimer's patients as compared to the MMSE.[1] The study consisted of three groups: MCI, Alzheimer's disease (AD), and healthy controls. The authors found a significant effect for education, so they added a single point to the score if the patient had12 years or less of education. Sensitivity scores were higher in the MoCA than the MMSE for both the MCI and AD groups (90% and 100% respectively). In contrast, the MMSE had sensitivity of 18% and 78% for MCI and AD respectively.

Specificity was lower in the MoCA than the MMSE, with 87% as compared to the 100% identification of the controls with the MMSE. [1] It is important to note, however, that the majority of patients who were controls and had AD were correctly identified with both tests. However, three quarters of patients with MCI scored normally on the MMSE and abnormally on the MoCA. The authors concluded that the MoCA may be more powerful in the mild stages of cognitive impairment, while the MMSE is more powerful in the severe stages. This difference may be due to the more rigorous nature of the memory, executive functions, higher level language, and visuospatial processing assessed with the MoCA.

The most significant problem that has been found with the MMSE is low sensitivity due to ceiling effects in patients with mild cognitive impairment. A ceiling effect occurs when there is some factor that increases the score of a test. In this case, it has been shown the MMSE has poor sensitivity in detecting MCI due to a ceiling effect.[1,9] In other words, the MMSE incorrectly classifies MCI individuals as cognitively normal.. The MoCA reduces this effect. *Trzepacz et al*. performed a study to compare the score ranges of the two tests. They found the MOCA had a far broader range of scores for MCI than the MMSE. This shows a reduced ceiling, because the MMSE had a similar range between normal and abnormal Functional Activities Questionnaires, a tool that can be used to differentiate between MCI and demented status, and the MOCA had vastly different ranges.[6]

Literature Review

To equate the two scores, Monsell et al. proposed using equipercentile equating methods to match the scores based on quantiles. The authors used a study design that is fairly common throughout the literature. A group of patients is recruited with mild to moderate dementia and cognitive impairment symptoms. Then they are given the tests in a random order with an intermediary period depending on the test. Monsell uses a fairly specific population here that is designed to reflect a predetermined distribution of the scores, with more subjects scoring between 15 and 25 than those above and below.[3] Other studies have suggested using more generalized populations, but most use similarly specific populations as Monsell has done.[7,9]

Once the data is collected, the first step of the statistical analysis is to determine correlation. The method begins with determining the correlation between each of the two tests. Monsell argues for using a conservative approach for the cutoff for correlation of 0.6. Using this approach, they found a spearman correlation of 0.76 between the MMSE and the education adjusted MoCA, suggesting that it is appropriate for equating.[9] Similarly, a study equating the short MoCA and the MMSE found a higher correlation of 0.82.[4] In contrast, another study in a more general population used a different method of agreement called the concordance correlation coefficient, which only found a correlation of 0.57.[7] There is some evidence that the correlation may change depending on the disease state. For example, in the *Tzepac et al* paper, high correlations were found for all subjects and the AD and dementia subjects (0.84 and 0.86 respectively), and lower correlations were found in the healthy aging group and the MCI group (0.6 and 0.43 respectively).[6] This finding is consistent with the finding of the MoCA being able to better differentiate between MCI and healthy cognitive aging, however. Overall, there is

strong evidence for medium to high correlation between the two tests, making equating valid.

Monsell et al suggest using equipercentile weighting with log linear smoothing to equate the two scores of highly correlated tests, including the education adjusted MoCA and the MMSE.[3] Their study also used a validation subset to determine how far off their predictions were. Only 33% of the predictions for the MOCA were exactly the same as reported, while 61 and 83% were within one or two points.[3] Generally, higher scores on the MMSE were equated to lower scores on the MoCA which is consistent with findings that the MoCA is more sensitive to mild cognitive impairment than the MMSE. A few observations that were more than 5 points off were found (4%), and they were mostly found in the lower scores. This is not an unsubstantial number, and the authors mention imputation needs to be improved.

