Modeling Patient Satisfaction After Arthroscopic Partial Meniscectomy Using Knee Injury and Osteoarthritis Scores (KOOS) Measured with Error

by

Tucker J. Harvey

B.S., Allegheny College, 2020

Submitted to the Graduate Faculty of the

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Tucker J. Harvey

It was defended on

April 25, 2022

and approved by

Dr. Maria Brooks, PhD, Epidemiology

Dr. Jeanine Buchanich, PhD, Biostatistics

Dr. Jenna Carlson, PhD, Biostatistics

Thesis Advisor/Dissertation Director: Dr. Ada Youk, PhD, Biostatistics

Modeling Patient Outcome After Arthroscopic Partial Meniscectomy Using Covariates Measured with Error

Tucker J. Harvey, M.S.

University of Pittsburgh, 2022

Pain and reduced function resulting from tears of the meniscus can cause disability, compounded by a risk of osteoarthritis. The Knee Injury and Osteoarthritis Score (KOOS) is a tool for quantifying pain and functional deficits, but scores can vary even without underlying biological change. Arthroscopic Partial Meniscectomy (APM) is a minimally invasive surgery used to repair meniscal tears, but it's hypothesized that using improvements in KOOS pain or function sub-scores alone is too simplistic to determine whether patients are satisfied with their post-operative results. This project modelled post-operative satisfaction using KOOS and baseline demographic characteristics, while accounting for the intra-subject variability of KOOS. Logistic regression was used to model satisfaction with KOOS and demographic covariates. A backwards-selection technique with bootstrapping was used to quantify variable importance. Age, education, race, and mental health were identified as important covariates for a reduced model. A multiple imputation technique was used to simulate KOOS uncertainty, in which other covariates were used to impute potentially true values of the erroneously-measured variables using background information about the distribution of errors. This was followed by regression of satisfaction on these imputed values and computation of corrected regression coefficient estimates. Minor changes to regression coefficient estimates and odds ratios resulted, but the associated confidence intervals generally overlapped with the uncorrected estimates. While the effect of KOOS pain on satisfaction marginally decreased, the effect of KOOS function increased. Individuals who were worse off

(high pain and low function) at baseline, and those whose pain and function improved the most, had the highest probability of satisfaction. More research is needed to exactly explain the effect estimates for demographic predictors. Additionally, simulation studies would be useful to determine the performance of this measurement error correction method, as would a study of KOOS using validation data. This project has public health implications in educating clinicians and patients about what factors are important in determining satisfaction after APM, specifically how KOOS should be used. There are considerable ethical and financial benefits to more effectively identifying ideal candidates for surgery, while ruling out those unlikely to benefit.

Table of Contents

List of Tables

List of Figures

Preface

Definitions of most common abbreviations used (also defined as used throughout text):

- PASS: Patient Acceptable Symptom State, Outcome Variable
- KOOS: Knee Injury and Osteoarthritis Score, Primary Predictors of Interest
- APM: Arthroscopic Partial Meniscectomy Surgery
- OA: Osteoarthritis
- MIME: Multiple Imputation for Measurement Error

Definitions of most common variables used (also defined as used throughout text):

- Z: Variables measured with error (KOOS in this case)
    - Z_M: Erroneously measured values
    - Z_T: True values
- X: Covariates not measured with error
- Y: Outcome variable (PASS in this case)
- ME: Measurement error

I'd like to acknowledge Dr. Naveen Subhas and the Imaging Department at the Cleveland Clinic Foundation for allowing to conduct this project as a visiting researcher, providing me with access to this data, and supporting this work. I'd also like to acknowledge all University of Pittsburgh faculty members who served as my committee members and have provided me with support and guidance during this project.

1.0 Introduction

Modelling clinical outcomes is a challenging yet necessary task. Often, the goals are to predict patient outcomes using available data on personal and medical factors. While building a model which consistently makes accurate predictions is generally desirable, in some cases it may also be of interest to understand whether and how particular personal and medical factors can best be used to explain why a particular outcome occurs. This interpretive goal, of better understanding the relationship between predictor and outcome, and whether such a relationship is statistically significant, in addition to making accurate predictions overall, was a central aim of this project. Such an understanding is particularly important when considering the outcomes of costly procedures with an extended recovery period and variable results, as was the case with Arthroscopic Partial Meniscectomy (APM) knee surgery which was considered here. An additional challenge can result when patient factors potentially associated with clinical outcomes are those perceived by patients rather than measured strictly. However, patient perception of deficits prior to surgery, the resulting perceived improvements in those deficits, and patients' overall feelings of satisfaction associated with their treatment experience and outcome should be of foremost consideration, as was the case in this modelling project. A deeper understanding of the association between such perceived deficits and clinical outcome by both physicians and patients could be highly beneficial in better tailoring treatment decisions to individual cases of knee injury.

1.1 Meniscal Injury

Knee injury is an extremely common cause of discomfort which often leads to disability, particularly in cases of severe and/or chronic pain. Severe disability is associated with increased age, as well as a variety of other physical and personal factors (Baldwin et al. 2017). The high prevalence of debilitating knee pain can partially be attributed to osteoarthritis (OA), for which torn meniscus (cartilage of the knee joint) is a strong risk factor (Englund, Roos, & Lohmander, 2003). Patients with symptomatic OA alone can often be treated with physical therapy and other non-surgical treatment, whereas individuals with meniscal tears may require arthroscopic partial meniscectomy (APM) (Samson et al. 2007). APM is a very common treatment for meniscal tears, but the benefit of surgery over other treatment is questionable for some individuals (Katz et al., 2013). Increased age and several health comorbidities, including OA, have been reported to be associated with decreased surgical benefit and adverse outcomes (Abram et al., 2018) (Englund, Roos, & Lohmander, 2003). Understanding the patient-specific personal and medical factors associated with successful APM is necessary to both inform its clinical use and to adequately educate patients about treatment options. Patient-perceived satisfaction with surgical results should be of principal importance in determining procedure success.

The Knee Injury and Osteoarthritis Score (KOOS) is derived from a questionnaire aimed at assessing patients' opinions of their own knee-associated problems. These are well-established metrics used for both research and clinical purposes (Roos & Lohmander, 2003). KOOS is intended for use in knee-injury cases which may result in OA (such as meniscus injury) and can serve to both comprehensively assess patient experience at initial consultation and to monitor individuals over time (Roos & Lohmander, 2003). KOOS combines daily function and disability with recreational abilities, increasing validity for a wide range of patient lifestyles. Completion of

the KOOS questionnaire results in scores for multiple subsections. This project focused on baseline (prior to surgery) KOOS pain and function subsections, as well as the change in both pain and function between baseline and a 1-year post-operative follow-up point. These four scores served as the primary predictors of interest in this project. KOOS pain and function are both measured with a range of [0, 100], whereas change in pain and function have a potential range of [-100, 100]. High baseline scores represent severe patient-reported pain and greater levels of perceived functioning (low disability), whereas higher values for both change in pain and change in function represent greater improvement (decreased pain, improved function).

1.2 Analytical Challenges

1.2.1 Model Selection and Variable Importance

Covariate selection is a seemingly simple yet important task in modeling, particularly in clinical settings. While models built with the primary goal of maximizing predictive accuracy may benefit from including as many predictors as the sample size will allow, those built with goals of understanding and testing the effects of individual covariates on the outcome, as was the case with this project, may wish to build simpler and/or smaller models (James et al., 2017). In some cases, inclusion of many non-essential predictors in a logistic regression model has been reported to weaken estimates of the true associations between predictors and outcome, lead to imprecise estimates, and increase type II error (Ranganathan, Pramesh, & Aggarwal, 2017). Often, univariate analysis precedes modeling, in which the association of each variable with the outcome is first tested without the presence of other variables, and additionally considered for its scientific

plausibility (Ranganathan, Pramesh, & Aggarwal, 2017). Following such preliminary steps, backwards stepwise selection is one of the most common and straightforward methods by which important covariates are selected from a high-dimensional dataset, one with many variables. In the case of logistic regression, variables are typically removed one at a time, selected by minimizing model deviance at each step, and then models of different sizes can be compared in several ways (James et al., 2017). Backwards selection has been shown to have weaknesses when used in place of purposeful selection methods (Bursac et al., 2008). Once a variable is excluded in the stepwise process, it is not allowed to return, even if it becomes more useful in a smaller model. For this reason, one "best" subset of predictors is not guaranteed to be found (James et al., 2017). Additionally, limited information is provided about the importance of variables by this method. Variables are either removed or not, which may be unique to the sample being considered, and prevents the degree of their usefulness from being ascertained.

The bootstrap is a tool which is widely applicable to situations where it is desirable to quantify the uncertainty associated with a statistical method (James et al., 2017). It overcomes the limitations that result from having only one sample with which to draw inferences about a sampling distribution. In bootstrapping, observations are randomly sampled from a main dataset of interest, usually with replacement, to create a new bootstrapped sample. By repeating this sampling process many times, slightly different bootstrapped samples can be obtained, creating a simulated sampling distribution. Often, a statistical process is performed on each bootstrapped sample, each time resulting in an estimate of interest, and their variance can then be computed (James et al., 2017). In our case, bootstrapping was used to overcome the limitations of backwards selection. By performing this variable selection technique on multiple bootstrapped samples, a more robust understanding of the frequency with which variables were selected and the consistency with which

they were useful in our logistic regression model was obtained. This information was then used to make a better-informed judgement about variable importance in our specific case.

1.2.2 Measurement Error

Following an investigation of variable importance and covariate selection, the primary analytical challenge associated with the current project is that the pain and function scores derived from the KOOS questionnaire are measurements that suffer from intra-subject variability. The test-retest variance for the KOOS pain score has been reported to be 4.8, and 8.3 for the KOOS function score (Collins et al. 2011). For example, if the KOOS pain was measured for a patient on multiple consecutive days where there was no biological change, the variance of those scores is 4.8 (Collins et al. 2011). Therefore, the analyses used in this project focus on methods attempting to correct for such within-subject uncertainty. This uncertainty is effectively a type of measurement error, as the subject reported score varies somewhat from the *true* measure of pain or function.