However, there are a few problems with this method. The only method of controlling for confounders such as education and age involves reducing the scores by a specified amount, creating a confounder adjusted score. For example, the education adjusted MoCA score increases the score by one point if the patient has <=12 years of education.[1] The issue with this is that both the 12 years of education and the one-point reduction are fairly arbitrary. Instead, it may be better to perform a regression analysis to control for confounders in a more dynamic manner.

Wong et al. used Poisson regression to compare the MMSE and MoCA in stroke patients.[8] While it is a different disease of interest, it is one of the only papers to perform equating with regression techniques. The study compared equipercentile weighting with log linear smoothing against a Poisson regression controlling for education and age. The

regression model performed well, with 44.5% of the patients one point away from the true value, and 63.2% within two points. In contrast, the equipercentile method produced 39.1% and 60.9% for one and two points away respectively.[8] Thus, in this dataset, the regression method controlling for the two confounders performed better than the equipercentile weighting method.

There are issues with the Wong study, however. Poisson regression may not be appropriate for predicting a test that is only on a scale of 0 to 30. The vast majority of the patients are considered normal, so their scores are above 26 points, skewing the distribution.[8] The paper does not provide regression diagnostics so it is difficult to tell whether or not their model fits the assumptions of the Poisson regression. A continuous model may be more appropriate. This paper suggests that controlling for age and education is appropriate and may produce better predictions.

In addition to potential errors in the model assumptions, Wong did not control for race as a potential confounder. The research regarding this variable is sparse. *Rossetti et al* studied a population of community dwelling African Americans, and recorded their MoCA scores. On average, the patients scored a 22.01 which is much lower than the suggested 26 cutoff from the MoCA. Even after adjusting for education, the study still found the mean to be lower.[10] However, there was another study performed to validate the equating in a more diverse population. Falkowski et al studied 185 patients and gave them both tests, and used the crosswalk provided by Roalf et al[9] to derive predicted scores. They found they all had high intra class correlations across racial groups.[11] Thus, it is unclear whether or not race will have a significant effect, and it is worth studying.

Methods

Study Population and Design:

As stated previously, the dataset was collected from a group of patients at the ADCs who were given a multitude of different tests including the MOCA and the MMSE. The dataset consists of both new and returning patients. Ideally, the study would have only consisted of new patients; however, due to time constraints, both tests were given to both new and returning patients.[3] For this analysis we are only interested in the study visits during which the Crosswalk study was completed for the first time. If the returning patients have more than one visit, the extra visits will be ignored. Patient recruitment was stratified to ensure the majority of patients were either MCI or mildly demented, with ADCs being instructed to recruit the majority of patients with MMSE scores between 16 and 26 points. No scores below 10 points were included due to the difficulty of administering the tests at that severity of cognitive impairment.[3] Each patient received the tests in a random order to ensure no learning effect was present. The authors found that this randomization was successful. More details can be found in *Monsell et al*.

Statistical Analyses:

The main statistical questions of interests are as follows: (1) Are education, age, and race associated with MOCA and MMSE in a nonlinear fashion? (2) Does MMSE have a nonlinear relationship with MOCA due to ceiling and floor effects, controlling for the above variables? (3) Do age, education, and race have a significant effect on the equating between MMSE and MOCA? These questions have not been addressed adequately in the previous literature, due to the inability of equipercentile weighting to adequately adjust for more than one potential confounder and explore nonlinear effects.

We propose a set of regression methods to explore the dynamic relationships between the tests and the three potential confounders.

To explore the potential non-linear relationship between MOCA, MMSE and the three potential confounding variables: race, education, and age, we propose two quantile regression models. Quantile Regression allows us to evaluate how the effects of these variables change over the different quantiles of the outcome variables of interest: MOCA and MMSE. In other words, are the effects of these variables more extreme in patients who scored higher or lower in the MOCA or MMSE, and how is this change different between the two tests?

Variables were coded in the following manner. Age is centered and standardized for ease and interpretability of computation by subtracting its median of 75 and by dividing by 10. Race is coded into two categories, white and asian vs black, due to the low number of Asians, Hispanics and other race groups. Education is recorded as the years of education obtained, and then separated into three categories: less than 12 years, 12 to 16 years, and greater than 16 years. These groupings approximately correspond to a high school, college, or graduate education. In addition, data points with missing values for any of the above variables were removed from the dataset.

We define the quantile regression as follows. Define Y as the outcome variable, X the design matrix, and τ the quantile of interest. We wish to estimate β_τ the vector of regression coefficients for the model of the τth quantile. Then, we can define the following function:

$$Q_Y(\tau) = F^{-1}(Y) = \inf\{y : F(y) > \tau\}$$

Where F is the cumulative distribution function of Y. This is the quantile function, which is the minimum value of Y such that Probability of Y is greater than tau. For regression purposes we use the conditional quantile function as the following:

$$Q_{(Y|X)}(\tau) = X\beta_\tau$$

This is the formulation for the Model equation for quantile regression. The goal is to estimate β_τ using the following equation.

$$\widehat{\beta_\tau} = \arg\min \sum_{i=1}^{n} \rho_\tau(\mathrm{Y} - \mathrm{X_i}\beta_\tau)$$

The function ρ_τ is defined as the expected loss function, which can be written as:

$$E\big(\rho_\tau(\mathrm{Y} - \mathrm{X_i}\beta_\tau)\big) = (\tau - 1) \int_{-\infty}^{X\beta} (y - X\beta) dF_Y(y) + \tau \int_{X\beta}^{\infty} (y - X\beta) dF_Y(y)$$

Here we are trying to minimize the expected loss between the predicted and true values for each value of tau, which gives us the result of $\widehat{\beta_\tau}$.

In this analysis we will define two models. They will both use the centered age, categorical education, and binary race as defined above. The first model will use MMSE as an outcome variable, while the second will use MOCA. We will regress over 7 quantiles ranging from 0.2 to 0.8 in increments of 0.1. This will give us a wide view of how the effects change over a range of quantiles.

To access the difference between the quantile regression and ordinary least squares, we will create a coefficient plot using R. This will plot each of the coefficients for our predictors over our quantiles of interest. These plots will also include the OLS estimate as a comparison. If the quantile regression model suggests that the linearity

assumptions hold, we will use parametric linear regression to equate MMSE and MOCA, adjusting for age, education and race.

The parametric model will include a linear link function of MOCA as its outcome variable, and use MMSE as its main predictor of interest. We will split the data into training and validation datasets, which will consist of 70% and 30% of our data respectively. Due to ceiling and floor effect present in the MMSE, there may still be a non-linear effect present. As such, we will include a knot at an MMSE value of 24, as defined by *Trzepacz et al*. as the cutoff value between MCI and normal aging.[6] We will use a cubic spline for this model. This will allow the effect of MMSE on MOCA to change between MCI and healthy individuals. The three predictor variables: race, education, and age are coded the same way as before. MOCA and MMSE will be centered and standardized for ease of computation and interpretation, by subtracting the median and dividing by 10. If this yields poor results, we will use non linear regression, such as a three parameter logistic link instead.

When the models are built we will use the validation dataset to evaluate the accuracy of the predicted MOCA values. This model does not place any asymptotic restrictions on the predicted values. Therefore, scientifically absurd values may be expected. We will report how far the predicted values deviate from the true values and compare these against the equipercentile weighting method used by *Monsell et al.*

Results Section

Study Population

After removing data points with missing values in our variables of interest, 927 subjects remained. Descriptive statistics of this population can be found in Table 1.