Measurement error is a widespread problem which is particularly present in observational studies involving inaccurate instruments, expense associated with exact measurement, subjectivity associated with some self-reported measurements, and variables for which exact measurements cannot be achieved (Guolo, 2008). Failure to correct for measurement error can result in biased regression coefficient estimates, both for the variable(s) measured with error, and for other covariates in the model, either towards or away from the null value (Rosner, Spiegelman, & Willett, 1990) (Thürigen et al., 2000). Measurement error can also result in unreliable confidence interval spread and reduced statistical testing power (Armstrong, 2003). Mis-measured values sometimes show bias through systematically overestimating or underestimating true values (Thürigen et al., 2000). In the present case, KOOS pain and function measures represent patient

perception, and therefore "true" values are unattainable, but their error is believed to have a mean of zero. If Z is a variable measured with error, the true value of that variable, Z_T, is equal to the sum of the measured value (Z_M) and the error associated with that measurement (ME) (Keogh & White, 2014). In the present case, ME can be considered a random variable which follows a normal distribution with known variance (σ^2).

$$\boldsymbol{Z_T = Z_M + ME} \quad \text{Equation 1}$$

$$\boldsymbol{ME \sim N(0, \sigma^2)} \quad \text{Equation 2}$$

Although several methods have been proposed to correct for such problems associated with measurement error, they are infrequently used (Thürigen et al., 2000).

Corrective methods for measurement error have been developed for multivariable regression models including logistic regression, which is the focus of this project. In models with a binary outcome, Y, the outcome model is defined as the probability of having a positive outcome (or success), conditional on the true values of predictor variables measured with error, Z_T, and any number of other non-erroneous predictors, X, which is equal to a function of Z_T, X, and a vector of parameter estimates for the outcome model, β. (Equation 3).

$$\boldsymbol{P(Y = 1|Z_T, X) = h(Z_T, X, \beta)} \quad \text{Equation 3}$$

A model for the true values of the mis-measured variable, Z_T, can be defined as a function of the measured values, Z_M, and vector of parameter estimates for those measured values, λ, plus an error term for this measurement error model, δ.

$$\boldsymbol{Z_T = f_Z(Z_M, \lambda) + \delta} \quad \text{Equation 4}$$

Corrective techniques generally rely on the assumption that the measured values, Z_M, and the outcome, Y, are conditionally independent, given the true values of mis-measured variables, Z_T. (Thürigen et al., 2000). Practically, the error, ME, does not explain anything about Y beyond what is explained by the true values of Z.

$$\boldsymbol{f(Y|Z_T, Z_M, X) = f(Y|Z_T, X)} \quad \text{Equation 5}$$

Many correction methods intended to account for measurement error involve using a validation dataset, in which both the true value of the mis-measured variable, Z_T, and the value measured with error, Z_M, are present in either a sub-set of observations, or a related dataset (Rosner, Spiegelman & Willett, 1990). This data structure allows for an understanding of the relationship between the true and measured values, which is then used for calculation of corrected regression coefficient estimates. These methods assume that the probability of an observation/subject being present in the validation dataset is independent of any variables in the dataset (Thürigen et al., 2000). Finally, methods rely on the assumption of transferability, that the parameter estimates describing the relationship between the true and measured values from a validation dataset will converge on those of the main study, or that the measurement error model (equation 4) is transferable between the validation data and main study data (Thürigen et al., 2000).

Regression calibration methods of measurement error correction for logistic regression were developed in the late 1980's (Rosner, Spiegelman & Willett, 1989) (Armstrong, Whittemore, & Howe, 1989), and have been since used in and modified for multiple applications (Guolo, 2008). Here, a parametric model is constructed from validation data describing the association between true and measured values. In most cases, this validation (or measurement error) model is used to directly correct the coefficient estimates from an uncorrected model predicting the outcome of

interest, built using the main study dataset. Rosner et al. describes a common version this method for the case of multiple covariates measured with error (1990):

1. An uncorrected logistic regression model is built using the main study dataset, where mis-measured values and other variables are used to predict outcome (Y ~ Z_M + X). These uncorrected coefficient estimates are stored. Let β_Z be a vector of coefficient estimates for variables measured with error and β_X be a vector of coefficient estimates for non-erroneous variables.

$$\boldsymbol{logit}[\boldsymbol{P}(\boldsymbol{Y}|\boldsymbol{Z_M}, \boldsymbol{X})] = \boldsymbol{\alpha} + \boldsymbol{\beta_Z Z_M} + \boldsymbol{\beta_X X} \quad \text{Equation 6}$$

2. In the case of continuous variables measured with error, a multivariate linear regression model is fit using the validation dataset, with the true values of all covariates measured with error are regressed on the erroneous measurements and other variables. Let λ_Z be a matrix of regression coefficient estimates for variables measured with error and λ_X be a matrix of regression coefficient estimates for non-erroneous variables.

$$\boldsymbol{Z_T} = \boldsymbol{\alpha'} + \boldsymbol{\lambda_Z Z_M} + \boldsymbol{\lambda_X X} + \boldsymbol{e} \quad \text{Equation 7}$$

3. A vector of uncorrected coefficient estimates from step 1 is multiplied by a matrix of the coefficient estimates from step 2, augmented by the identity matrix and zero matrix, to calculate a vector of the corrected regression coefficients:

$$\boldsymbol{\beta^*} = \boldsymbol{\beta \lambda^{-1}} \quad \text{Equation 8}$$

Where β is a vector of uncorrected regression coefficients:

$$\boldsymbol{\beta} = (\boldsymbol{\beta_Z}, \boldsymbol{\beta_X}) \quad \text{Equation 9}$$

And β^* is a vector of corrected regression coefficients:

$$\boldsymbol{\beta}^* = (\boldsymbol{\beta_Z}^*, \boldsymbol{\beta_X}^*) \quad \text{Equation 10}$$

And letting *I* be the identity matrix, the matrix λ is given by:

$$\boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda_Z} & \boldsymbol{\lambda_X} \\ \mathbf{0} & \boldsymbol{I} \end{pmatrix} \quad \text{Equation 11}$$

Similar methods of regression calibration use a model built on the validation dataset to predict the true values, Z_T, using measured values and other variables ($Z_T \sim Z_M + X$), but then apply this model to the main dataset to replace the erroneous measurements with predictions of Z_M, adjusted values which can then be used directly to model the outcome (Guolo, 2008). Alternative models have also been used to calculate adjusted values. For instance, an empirical Bayes estimate of the true value conditional on the outcome in moment reconstruction methods (Kromhout, 2009), maximum likelihood approaches, which maximize a likelihood function of Y given Z_T and Z_M, semi- and non-parametric methods, which don't specify the distribution of the measurement error or measurement error model, and Bayesian methods could be considered (Thürigen et al., 2000).

In cases where true measurements are unavailable (or not applicable, as in our case of pain and function) a regression calibration model can be indirectly estimated using repeated measurements taken with error for some or all observations/subjects, and use of this type of correction often has the effect of strengthening the relationship between predictor and outcome (Freedman et al. 2004) (Wong et al., 1999). Such methods have been developed in cases where repeated measures are taken of erroneous variables derived from questionnaires, and where other variables may also be present, which is closely related to the present case regarding the KOOS predictors. Wong et al. describes how a correction factor for the coefficient of the erroneous variable can be calculated from the variance of the repeated measures and the correlation between them (1999). Bashir and Duffy also review various methods of dealing with data containing

predictors with repeated erroneous measurements, including linear imputation methods and a discriminant analysis method, in which it is assumed that repeated errors are independent and normally distributed (1997).

Imputation methods involve making predictions for the true values of the erroneous variable, Z_T, and treating measurement error as a missing data problem. Here, other variables, including the outcome, can be used fill in a missing (or mis-measured) value (Padilla et al., 2009). Similar to regression calibration, these methods typically rely on either validation data (Cole, Chu, & Greenland, 2006) or repeated measures taken on the same observation (Keogh & White, 2014) (Bashir & Duffy, 1997). When performing multiple imputations of a mismeasured variable, imputed values are treated as potentially true values, rather than a single true value, having the benefit of accounting for uncertainty about the actual value. This may prevent underestimation of standard errors and inflated type I error rates and may result in more conservative confidence intervals (Padilla et al., 2009). The following is a common version of a multiple imputation approach described generally in a review by Padilla et al. (2009):

1. A regression model is created predicting the variable measured with error using other variables in the dataset as predictors.
2. The coefficient estimates from this model, along with their variances-covariance matrix is used to generate a new set of M random parameter estimates. It is generally assumed that the generated estimates follow a normal distribution around the original estimate for each variable from step 1.
3. For each of the M sets of generated coefficient estimates, a prediction (or imputation) of a potential true value of the mis-measured variable is made.

4. The model of interest is fit using the M potential true values of the mis-measured variable.
5. The M sets of coefficient estimates from the model in step 4 are combined and their total variance is calculated as described by Rubin (1987).

Here, a similar method using multiple imputation to correct for measurement error was used but modified for our specific case. The dataset available for use in this project did not contain either validation data or repeated measures of mis-measured variables, preventing the use of standard regression calibration methods. However, limited information was known regarding the distribution of the measurement error typically associated with each KOOS predictor individually, described above in equation 2. Unfortunately, the covariance between measurement errors in KOOS predictors was unknown. For example, it was unclear how much of the error in a measurement of baseline pain can be explained by the error in a measurement of baseline function for the same subject. For this reason, the multiple imputation method used here could not involve the use of a variance-covariance matrix to generate imputations. A conservative approach was instead used in which there was no correlation between the errors in different KOOS variables. Rather than using the variability inherent in the measurement error model to randomly generate potentially true associations between mis-measured variables and other variables (as in step 2 above), we simply used the least-squares estimates from our measurement error model, but subsequently added in the known error distribution directly when imputing possible true values for the KOOS variables. While this approach was simple compared to many of those described, it retained the benefit of using multiple imputations to simulate the uncertainty inherent in measurement error (rather than attempting to calculate a single "corrected" value) and made the best use of our limited background information regarding the measurement of KOOS.

1.3 Objectives

This project sought to better understand patient outcome after APM. Specifically, we were interested in modeling whether patients are satisfied with their results at the 1-year follow-up point. Although a 10-unit improvement in pain (KOOS) is commonly used to indicate that treatment was a success, we posit that this is too simplistic since the relationship between KOOS and overall satisfaction is complicated, likely involving other potentially confounding factors which are necessary to take into account and control for in a potential model. It has been observed that some patients with great improvement in pain and function are not satisfied with their outcome, while others with moderate function at baseline, and who show only modest improvements in pain and function, are satisfied with their outcome. This project sought to identify what factors are most important in determining overall patient satisfaction. It was hypothesized that demographic predictors most closely related to other health comorbidities, such as age and BMI, would be most important to control for. This project also sought to determine how KOOS pain and function could best be used to predict whether patients will be satisfied with their surgical outcome. It was hypothesized that all four KOOS variables would be important in predicting ultimate patient satisfaction, but that this association would be less strong after considering the uncertainty associated with measurement error. It was our hope that by being able to better understand and predict patient outcomes, physicians could better identify ideal candidates for surgery and may be able to rule out those least likely to benefit.

2.0 Methods

2.1 Data

Data came from a study population taken from the OrthoMiDaS Episode of Care (OME) prospective surgical cohort at the Cleveland Clinic. The OME cohort, which includes surgical and outcome data from elective knee, hip, and shoulder surgeries performed at all Cleveland Clinic sites, was collected from February 2015 to December 2017. Out of this cohort, only APM surgical patients (with meniscal tears) who had valid MRI data were selected for this project. Patients under the age of 45, as well as those who had received previous knee surgery or who underwent another concurrent surgery were additionally excluded, so as to focus on typical cases in individuals with a higher prevalence of osteoarthritis (OA) (N=924).

The primary outcome of interest was the binary variable Patient Acceptable Symptom State (PASS), whether or not the patient reported overall satisfaction with their surgical outcome at the 1-year follow-up timepoint after APM (Wright et al., 2015). Of primary interest as predictors were the continuous patient-reported KOOS variables: baseline pain, baseline function, 1-year change in pain, and 1-year change in function (Roos & Lohmander, 2003). Also of interest were a selection of demographic variables, summarized in Table 1. Categorical variables were summarized with counts and percentages, and continuous variables were summarized with mean, standard deviation (SD), and range. Clinical recommendations were considered in cases where the linear form of a continuous variable was deemed not appropriate. Specifically, the variables body mass index (BMI), area deprivation index (ADI), and mental component summary (MCS) were used in both their quadratic (squared) and original linear forms. Additionally, the variables age, years of

education, and comorbidity index were each categorized based on clinically recommended split-points.

2.2 Descriptive Statistics

The primary predictors of PASS, the KOOS variables, were graphically compared using side-by-side boxplots, demonstrating the difference in distribution of baseline pain, baseline function, change in pain, and change in function by overall patient satisfaction. Each boxplot was visually assessed for normality. A 2-sample t-test with unequal variances was performed for each KOOS variable under the null hypothesis that the mean scores were equal in satisfied and unsatisfied patients. The assumptions of this test were verified; subjects were independent, and the central limit theorem was satisfied by the large size of each group (considerably greater that n=30).

Continuous demographic variables were similarly graphically summarized by PASS outcome status using side-by-side boxplots. Categorical predictor variables were graphically summarized by PASS outcome status using side-by-side bar graphs. This preliminary examination of plots was intended to qualitatively assess the relationship between each predictor and patient satisfaction.

Interactions between baseline pain with change in pain, baseline function with change in function, and baseline function with baseline pain were also assessed. Scatterplots were used to visually examine the relationship between each predictor in the interaction and stratified by outcome status. Flexible lines (using a generalized additive model) were fit to each stratum and the relationship between the slopes of each line was examined to assess the extent of the interaction. Logistic regression, using PASS as the outcome and each interaction, along with the

two individual variables, as the only predictors, was used to assess the significance of each interaction individually, using the p-value from the Wald test.

2.3 Modeling PASS

Logistic regression (James et al., 2017) was used to model patient satisfaction. All covariates were regressed on PASS using a logit link in order to model the probability of each patient being satisfied. Let Y be the outcome variable PASS and X a vector of covariates.

$$\boldsymbol{p(Y) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}} \qquad \text{Equation 12}$$

Or equivalently:

$$\boldsymbol{\log\left(\frac{p(Y)}{1-p(Y)}\right) = \beta_0 + \beta_1 X} \qquad \text{Equation 13}$$

Coefficients were estimated using the general method of maximum likelihood, such that the predicted probability of satisfaction for each subject is as close as possible to that individual's satisfaction status.

This full uncorrected model contained all covariates shown in table 1. Based on clinical recommendations provided at the time of data acquisition, both the linear and second-order polynomial transformations are used in the model for the variables body mass index (BMI), National Area Deprivation Index (ADI), and Mental Health Score (MCS). In the case of categorical predictors, dummy variables were created for all but the first (or lowest) level of each variable, such that each could only take on two possible numeric values (i.e. 1 or 0 corresponding to yes or

no), and these were used as predictors in the regression model. Interactions of potential clinical relevancy were also included between the primary (KOOS) variables of interest, specifically interactions between baseline pain with change in pain, baseline function with change in function, and baseline function with baseline pain.

Model fit was assessed using several diagnostic methods. To evaluate overall model fit, the scaled deviance and pseudo-R^2 metrics were calculated, and the Hosmer-Lemeshow goodness-of-fit test was performed under the null hypothesis that the predicted and observed probabilities were similar (i.e. high model fit). A residual plot was examined to assess the assumption that the logistic model was correctly specified, or that data fits linearly on the logit scale. The presence of outliers was investigated using a standardized Pearson residual plot. A Pregibon leverage plot was examined to assess whether observations had extreme covariate patterns, a plot of Pregibon delta-beta (dbeta) statistics was used to detect high influence, and variance inflation factors (VIF) were used to assess multicollinearity.

2.3.1 Variable Selection

A reduced model was constructed to eliminate variables not contributing significantly to the prediction of PASS, to improve the overall interpretability of the model, and to prevent overfitting. A reverse stepwise selection technique was employed. Starting with the full model, reduced models of size p-1 were built by removing each variable one at a time, where p is the total number of predictors. The best of these was selected based on that which had the smallest deviance. Then, beginning with the p-1 sized model, variables were again removed one at a time and a best model of size p-2 was chosen in the same fashion. This was repeated until the best single-variable model was reached. The result was a selection of p+1 models with the number of variables ranging

from the null model (0 predictors) to the original full model (p predictors). The model with the lowest AIC was then selected among these p+1 best models of different sizes.

To better understand the importance of each variable and the consistency with which it could be used predict patient satisfaction, a bootstrapping technique was employed. Out of the 828 observations in the dataset, a sample of those observations of size N=828 was randomly drawn 50 times with replacement. Each time, the best p+1 models were selected based on deviance, and the best size model was selected based on AIC. Using this resampling technique, the percentage of times (or the percentage of bootstrapped samples) in which each variable was selected was calculated. Similarly, the percentage of times in which each variable was significant was calculated. Based on the combination of these findings and clinical judgement about the relevance of each variable in predicting patient satisfaction, a selection of variables was made, and a reduced uncorrected model was constructed. Particularly in the case of interactions, the meaningfulness of these covariates on the probability of satisfaction, in addition to statistical significance, was also taken into consideration.

The diagnostic methods of evaluating model fit and the presence of problematic points used for the full uncorrected model was again used to assess this reduced uncorrected model. Additionally, the performance of full and reduced models was compared. A c-statistic was calculated to measure discrimination between satisfied and unsatisfied patients, which was graphically represented using an ROC curve (James et al., 2017). A higher c-statistic (closer to 1) indicates stronger model discrimination. Brier score (BS), the mean of the square of the differences between the true PASS outcome and predicted probability of satisfaction for each observation, was calculated for each model (Wu & Lee, 2014). Here, a lower BS (close to 0) indicates more accurate

predictions. Where t is the observation number from 1 to N, f_t is the probability forecasted by the model for the t-th observation, and o_t is the actual outcome for that observation, the brier score is:

$$\boldsymbol{BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2} \qquad \text{Equation 14}$$

Additionally, AIC and pseudo-R^2 was used to compare models.