Overall, the Mean on the MMSE is larger at 26.31 than the MOCA at 21.91, which is fairly consistent with other findings. Our population is largely white or Asian as well as highly educated, with only 17% representing African Americans, and only 19% having less than a high school education. Due to the few lower educated and non-white participants in our study, we are not able perform an interaction analysis between education and race. It may be of interest in a later study, however. Further descriptive statistics can be found in *Monsell et al.*

We divided these subjects into a training and validation dataset resulting in 648 in the training set and 279 in the validation dataset. This was performed using a regular simple random sample. As shown in table 1, the randomization into the two datasets was fairly successful, with the proportions of the variables of interest being representative of the population as a whole. Some differences were seen it the validation dataset; however, they were not extreme, and this is to be expected with a smaller sample size.

Variable	**Level/units**	**Overall Descpritive Statistics (n=927)**	**Training Descriptive Statistics (n=648)**	**Validation Descriptive Statistics (n=279)**
MMSE	Mean (SD)	26.31 (4.92)	26.45 (4.78)	26.00 (5.05)
MOCA	Mean (SD)	21.91 (6.42)	22.03 (6.29)	21.65 (6.70)
Race				
	White n (%)	766 (82.6)	528 (81.5)	238 (85.3)
	Non-white n (%)	161 (17.4)	120 (18.5)	41 (14.7)
Education				
	<12 years n (%)	165 (18.8)	110 (17.0)	55 (19.7)
	13-16 years n (%)	391 (42.2)	266 (41.0)	125 (44.8)
	>16 years n (%)	371 (40.0)	272 (42.0)	99 (35.5)

Age	Mean (SD)	74.99 (9.60)	75.41 (9.32)	74.03 (10.17)

Table 1 Descriptive Statistics: Presents descriptive statistics for variables of interest on full dataset as well as validation and training sets.

Quantile

Regression of MMSE and MOCA

We constructed two models to evaluate the relationships between MMSE and MOCA, and three confounders: race, education, and age. Quantile Regression estimates were made over 7 quantiles ranging from 0.2 to 0.8 in increments of 0.1. Age was centered around its median and divided by 10 for computational purposes. The reference group for education was less than 12 years, whereas the race variable used the white and Asian population as the reference group. Results of the regression for two values of tau, 0.3 and 0.7, are given below in table 2.1 and 2.2.

Variable	Coefficient for tau=0.3	95% CI for tau=0.3	Coefficient for tau=0.7	95% CI for tau=0.7
Race	0.62	(-.39, 1.35)	-0.31	(-0.75,0.70)
Centered Age	-0.77	(-1.09,-0.30)	-.22	(-0.51,0.34)
Education 12-16 years	2.23	(0.48, 4.11)	0.69	(0.35, 1.08)
Education >16 years	3.92	(2.48, 5.80)	1.33	(0.68,1.70)

Table 2.1 Quantile Regression on MMSE Results: These coefficients for the 30^{th} percentile can be interpreted as follows. For a continuous variable such as Age, for a 1 unit increase in centered Age, the MMSE decreases by 0.77 in the 30^{th} percentile. For a categorical variable such as Race, MMSE is 0.62 higher for Non-white individuals, than white or asian individuals. Interpretations for the 70^{th} percentile are similar.

The effects of age and education seem to decrease between the two Quantiles, and all three are significant, in all cases except for age at the 70th percentile. Race is not a significant factor for MMSE at either tail, as well. A coefficient plot for MMSE can be found in figure 1. Graphed here are the different coefficients from the quantile regression mapped over the quantiles 0.2 to 0.8. The OLS estimate and itx confidence interval is found as well. In the MMSE plot, there does not appear to be any difference between the OLS estimate and the Quantile regression estimates regardless of which quantile we look at. This can be determined by the overlap of the quantile regression intervals and the OLS intervals. Overall, there does not appear to be a difference between the OLS and the Quantile regression estimates.