2.3.2 Measurement Error

To correct for the potentially biased regression coefficient estimates and the uncertainty of these estimates due to the presence of within-subject variability in baseline and change in pain and function KOOS variables, a multiple-imputation method of correcting for measurement error (MIME) was used to calculate corrected regression coefficient estimates. The following planned procedure was modified from that used by Cole et al. (2006), Padilla et al. (2009), and Keogh & White (2014), to accommodate the lack of a validation study or repeated measures, but with prior knowledge of the distribution of measurement errors:

1. Let Z_1, Z_2, Z_3, and Z_4 be the recorded values of four variables subject to measurement error, Y be the outcome, and X_i be a vector of all other predictors free of measurement error. Each erroneous variable of interest, Z, was regressed on all other predictors, X_1,…, X_p, and other Z_i, using the linear model in equation 4.

$$\boldsymbol{Z_1 \sim Z_2 + Z_3 + Z_4 + X_1 + \cdots + X_p} \qquad \text{Equation 15}$$

2. Let Z* be the predicted value of Z, α_0,…, α_p, and α_Y be the regression coefficient estimates from the linear model in equation 4, and e be the error with which Z was measured. It is

assumed that e follows a normal distribution with mean zero and known variance, σ^2, such that:

$$\boldsymbol{e \sim N(0, \sigma^2)} \qquad \text{Equation 16}$$

The regression coefficients estimated in step 1 were stored and used to predict the value of each Z_i using equation 6. The resulting predictions, Z_i*, were considered to be possible true values of Z_i, which have been subjected to measurement error with variance σ^2.

$$\begin{aligned}
\boldsymbol{Z_1^*} &= \boldsymbol{\alpha_0 + \alpha_1 Z_2 + \alpha_2 Z_3 + \alpha_3 Z_4 + \alpha_4 X_1 + \cdots + \alpha_{p+3} X_p + e} \\
\boldsymbol{Z_2^*} &= \boldsymbol{\alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_3 + \alpha_3 Z_4 + \alpha_4 X_1 + \cdots + \alpha_{p+3} X_p + e} \\
\boldsymbol{Z_3^*} &= \boldsymbol{\alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_4 + \alpha_4 X_1 + \cdots + \alpha_{p+3} X_p + e} \\
\boldsymbol{Z_4^*} &= \boldsymbol{\alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \alpha_4 X_1 + \cdots + \alpha_{p+3} X_p + e}
\end{aligned} \qquad \text{Equation 17}$$

3. Step 2 was repeated 10 times, each time with a randomly simulated value of e, under the distribution in equation 5, resulting in 10 probable true scores for each Z_i. Here, the first subscript indicates which variable and the second indicates which imputation.

$$\begin{gathered}
\boldsymbol{Z_{1,1}^*, Z_{1,2}^*, \ldots, Z_{1,10}^*} \\
\vdots \\
\boldsymbol{Z_{4,1}^*, Z_{4,2}^*, \ldots, Z_{4,10}^*}
\end{gathered} \qquad \text{Equation 18}$$

The similarly between the measured values, Z, and imputed predictions, Z*, was calculated using the mean squared error (MSE). Here, i is the ith KOOS variable, j is the jth imputation, and k is the kth observation.

$$\boldsymbol{MSE_{i,j,k} = \frac{1}{10} \sum_{j=1}^{10} \frac{1}{N} \sum_{k=1}^{N} (Z_i - Z_i^*)^2} \qquad \text{Equation 19}$$

4. The outcome Y was regressed on all predictors, including Z_1*, …, Z_4*, and X_1,…, X_p, in a logistic regression model. This was repeated 10 times, once for each set of imputed Z^*.

$$\boldsymbol{Y \sim Z^*_{1,1} + \cdots + Z^*_{4,1} + X_1 + \cdots + X_p}$$

$$\vdots$$

$$\boldsymbol{Y \sim Z^*_{1,10} + \cdots + Z^*_{4,10} + X_1 + \cdots + X_p} \qquad \text{Equation 20}$$

Let β_0, β_1, …, β_p, and β_Z be the resulting regression coefficients of the above model (equation 19). Different values of Z* for each of the 10 imputations gave slightly different sets of regression coefficients:

$$\begin{pmatrix} \widehat{\beta_0} \\ \vdots \\ \widehat{\beta_p} \end{pmatrix}_1, \ldots, \begin{pmatrix} \widehat{\beta_0} \\ \vdots \\ \widehat{\beta_p} \end{pmatrix}_{10}$$

Each with estimated standard errors:

$$\begin{pmatrix} \widehat{\sigma_0} \\ \vdots \\ \widehat{\sigma_p} \end{pmatrix}_1, \ldots, \begin{pmatrix} \widehat{\sigma_0} \\ \vdots \\ \widehat{\sigma_p} \end{pmatrix}_{10}$$

5. An estimate of the corrected regression coefficient for each variable was the average of the m estimates for β_i, as described by Yuan (2010). The subscript *i* is an indicator for all variables being considered in this model (1 through p for X variables, and Z1 through Z4 for KOOS variables), and the subscript *j* is an indicator for imputations 1 through 10.

$$\boldsymbol{\overline{\beta_i} = \frac{1}{10}\sum_{j=1}^{10} \widehat{\beta_{ij}}} \qquad \text{Equation 21}$$

The final corrected model was constructed using these averaged regression coefficients:

$$\log\left(\frac{Y}{1-Y}\right) = \overline{\beta}_0 + \overline{\beta}_{Z1}Z_1 + \cdots \overline{\beta}_{Z4}Z_4 + \overline{\beta}_1 X_1 + \cdots + \overline{\beta}_p X_p \quad \text{Equation 22}$$

6. An estimate of the variance associated with each corrected coefficient was also calculated (Yuan, 2010) as both a function of both the within-imputation variance, σ_w^2, and between imputation variance, σ_b^2, giving the total variance of the regression coefficient estimate, σ_T^2. For these variance terms, the associated subscript *w* stands for "within", *b* stands for "between", and *T* stands for "total".

$$\sigma_{i_w}^2 = \frac{1}{10}\sum_{j=1}^{10} \widehat{\sigma_{i_j}} \quad \text{Equation 23}$$

$$\sigma_{i_b}^2 = \frac{1}{10-1}\sum_{j=1}^{10} (\beta_{i_j} - \overline{\beta_i})^2 \quad \text{Equation 24}$$

$$\sigma_{i_T}^2 = \sigma_{i_w}^2 + (1 + \frac{1}{10})\sigma_{i_b}^2 \quad \text{Equation 25}$$

7. A test statistic for β_i was constructed as described by Rubin (1987), which follows a t-distribution:

$$(\beta - \overline{\beta_i}) \Big/ \sqrt{\sigma_{i_T}^2} \sim T(v) \quad \text{Equation 26}$$

With *v* degrees of freedom:

$$v = (10-1)(1 + \frac{\sigma_{i_w}^2}{(1 - 1/10)\sigma_{i_b}^2})^2 \quad \text{Equation 27}$$

The adjusted coefficients of this final corrected reduced model were compared to those of the uncorrected reduced model. Additionally, estimates for the odds ratios associated with an increase in pain or function were also calculated, compared to the uncorrected model, and used to interpret the effects of these variables on patient satisfaction. The statistical significance of each corrected regression coefficient estimate was calculated using the test described above and compared to the Wald test results from the uncorrected model. Finally, predictions of the probability of patient satisfaction were made based on this corrected model. These were used to calculate a C-statistic, brier score, and ROC curve, in order to compare the performance of models.

2.4 Statistical Details

For all hypothesis tests, a p-value of 0.05 was used as the threshold for statistical significance. Such tests include the t-tests and chi-squared tests used in preliminary analysis, as well as Wald tests and likelihood ratio tests from various logistic regression models. All applicable assumptions of such tests were checked using visual inspection or knowledge of sample characteristics when appropriate. R-studio software was used for all analyses.

3.0 Results

3.1 Descriptive Statistics

3.1.1 Univariate Summary

Data, including primary KOOS predictors and demographic variables, were summarized in table 1. Counts with percentages were presented for categorical predictors and means with standard deviation (SD) and ranges were presented for continuous predictors. The continuous variables age, education, and comorbidity index were additionally categorized based on clinical recommendations.

Table 1. Summary of Demographics and KOOS in Study Sample

Continuous Variables	**Mean (Standard Deviation)**
Baseline Pain (KOOS)	46.8 (17.2)
Baseline Function (KOOS)	45.7 (16.5)
Change in Pain (KOOS)	29.0 (22.1)
Change in Function (KOOS)	19.0 (20.1)
BMI	30.4 (6.4)
Income	79564 (37638)
Mental Component Summary (MCS) Score	54.0 (9.8)
National Area Deprivation Index (ADI)	42.5 (25.2)
Categorical Variables	**Count (Percent of Sample)**
Age	
<50	167 (20.2%)

50-70	611 (73.8%)
>70	50 (6.0%)
Sex	
Female	410 (49.5%)
Male	418 (50.5%)
Race	
White	735 (88.8%)
Non-white	93 (11.2%)
Ethnicity	
Non-Hispanic or -Latino	797 (96.3%)
Hispanic or Latino	31 (3.7%)
Years of Education	
≤12 years	327 (39.5%)
13-16 years	354 (42.8%)
>16 years	147 (17.7%)
Smoking status	
Never or quit>6m	482 (59.1%)
Quit<6m or current	336 (40.9%)
Insurance	
Medicaid or Uninsured	185 (22.3%)
Medicare or Private	643 (77.7%)
Comorbidity index	
0: No comorbidities	551 (65.9%)
1 or 2: Low comorbidities	229 (27.5%)
>2: High comorbidities	48 (5.8%)

3.1.2 Bivariate Summary

3.1.2.1 KOOS Predictors by PASS Status

Data were then stratified by outcome (PASS) status. There were 292 subjects who reported being unsatisfied with their surgical outcome and 617 subjects who reported being satisfied. Means and standard deviations (SD) of the primary KOOS predictors, stratified by outcome status, are shown in table 2. Also shown were corresponding p-values from 2-sample t-tests for evaluating whether the mean scores were different between the outcome groups (table 2). In all cases, these results were highly statistically significant. Patients who had higher baseline pain and lower baseline function, as well as those who experienced a greater reduction of pain and greater improvement in function after surgery were more likely to be satisfied by their outcome.

Table 2. KOOS Predictors by Outcome Status

	Unsatisfied, n = 267	**Satisfied, n = 561**	
KOOS Variables	**Mean (SD)**	**Mean (SD)**	**P-value**
Baseline Pain	42.1 (16.1)	49.5 (17.1)	<0.001
Baseline Function	50.3 (17.3)	43.2 (15.6)	<0.001
Change in Pain	12.4 (20.4)	36.9 (18.1)	<0.001
Change in Function	6.2 (18.9)	25.0 (17.6)	<0.001

Figure 1 shows the distribution of each KOOS predictor by outcome status. These figures demonstrate the same associations between KOOS and PASS described and tested above.

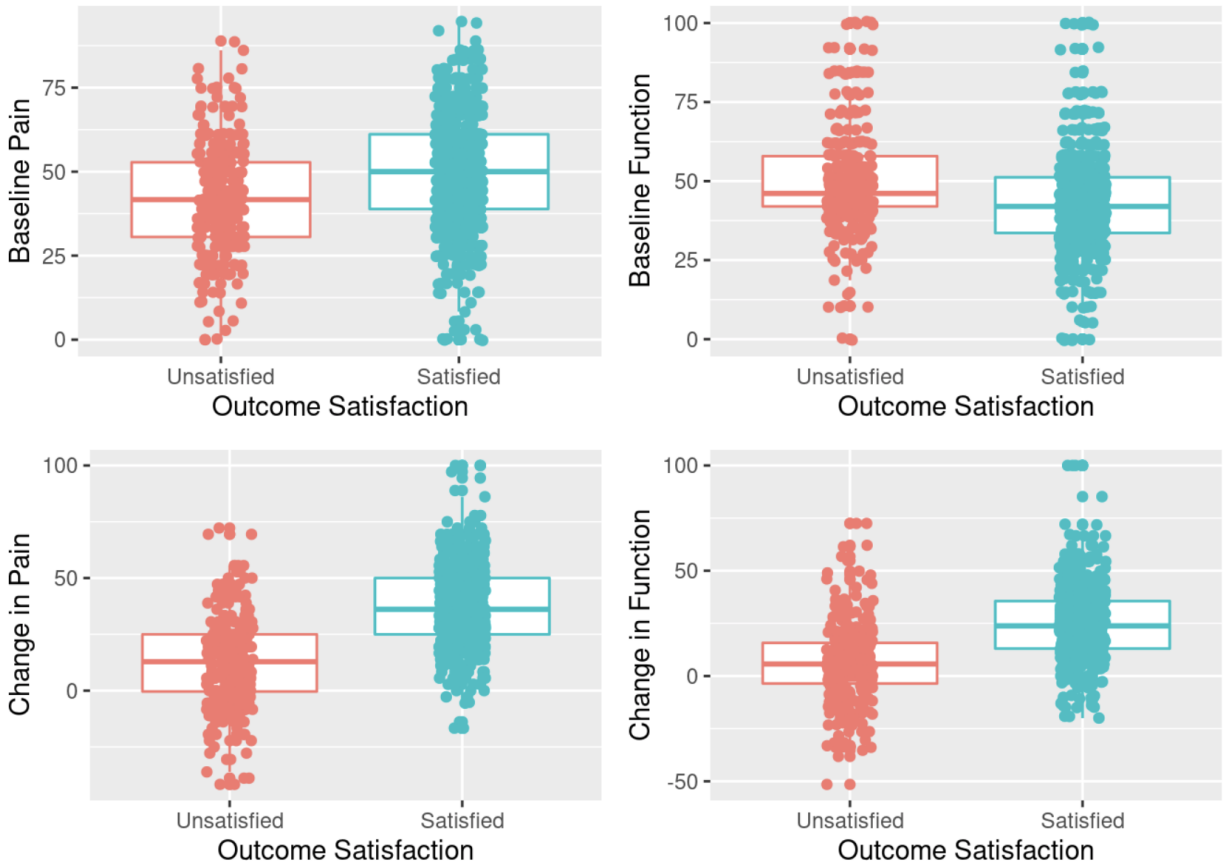

Figure 1. Current Satisfaction vs. Pain and Function (KOOS).

Outcome status (PASS) is used to separate patients based on whether they were satisfied (blue) or unsatisfied (red). Boxplot is overlayed, showing mean (center) and 1st and 3rd quartiles of pain and function distributions.

3.1.2.2 Demographic Predictors by PASS Status

Demographic predictors were stratified by outcome status. Means and standard deviations in each outcome category are shown for continuous variables, while counts in each outcome category are shown for categorical variables in table 3. Two-sample t-tests were used for continuous predictors, while chi-squared tests were used for categorical predictors, to determine whether the means and distributions of these predictors differed by outcome status. Sex, BMI, education, income, MCS score, comorbidities, and national ADI were significantly associated with surgical satisfaction (table 3).

Table 3: Demographic Predictors by Outcome Status

	Unsatisfied, n = 267	**Satisfied, n = 561**	
Continuous Variables	**Mean (SD)**	**Mean (SD)**	**P-value (t-test)**
BMI	31.7 (6.6)	29.9 (6.2)	<0.001
Income	74,787 (36517)	81,820 (37977)	0.008
Mental Component Summary (MCS) Score	51.3 (12.3)	55.3 (7.9)	0.001
National ADI	46.6 (26.7)	40.6 (24.1)	0.001
Categorical Variables	**Count (percentage of unsatisfied)**	**Count (percentage of satisfied)**	**P-value (Chi-squared)**
Age			0.374
<50	59 (22.1%)	108 (19.3%)	
50-70	189 (70.8%)	422 (75.2%)	
>70	19 (7.1%)	31 (5.5%)	
Sex			0.003
Female	152 (56.9%)	258 (46.0%)	
Male	115 (43.1%)	303 (54.0%)	
Race			0.238
White	232 (86.9%)	503 (89.7%)	
Non-white	35 (13.1%)	58 (10.3%)	
Ethnicity			0.694
Non-Hispanic or -Latino	256 (95.9%)	541 (96.4%)	
Hispanic or Latino	11 (4.1%)	20 (3.6%)	
Education			0.014
≤12	109 (38.2%)	218 (38.9%)	
13-16 years	98 (36.7%)	256 (45.6%)	
>16 years	60 (22.5%)	87 (15.5%)	
Smoking status:			0.688
Never or quit>6m	156 (58.4%)	336 (59.9%)	

quit<6m or current	111 (41.6%)	225 (40.1%)	
Insurance			0.340
Medicaid or Uninsured	65 (24.3%)	102 (18.2%)	
Medicare or Private	202 (75.7%)	441 (81.8%)	
Comorbidity Index			0.219
No comorbidities	167 (62.5%)	384 (68.4%)	
Low comorbidities	84 (31.5%)	145 (25.8%)	
High comorbidities	16 (6.0%)	32 (5.7%)	

The distributions of continuous predictors by outcome status are shown in figure 2. Following the results in table 3, the mean BMI of unsatisfied patients was higher than that of satisfied patients, the mean income of unsatisfied patients was lower than that of satisfied patients, the mean MCS score of unsatisfied patients was lower (worse mental health) than that of satisfied patients, and the mean national ADI of unsatisfied patients was (greater deprivation) than that of satisfied patients.

The distributions of categorical predictors by outcome status are shown in figure 3. Following the results in table 3, the proportion of satisfied individuals was similar across age, race, ethnicity, smoking insurance, and comorbidity groups. A greater proportion of females were unsatisfied than males, and the proportion of satisfied patients in the medium (13-16 year) educated group was less than the proportion of satisfied patients in the highly educated group (>16 years).

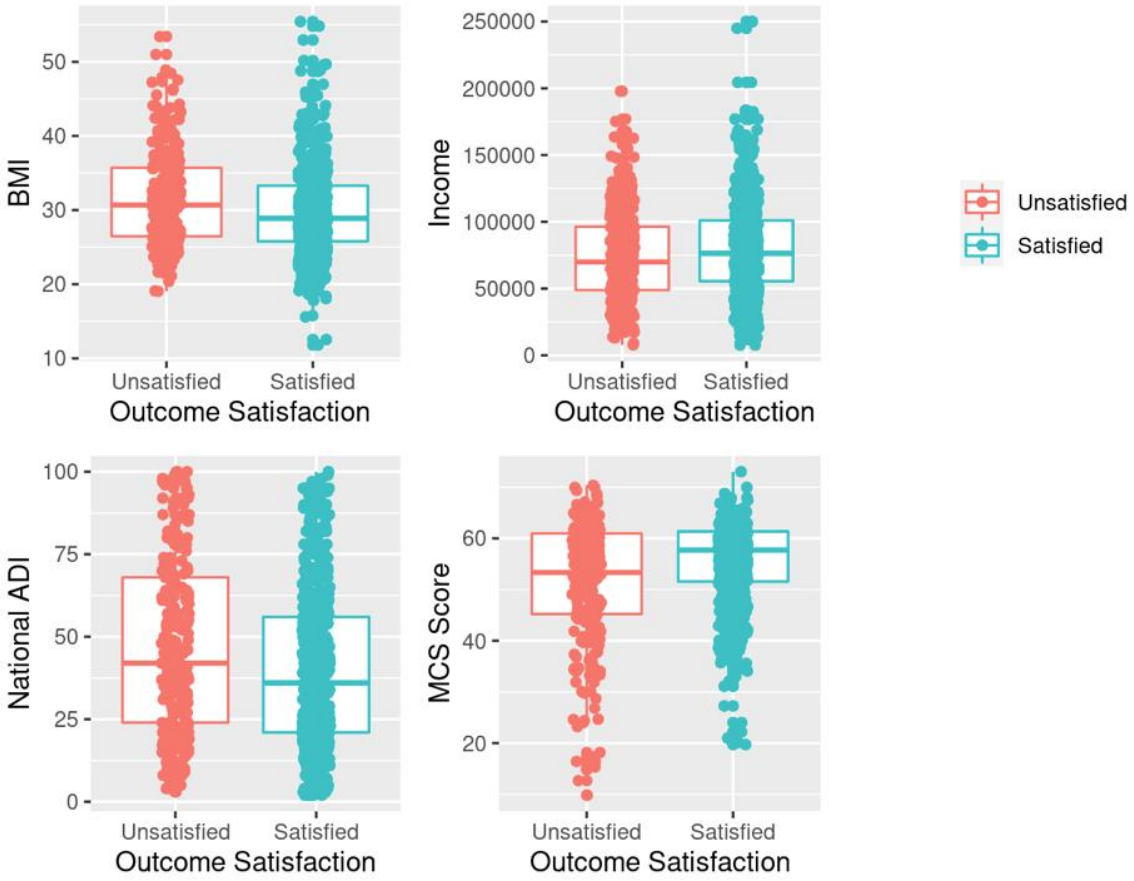

Figure 2. Current Satisfaction vs. Continuous Demographic Predictors.