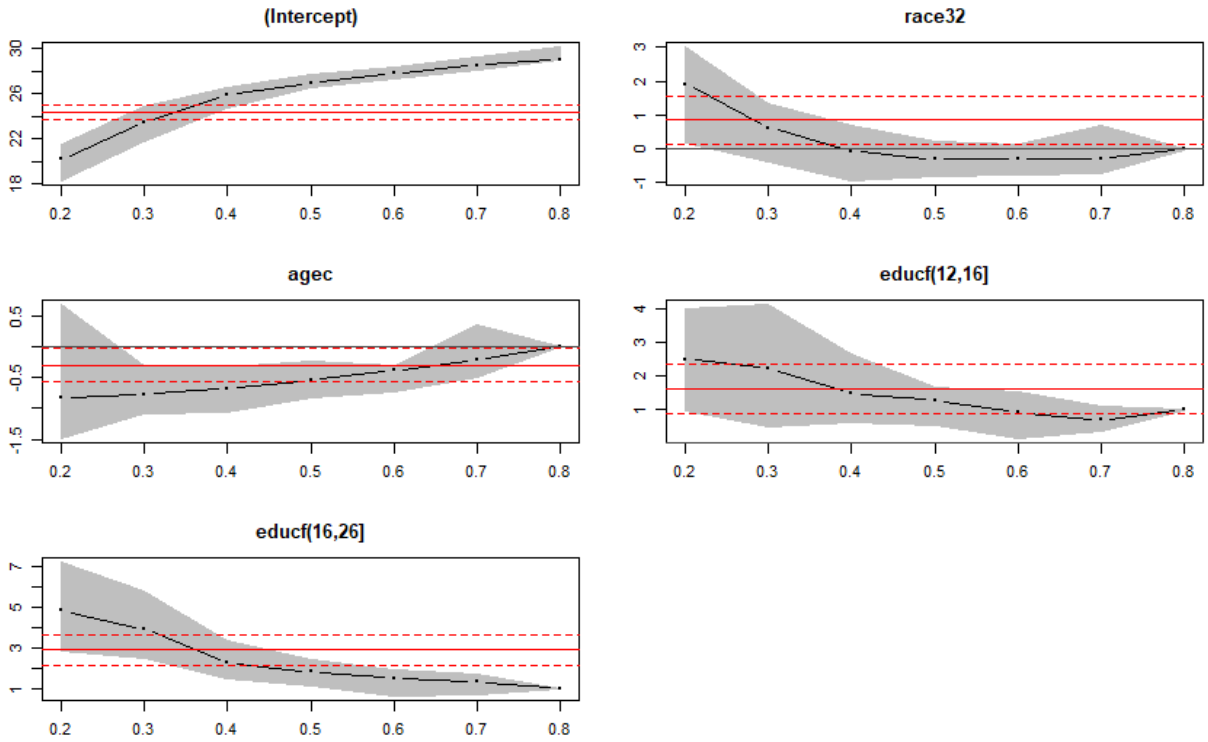

Figure 2.1 Coefficient Plot for MMSE: The black line represents each of the coefficient estimates at each of the values of tau on the x axis. The y-axis is the values of the coefficients. The red line is the OLS estimate and corresponding confidence interval.

The Quantile Regression results for the MOCA are reported in table 2.2. Overall the results for the MOCA are larger than the effects on the MMSE. In addition, Race is significant at the 70th percentile, unlike the MMMSE. However, education at 12-16 years is not significant compared to less than 12 years of education. A coefficient plot is included in figure 2.2. Similar to the above there does not appear to be evidence that there is a difference between the quantile regression estimates and the OLS estimates.