Outcome status (PASS) was used to separate patients based on whether they were satisfied (blue) or unsatisfied (red). Boxplot is overlayed, showing mean (center) and 1st and 3rd quartiles of pain and function distributions.

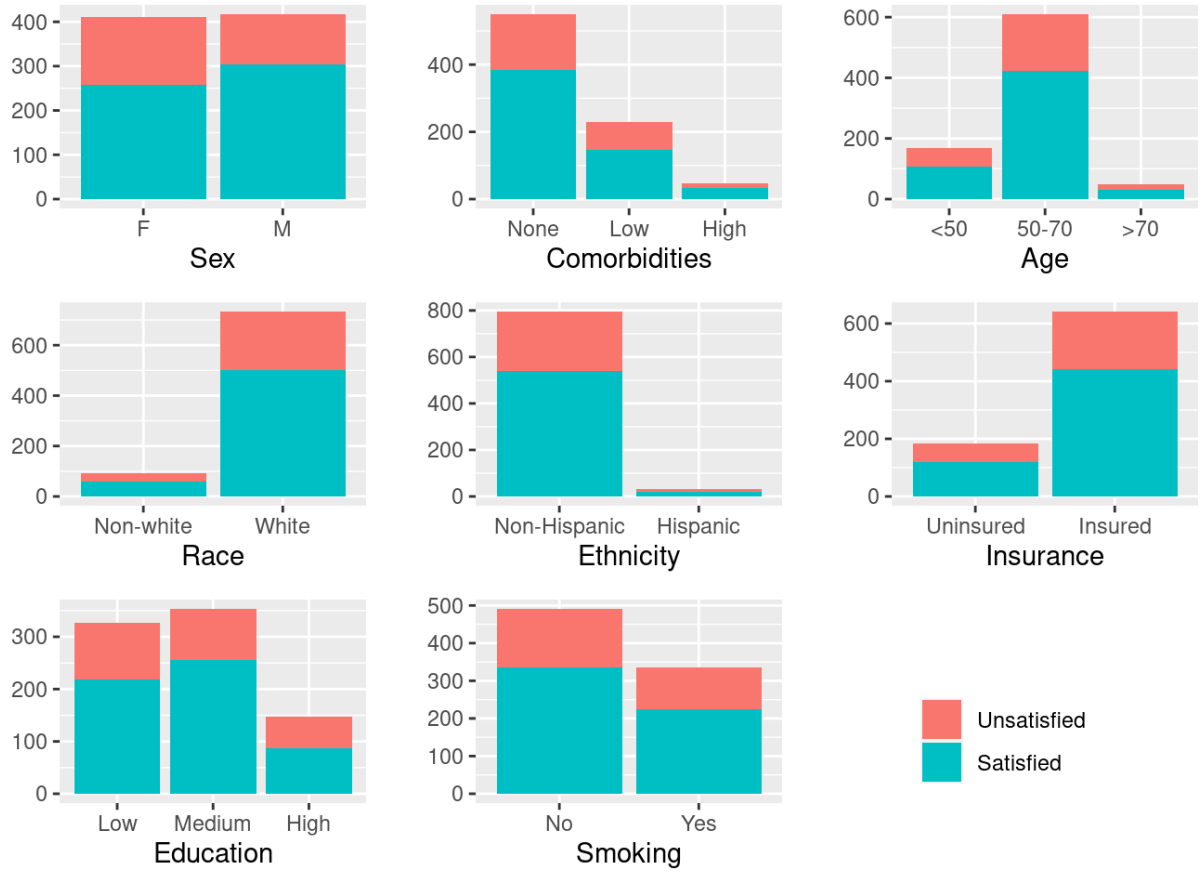

Figure 3. Current Satisfaction vs. Categorical Demographic Predictors.

Outcome status (PASS) was used to stratify patients based on whether they were satisfied (blue) or unsatisfied (red). Total height of bars represent the number of subjects in each group, further divided into colored portions of bars representing the number of subjects in each outcome status of each group.

3.1.3 Potential Interactions

Potential clinically suggested interactions between the primary KOOS predictors were examined prior to modelling. Specifically, interactions between baseline pain and function, between baseline function and change in function, and between baseline pain and change in pain are shown in figure 4. In each case, there were no more than minor differences identified between

outcome statuses (satisfied and unsatisfied subjects) in the slope of a line of best fit graphed between two KOOS predictors, suggesting that interactions were minimal (Figure 4). Logistic regression with each interaction (and both main effects) as the covariates revealed that the interaction between baseline pain and baseline function was not statistically significant ($p = 0.94$), while the interactions between baseline pain and change in pain, and between baseline function and change in function were statistically significant (both $p<0.001$).

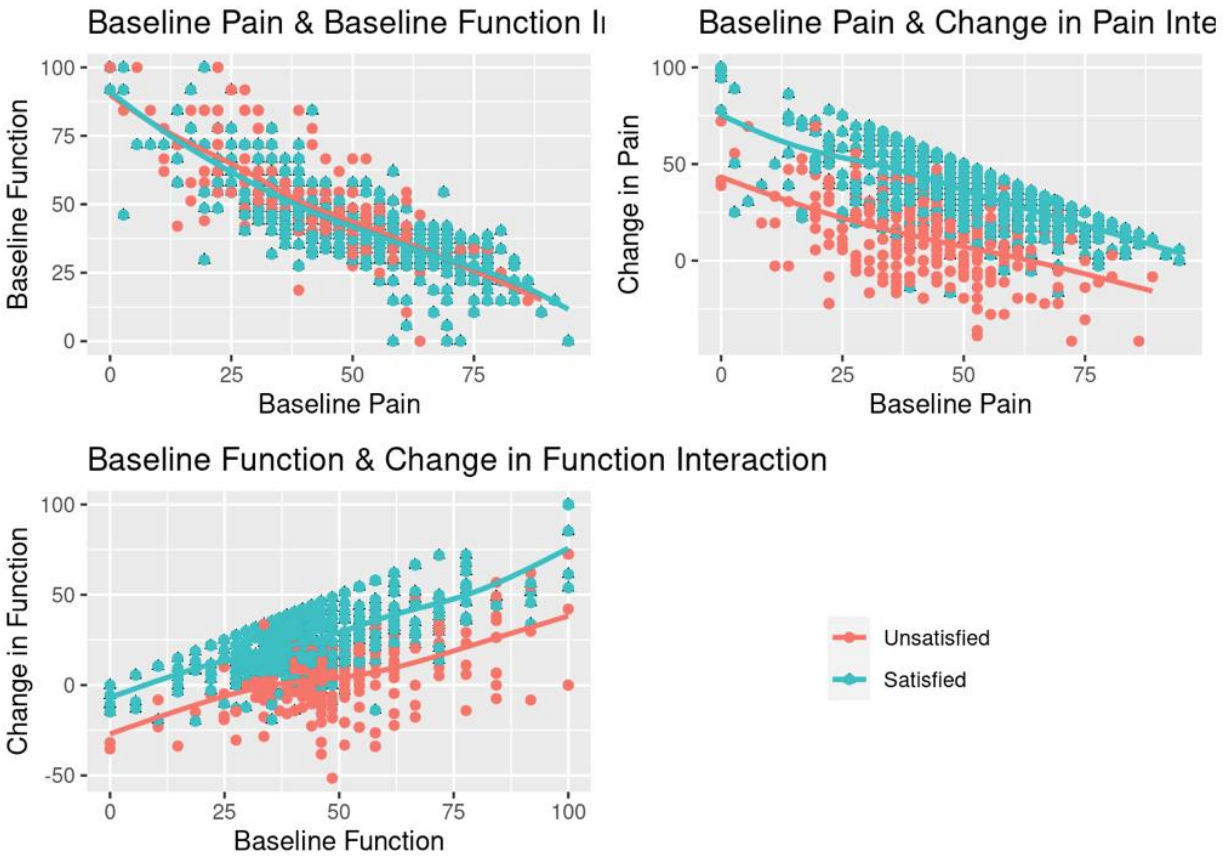

Figure 4. Interactions Between KOOS Variables.

A scatterplot between two KOOS predictors is shown. Points were colored based on outcome status (blue: satisfied, red: unsatisfied). A flexible line of best fit was fitted to each outcome status using a generalized additive model (GAM). Differences in the slopes between lines at the same point indicate the effect of one predictor differs based on the value of the other.

3.2 Modeling PASS

3.2.1 Full Uncorrected Model

The full uncorrected model was built using logistic regression and contained all KOOS variables, interactions, and demographic variables. As no correction was applied for the KOOS variables measured with error, the estimates for those predictors may be biased, which could also slightly alter the estimates of other predictors. Additionally, although all KOOS predictors were significant, when considered as part of interaction(s), many of the demographic predictors were not (Table 4).

The full model was then evaluated for performance and fit (Table 7). In the overall likelihood ratio test, this model predicted the outcome statistically significantly better than the null model (538.8, df=26, $p<0.0001$). The Hosmer-Lemeshow Goodness of Fit test was not statistically significant, indicating no significant deviation of fit (8.48, df = 8, p = 0.388). An ROC curve suggested discrimination was very strong, with an AUC of 0.936. The Brier score is low (0.092), indicating that the predicted probabilities of PASS are close to the actual outcomes. Additionally, AIC and BIC were 554.3 and 656.8 respectively, and pseudo-R^2 was 0.518. Additional model diagnostics and plots, shown in appendix C, did not reveal substantial deviations of model fit.

Table 4. Logistic regression results of full uncorrected model

	Coefficient Estimate	**P-value (Wald test)**
Intercept	-12.18	0.002
Baseline Pain (KOOS)	0.121	<0.001
Change in Pain (KOOS)	0.0632	0.002

Baseline Function (KOOS)	0.0281	0.317
Change in function (KOOS)	0.0859	<0.001
Age (reference category <45)		
[50,70)	0.157	0.595
[70,inf)	0.686	0.245
Sex (reference Female)		
Male	0.0437	0.858
BMI	0.108	0.442
BMI^2	-0.00148	0.479
Race (reference white)		
Non-white	-0.634	0.147
Ethnicity (reference non-hispanic)		
Hispanic or Latino	-0.513	0.359
Insurance (reference insured)		
Medicaid or uninsured	-0.0490	0.869
Income	-0.00000219	0.740
Comorbidities (reference Medium)		
Low	-0.177	0.496
High	0.463	0.353
Education (reference <13 years)		
13-16 years	-0.667	0.018
>16 years	-1.34	<0.001
National Area Deprivation Index (ADI)	-0.000676	0.980
ADI^2	0.0000552	0.801
Mental Health Score (MCS)	0.193	0.008
MCS^2	-0.00199	0.008
Smoking (reference No)		
Yes	-0.00969	0.967
Baseline pain x change in pain	0.000955	0.026
Baseline function x change in function	-0.000822	0.026

Baseline pain x baseline function	-0.00121	0.013

3.2.2 Variable Selection and Reduced Model

Backwards stepwise selection was performed on 50 bootstrapped samples from the original dataset. Each time, all variables in the full model were considered for removal. The percentage of time, out of the 50 bootstrapping iterations, that each variable was selected for the "best" model is shown in table 5. Additionally, the percentage of the 50 "best" models in which each variable was significant is shown in table 5.

Table 5. Bootstapping and Model Building Results.

	Covariates Selected (%)	**Significant (%)**
Baseline Pain	100	100
Change in Pain	100	88
Baseline Function	100	28
Change in function	96	100
Age	36	83.3
Sex	20	60
BMI BMI^2	30 24	53.3 58.3
Race	46	69.6
Ethnicity	46	65.2
Insurance	30	66.7
Income	20	50
Comorbidities	38	36.8
Education	94	97.9

National Area Deprivation Index (ADI) ADI^2	44 40	63.6 60.0
Mental Health Score (MCS) MCS^2	88 88	86.4 86.4
Smoking	16	50
Baseline pain x change in pain	74	83.8
Baseline function x change in function	68	79.4
Baseline pain x baseline function	74	75.7

The four KOOS variables, baseline and change in pain function, were the top performing variables in this procedure and selected for a reduced model. They were each selected for the best model in over 96% of bootstrapped samples. Baseline pain and change in function were statistically significant 100% of the time, whereas baseline function and change in pain were statistically significant 88% and 28% of the time respectively. These lower significance percentages were due to the presence of interaction terms which were selected only approximately 72% of the time. In cases where interactions were present, this considerably reduced the significance level of each KOOS variable in that interaction. Because of this, all four KOOS variables were deemed important and necessary individually, even though their significance percentages were influenced by the inclusion of interactions.

The top performing demographic variables selected for a reduced model were education, which was selected 94% of the time and significant 98% of the time; MCS and the quadratic MCS^2, which were each selected 88% of the time and significant 86% of the time; race, which was only selected 46% of the time but significant 70% of the time; and age, which was only selected 36% of the time but significant 83% of the time. These decisions made regarding which variables were top performing were somewhat subjective, considering both selection and significance percentage

metrics, significance level in the full model, descriptive statistics, and clinical recommendation. It is often desirable to control for age, for example, in many clinical modeling situations. Finally, no demographic variables were included which were expected to have substantial overlap, such as race and ethnicity for example.

Although interactions performed relatively well, selected approximately 72% of the time, and significant approximately 80% of the time, the decision was made to exclude these from the final model. The magnitude of these coefficients, or their difference from zero, was much smaller than the coefficients of the individual KOOS variables, by approximately a factor of 10 (Table 4). Considering the large sample size, these interactions were likely large enough to be statistically significant, but not clinically meaningful. The same conclusion was apparent in the interaction plots (figure 4). Additionally, the interpretation of the effect of each KOOS variable on the odds of satisfaction would be complicated if that variable were present in an interaction, and therefore interactions were not optimal for our interpretative objectives.

A reduced uncorrected model is shown in table 6. The Wald tests for all the selected variables described above were statistically significant in this model, except for age. In the overall likelihood ratio test, this model predicted the outcome statistically significantly better than the null model (520.1, df=12, $p<0.0001$). The Hosmer-Lemeshow Goodness of fit test was not statistically significant, indicating no significant deviation of fit (9.482, df = 8, $p = 0.303$). Discrimination was very strong, with an AUC of 0.930. The Brier score is low (0.098), indicating that the predicted probabilities of PASS are close to the actual outcomes. AIC and BIC were 545.1 and 588.2 respectively, and pseudo-R^2 was 0.50. These metrics are compared to the full model in table 7, and indicate that there was almost no sacrifice in performance when less-important demographic

predictors were removed. Additional model diagnostics and plots, shown in appendix C, did not reveal substantial deviations of model fit.

Table 6. Reduced Uncorrected Model

	Coefficient Estimate	**P-value (Wald test)**
Intercept	-7.899	<0.001
Baseline Pain*	0.080	<0.001
Change in Pain*	0.095	<0.001
Baseline Function*	-0.040	0.016
Change in Function*	0.047	0.002
Age (reference category <45)		
[50,70)	0.086	0.761
[70,inf)	0.619	0.256
Race (reference white)		
Non-white	-0.922	0.016
Education (reference <13 years)		
13-16 years	-6.68	0.012
>16 years	-1.38	<0.001
Mental Health Score (MCS)	0.215	0.005
MCS^2	-0.002	0.004

* Measured with error/variability

Table 7. Full and Reduced Model Comparison

	Full Model	**Reduced Model**
LRT (χ^2)	538.8 (df = 26) (p = 6.64e-98)	520.1 (df = 12) (**p = 1.66e-104**)
HL GOF Test (χ^2)	8.485 (df = 8) (**p = 0.388**)	9.482 (df = 8) (p = 0.303)
Pseudo-R2	**0.518**	0.500

AIC BIC	554.3 656.8	**545.1** **588.2**
AUC	**0.936**	0.930
BS	0.092	0.098

3.2.3 Correcting for Measurement Variability

The multiple imputation method of correcting for measurement error was performed with 10 imputations, using only the variables from the reduced model. The MSE describing the average squared difference between measured and imputed values for baseline pain was 61.0, or 7.8 describing the unsquared difference. The MSE describing the average squared difference between measured and imputed values for baseline function was 57.2, or 7.5 describing the unsquared difference. The MSE describing the average squared difference between measured and imputed values for change in pain was 102.2, or 10.1 describing the unsquared difference. The MSE describing the average squared difference between measured and imputed values for change in pain was 97.7, or 9.9 describing the unsquared difference. The coefficients from the model using imputed values for KOOS were used to make predictions of PASS and these predictions were compared to those of the uncorrected reduced models. The corrected model had an AUC of 0.927 and Brier score of 0.10, indicating almost no decrease in predictive performance compared to the uncorrected model.

A comparison between the corrected and uncorrected coefficients is shown in table 8, along with the p-values testing whether each coefficient was different from zero, odds ratios associated with a 10-point increase in KOOS, and 95% confidence intervals for each odds ratio. The estimates

for all four KOOS predictors changed somewhat, although in all cases the odds ratio confidence intervals from the uncorrected model have some overlap with those from the corrected model. Whereas the corrected coefficients for both KOOS pain variables moved towards the null value, corrected coefficients for both KOOS function variables moved away from the null value. However, both the uncorrected and corrected coefficients were statistically significant for all four KOOS predictors and the direction of the effect remained the same. The corrected coefficient for baseline function was farthest from the null, followed by change in function. Similarly baseline function have an effect which was larger in magnitude than baseline pain, and change in function had a larger effect than change in pain, suggesting that function scores were additionally important in driving ultimate satisfaction.

In the corrected model, for each 10-point increase in baseline pain, the odds of a patient reporting overall satisfaction were 1.6 times greater, and we are 95% certain this was between 1.17 and 2.19. For each 10-point increase in baseline function, the odds of a patient reporting overall satisfaction were 0.36 times as great, and we are 95% certain this was between 0.27 and 0.49. For each 10-point increase in pain improvement the odds of a patient reporting overall satisfaction were 2.18 times greater, and we are 95% certain this was between 1.74 and 2.74. And for each 10-point increase in functional improvement, the odds of a patient reporting overall satisfaction were 2.38 times greater, and we are 95% certain this was between 1.85 and 3.06. The 10-unit increase for KOOS odds ratios is chosen here because this is an amount often used to signify a clinically relevance change in pain, function, etc.

For these KOOS predictors, it was generally observed that the width of the odds ratio confidence intervals from the corrected model remained generally consistent with those of the uncorrected model, likely because the measurement variability being considered was small. The

general pattern of consistent effect direction with only minor changes in magnitude, and comparable estimation of uncertainty, was largely observed across demographic predictors as well (table 8).

Table 8. Coefficients and Estimates from Corrected Model

	Uncorrected	**Corrected**
Brier Score	0.098	0.100
AUC	0.930	0.927
Variable	**Coefficient (p-value)** **OR (95% CI)**	**Coefficient (p-value)** **OR (95% CI)**
Intercept	-7.899 (p < 0.001)	-3.216 (p = 0.149)
Baseline Pain	0.080 (p < 0.001) *2.22 (1.67, 2.96)	0.0471(p = 0.0033) *1.60 (1.17, 2.19)
Baseline Function	-0.040 (p = 0.016) *0.67 (0.49, 0.93)	-0.101(p < 0.001) *0.36 (0.27, 0.49)
Change in Pain	0.095 (p < 0.001) *2.58 (2.02, 3.31)	0.078 (p < 0.001) *2.18 (1.74, 2.74)
Change in Function	0.047 (p = 0.002) *1.60 (1.19, 2.13)	0.087 (p < 0.001) *2.38 (1.85, 3.06)
Age 50-70 (vs. <50 group)	0.086 (p = 0.761) 1.09 (0.63, 1.89)	0.074 (p = 0.795) 1.08 (0.62, 1.88)
Age >70 (vs. <50 group)	0.619 (p = 0.256) 1.86 (0.64, 5.40)	0.735 (p = 0.172) 2.08 (0.73, 5.99)
Race non-white (vs. white)	-0.923 (p = 0.016) 0.40 (0.19, 0.84)	-1.11 (p = 0.0049) 0.33 (0.15, 0.71)
Education Medium (vs. Low)	-0.668 (p = 0.012) 0.51 (0.30, 0.87)	-0.823 (p = 0.0029) 0.44 (0.26, 0.75)
Education High (vs. Low)	-1.385 (p < 0.001) 0.25 (0.13, 0.47)	-1.473 (p < 0.001) 0.23 (0.12, 0.44)

Mental Health Score (MCS)	0.215 (p = 0.005) 1.24 (1.07, 1.44)	0.170 (p = 0.027) 1.19 (1.02, 1.38)
MCS^2	-0.0022 (p = 0.004) 0.997 (0.996, 0.999)	-0.0019 (p = 0.014) 0.998 (0.996, 0.999)

*Odds ratios associated with a 10-point increase in KOOS score.