Variable	**Coefficient for tau=0.3**	**CI for tau=0.3**	**Coefficient for tau=0.7**	**CI for tau=0.7**
Race (Nonwhite, ref=White)	-.68	(-1.70, 0.23)	-1.84	(-2.87,-0.94)
Centered Age	-1.29	(-1.86,-0.44)	-1.05	(-1.73,-0.61)
Education 12-16 years	1.48	(0.61, 4.47)	1.26	(-0.07, 1.08)
Education >16 years	4.77	(3.33, 7.46)	2.16	(0.57,2.68)

Table 2.2 MOCA Quantile Regression Results: Coefficients and Confidence intervals for Regression of MOCA at tau=0.3 and tau=0.7

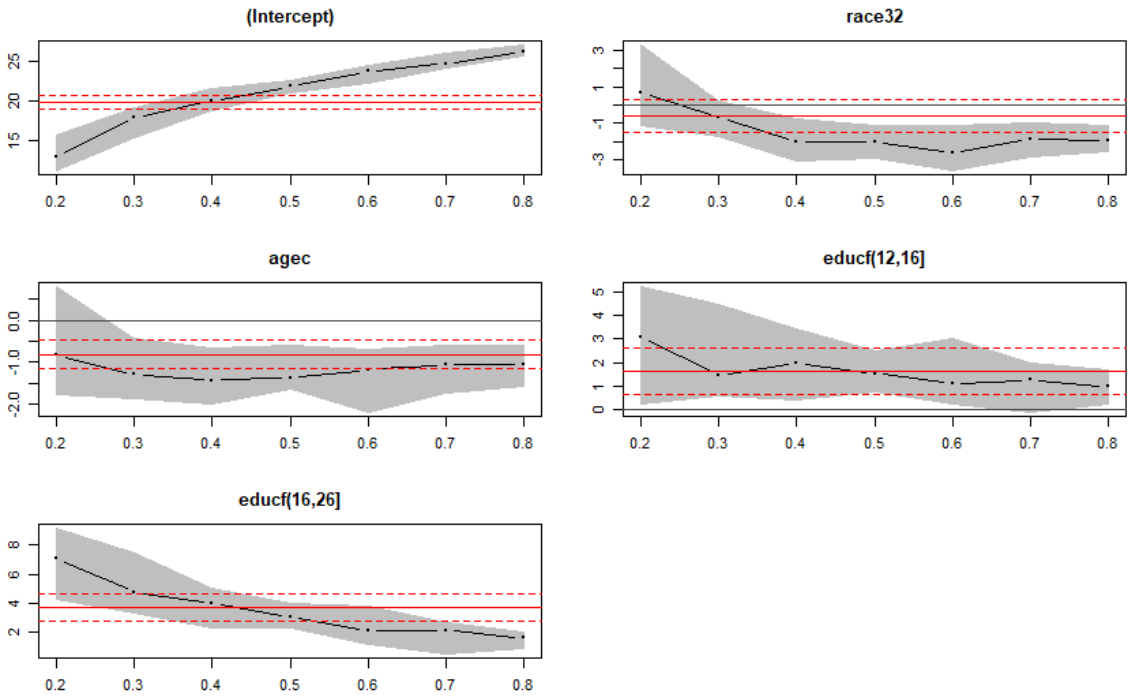

Figure 2.2 Coefficient Plot for MOCA: Black line represents coefficients. Red line is OLS

Overall, the effects are larger for our potential confounders in the lower quantiles than the upper quantiles, and they are more extreme in the MOCA than the MMSE. However, these estimates do not appear to differ significantly from the OLS estimates, so a parametric linear regression method can be used for equating the two tests.

Linear Regression

We then evaluated the impacts race, age and education have on the equating of MMSE and MOCA. We used linear regression methods with piecewise linear splines to evaluate the associations and perform prediction. The model was created using the training dataset, and prediction was evaluated using the validation dataset.

In this model, in addition to age, MOCA and MMSE were also centered around their median (22 and 26 respectively) and divided by 10. As such coefficients should be

interpreted in this manner. The table of the results for the linear regression can be found in table 2.3.

Variable	Coefficient	CI
Intercept	-2.63	(-2.82,-2.44)
MMSE	2.23	(2.00,2.46)
The MMSE 24-28	2.82	(2.63,3.01)
MMSE >28	3.08	(2.89,3.27)
Race (Nonwhite, ref=White)	-0.142	(-.203,-.08)
Centered Age	-.039	(-0.06,-0.01)
Education 12-16 years	0.01	(-0.06, 0.07)
Education >16 years	0.05	(-0.02,0.12)

Table 3 Results from the Linear Regression of MOCA and MMSE

As seen above, all linear splines were statistically significant with the difference between MOCA and MMSE, holding all else constant, increasing as the MMSE increases. This holds consistent with finding regarding the ceiling effect present in the MMSE. Interestingly, education was not found to be significant at any level. In contrast, Race and Age were found to be significant.

Using the validation dataset, we evaluated the predictions of MOCA from the model against the true values of MOCA. The results can be found in table 2.3. The majority of predictions (about 52%) were within two points of the true values, with over

80% being within 4 points of the true values. These differences were unstandardized back to values between 0 and 30 for this analysis, and should be interpreted as such. Only one value was found to be outside fo the normal range. This data point consisted of 0 for both the MMSE and MOCA.

Residual Size (n (%))	**+-1**	**+/- 2**	**+/- 3**	**+/- 4**
	66 (23.6)	144 (51.6)	184 (65.9)	225 (80.6)

Table 4- Prediction results from Linear Regression. Difference between true and predicted values of 1, 2, 3, 4 are represented

These values appear to be lower than the predictions made by *Monsell et al*. using equipercentile weighting, but they are not dramatically lower. It is difficult to compare between the two as well as we do not have whole number predictions. Our table rounds to the next largest whole number. This was done to be conservative. The model does not have any restrictions regarding keeping the predictions between 0 and 30. However, only one prediction was outside of this range. Overall, the predictions are good, but can be improved.

Discussion

In this paper, we have used regression methods to address questions regarding the effects important demographic variables have on equating MOCA and MMSE. =We studied how the demographic variables of age, race, and education affect the test scores individually, allowing for potential non-linear effects. Then we evaluated how each demographic variable changed what a score on MMSE would equate to on the MOCA.

By doing so, we have shown that previous methods are not adequate to adjust for multiple potential confounding effects.

In our quantile regression analysis, we found that multiple factors are associated with how high they scored on the test. These effects tended to be more extreme in the case of the MOCA. In both cases, there is some evidence the effects were higher among those who scored higher in each test. Intuitively, this makes sense due to ceiling effects that may be caused by these demographic variables. However, in both cases the coefficients were not significantly different than the OLS estimates, suggesting that using parametric assumptions may be adequate.

We chose to add piecewise linear splines to the model to evaluate how the effect of MMSE on MOCA changed over the with higher levels of MMSE. Splines were added at 24 and 28, resulting in significant increases in the effect. In other words, as MMSE increases, there is evidence to suggest that the gap between MOCA and MMSE increases. This reflects the ceiling effect in the MMSE, showing the difficulty in detecting MCI. Interestingly, we found no education effect, and instead found a race effect. It is unclear to what degree this is representative of an interaction effect, but we do not have the sample to evaluate this question. Age was also significant in the model. Overall, the linear model showed potential effects for race and age. Other studies using equipercentile weighting have not adjusted for these effects at all. This is the most important strength of our study.

There are multiple limitations to this study, however. Perhaps most important is that the crosswalk data is not a diverse sample. Under 20% of the study population was not white or had under 12 years of education. In addition, the population had a large

number of participant with graduate level education. In other words, our population was unrepresentative of the diversity of education and race in the US as a whole. This may be due to a problem in recruiting patients for these studies, in that the highly educated are more likely to participate, and the highly educated are more likely to be white or Asian.

Another downfall of this study is to the lack of adequate controlling for socioeconomic status. We have a mostly elderly population, with a median of 75. Education at this age may not be a good indicator of SES. Perhaps the reason why we did not see an education effect, where others have, is the inability to properly control for the SES with only education. A potential variable of interest to broaden our view of SES is income. Income at intake may be more indicative of a subjects SES at that age than education level. In addition to SES, as mentioned above, race/SES interaction effects could be of interest if we had a more diverse population.

A potential confounding variable of interest that we were not able to study is the presence of cerebrovascular disease. A lower score on either test could be a result of a number of vascular diseases, so controlling for the presence of this may be beneficial. One can measure this through the Hachinsky Ischemic scale. This scale is strongly associated with probable cerebrovascular disease and measures several factors such as: hypertension, history of smoking and cognitive status. This scale has been validated to be associated with vascular dementia.[12]

Another area for potential improvement is to build a stronger predictive model. While we shoed parametric methods may be useful, linear regression may not be the best model for prediction. We chose linear regression to focus on the covariate effects, and ensure interpretability. Using a model that restricts the potential predictions between two

values may be of interest. This will reduce the number of scientifically absurd predictions and may improve prediction accuracy.

Overall, our study showed that the previous method of equipercentile weighting may not be adequate. The only method of adjusting for confounders is to use crude changes based off of binary or categorical variables. We have shown that regression methods, which are fairly simple, can control for several important factors at a time. Race, Education and Age cannot be ignored as they can effect how one scores on the tests, and they need to be adjusted for. Complex regression methods may not be necessary, but regression methods in general should be used to provide adequate control for confounding variables.

References

1 Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society* **53**, 695-699, doi:10.1111/j.1532-5415.2005.53221.x (2005).

2 Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* **12**, 189-198, doi:https://doi.org/10.1016/0022-3956(75)90026-6 (1975).

3 Monsell, S. E. *et al.* Results From the NACC Uniform Data Set Neuropsychological Battery Crosswalk Study. *Alzheimer disease and associated disorders* **30**, 134-139, doi:10.1097/wad.0000000000000111 (2016).

4 Roalf, D. R. *et al.* Bridging cognitive screening tests in neurologic disorders: A crosswalk between the short Montreal Cognitive Assessment and Mini-Mental State Examination. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **13**, 947-952, doi:10.1016/j.jalz.2017.01.015 (2017).

5 Saczynski, J. S. *et al.* The Montreal Cognitive Assessment: Creating a Crosswalk with the Mini-Mental State Examination. *Journal of the American Geriatrics Society* **63**, 2370-2374, doi:10.1111/jgs.13710 (2015).

6 Trzepacz, P. T., Hochstetler, H., Wang, S., Walker, B. & Saykin, A. J. Relationship between the Montreal Cognitive Assessment and Mini-mental State

Examination for assessment of mild cognitive impairment in older adults. *BMC geriatrics* **15**, 107, doi:10.1186/s12877-015-0103-3 (2015).

7 Helmi, L. *et al.* Agreement and conversion formula between mini-mental state examination and montreal cognitive assessment in an outpatient sample. *World journal of psychiatry* **6**, 358-364, doi:10.5498/wjp.v6.i3.358 (2016).

8 Wong, A. *et al.* Converting MMSE to MoCA and MoCA 5-minute protocol in an educationally heterogeneous sample with stroke or transient ischemic attack. *Int J Geriatr Psychiatry*, doi:10.1002/gps.4846 (2018).

9 Roalf, D. R. *et al.* Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **9**, 529-537, doi:10.1016/j.jalz.2012.10.001.

10 Rossetti, H. C. *et al.* Montreal Cognitive Assessment Performance among Community-Dwelling African Americans. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* **32**, 238-244, doi:10.1093/arclin/acw095 (2017).

11 Falkowski, J. A. *et al.* Conversion of MoCA to MMSE scores. *Alzheimer's & dementia (Amsterdam, Netherlands)* **1**, 125, doi:10.1016/j.dadm.2015.02.001 (2015).

12 Moroney, J. T. *et al.* Meta-analysis of the Hachinski Ischemic Score in pathologically verified dementias. *Neurology* **49**, 1096-1105 (1997).