4.0 Discussion

4.1 Model Selection

Overall, results support the importance of careful and purposeful model selection (Ranganathan, Pramesh, & Aggarwal, 2017). Although many variables were removed in the model building process, some of which were individually associated with PASS, the predictive performance of the reduced model remained high, and very similar to that of the full model.

Results support the hypothesis that KOOS pain and function sub-scores, both at baseline and at the 1-year follow-up point, would be statistically significant and important predictors of PASS. This was repeatedly shown, both using t-tests in the preliminary analysis, and evidenced by the statistically significant logistic regression coefficient estimates from the full, reduced, and MIME-corrected models. Further, the bootstrapping procedure revealed that all four KOOS predictors were consistently chosen by backwards selection, across different bootstrap-simulated samples. These results are consistent with what has been previously reported about KOOS and its broad applicability and usefulness in assessing the impact of knee injury on patient experience (Roos & Lohmander, 2003). However, this is believed to be the first study that directly assesses the association between KOOS pain and function sub-scores and patient satisfaction after APM specifically, and its predictive usefulness.

Results only partially support the hypothesis that demographic predictors most closely related to other health comorbidities, such as age and BMI, would be most important to control for in modeling PASS. BMI statistically significantly predicted PASS individually, but was rarely chosen in the bootstrapping procedure, and produced inconsistent results in the models in which it

was included. Surprisingly, age was not statistically significant individually, nor was it often chosen in backwards selection. However, it was frequently statistically significant in the models in which it was selected. Age was also thought to be a potential source of confounding, having a plausible association with satisfaction, subjects' potential to recover from injury, and susceptibility to OA (Abram et al., 2017), and therefore selected in the reduced model out of the necessity to control for its effects on PASS. Also selected were MCS and education, due to their relatively high likelihood of being chosen by backwards selection. MCS is a general mental health score, linked to a number of health comorbidities (Paredes et al., 2020), which therefore aligned with our general hypothesis. Further, any number of factors related to mental health could have plausible associations with perceived pain, function, and recovery satisfaction, and therefore MCS was useful to include and control for in the final model. Education level has also been linked to health comorbidities (Clark & Royer, 2013), but its association with knee injury recovery and satisfaction after APM specifically are less well understood. Finally, racial disparities in healthcare, and surgical outcomes and recovery specifically, have been documented (Esnaola et al., 2008) (Egede, 2006). Therefore, although it only performed moderately well, race was selected as an important variable to control for in our reduced model.

4.2 Modelling 1-Year Satisfaction

4.2.1 Multiple Imputation for Measurement Error

The multiple imputation for measurement error (MIME) approach used here was intended to use the known information about the variability of KOOS measurements to simulate potentially

true values for pain and function. Then, it was hypothesized these imputed values could be used in modelling patient satisfaction, while taking into account the additional uncertainty involved with using multiple imputations for the same subject, better replicating the variability involved with using questionnaire-derived measurements. It was our hope that the resulting estimates of the regression coefficients for KOOS variables would show less bias towards the erroneously-measured values and that their variances would give a more realistic idea of their uncertainty, although confirmation of this behavior would require a study using simulated data, which was beyond the scope of this project.

Following the generation of imputed (or predicted) values for the pain and function values for each observation, using error simulated from the known measurement error distributions for each variable, the similarity of these imputations to the measured value was calculated using MSE, averaged across all 10 imputations. The average difference between imputed and measured values ranged from 7.5 to 10.1, which seemed reasonably close considering the moderate size of the errors used. Similarly, the overall predictive performance of the corrected model was virtually unchanged, judged by both Brier score and AUC, despite considerable differences from some of the coefficients, supporting the feasibility of this MIME approach.

Results only partially supported the hypothesis that the association between KOOS pain and function would be less strong after considering the uncertainty associated with measurement error. Interestingly, while the corrected coefficients for both baseline and change in pain moved closer to the null value of zero, both baseline and change in function moved farther from the null. This would suggest that, prior to considering the uncertainty associated with measurement error, models models may overestimate the effect of pain but underestimate the effect of function. It would also suggest that imputed values of KOOS function were more closely associated with the

outcome. It is additionally curious that the reported measurement variance was greater for function than it was for pain, and the imputation MSEs for function were higher. It was also noted that there was some correlation present between KOOS variables. Since the MIME procedure involves simulating additional variability in the imputation of each KOOS predictor individually, it is possible that some of the increases observed in coefficient magnitude could result from this collinearity being decreased. Despite the changes observed, the confidence intervals (before vs. after MIME) for overlapped for each KOOS variable, and the direction of the effects remained consistent. This provides additional validation for this procedure. Concerning covariates beyond KOOS, changes were generally less considerable, and the direction of effects and approximate level of statistical significance remained consistent. Since these variables were not imputed, the subtle changes likely only resulted from the larger changes to KOOS variable estimates.

4.2.2 Interpreting MIME Results

Beyond drawing comparisons between corrected and uncorrected models, evaluation of the corrected model provides some insights into the prediction satisfaction with surgical results. First, the effect direction of KOOS coefficients makes intuitive sense: individuals who are worse off to start (have greater pain and poorer function), and those who improve the most, have the greatest probability to be satisfied. There is a logical clinical explanation for the finding that function appears to be the most important in determining satisfaction. Pain, while important, is often treatable and manageable with medication. Furthermore, pain from meniscal injury specifically is often not chronic, and occurs primarily during times of heightened use of the knee joint. The KOOS questionnaire for function, however, targets daily activities specifically, focusing on the necessary lifestyle changes and deficits that result from injury (Roos & Lohmander, 2003). The impact on

daily function is much less easily treatable, often relying on frequent physical therapy for only marginal benefit, and likely a much stronger motivator for surgery.

Part of the overall objectives of this project was identify ideal, and non-ideal candidates for surgery using primarily KOOS questionaries. It was the hope that this could aid clinicians in ruling out individuals less likely to perceive a benefit. With this aim, the estimated probabilities of 1-year APM satisfaction for the individuals in the 25th 50th, and 75th percentiles of KOOS baseline pain and function were tabulated (shown in table 9) using the corrected coefficient estimates, with other variables held at the sample mean values or the most popular category (from table 1). Probabilities were calculated using equation 12 and the coefficients in table 8. As expected, most individuals have a high (>0.5) probability of satisfaction. However, that probability is highest for individuals in the 75th percentile for pain and 25th percentile for function, and lowest for individuals in the 25th percentile for pain and 75th percentile for function.

Table 9. Estimated Probabilities of Satisfaction

Baseline Pain	**Baseline Function**	**Estimated Probability of Satisfaction**
36.2 (25th PCTL)	37.5 (25th PCTL)	0.820
36.2 (25th PCTL)	44.0 (50th PCTL)	0.702
36.2 (25th PCTL)	56.6 (75th PCTL)	0.398
47.2 (50th PCTL)	37.5 (25th PCTL)	0.884
47.2 (50th PCTL)	44.0 (50th PCTL)	0.798
47.2 (50th PCTL)	56.6 (75th PCTL)	0.526
58.4 (75th PCTL)	37.5 (25th PCTL)	0.928
58.4 (75th PCTL)	44.0 (50th PCTL)	0.870

58.4 (75th PCTL)	56.6 (75th PCTL)	0.653

The coefficients of demographic control variables also provide some insight into predicting patient satisfaction. Table 9 could easily be expanded to include percentiles/categories for additionally relevant baseline characteristics, depending on clinical needs. Although neither the 50-70 or >70 age categories were statistically significantly different from the <50 category, the >70 group showed a trend level increase in the probability of satisfaction. This suggests that older patients may be on average slightly better candidates for surgery, although with considerable person-to-person variability. It is possible that the effects of meniscal injury and OA pose a heightened burden on elderly populations which is best relieved with surgery, as opposed to non-invasive therapies, although the exact reasons for this are unclear.

The statistically significant coefficient for race indicated that individuals identifying with non-white populations have poorer satisfaction after APM that those identifying as white. This was not surprising, as people of color have not only been reported to have poorer health outcomes overall, including after surgical procedures, and suffer a disproportionate burden from many comorbidities (Esnaola et al., 2008) (Egede, 2006). Such disparities may be linked to socioeconomic status, which was not included in this model but may account for slight differences in access to quality post-surgical rehabilitation. However, more work is necessary to fully understand the effects of race in the specific case of APM and meniscal injury.

Surprisingly, the coefficient estimates for both medium (undergraduate-level college) and highly (graduate-level) educated groups were negative and statistically significant, indicating that the probability of satisfaction was highest in those who were the least educated. Furthermore, the magnitude of this difference was greatest for the highly educated group. The reasons for this are

unclear, and more research is needed to both validate this finding in the specific case of APM, and to investigate potential explanations. It is speculated that individuals who are more highly educated would be able to best take advantage of appropriate physical therapy to regain function and medication to treat pain, therefore reducing the necessity of surgery in some cases and lessening its beneficial effect.

Finally, the coefficient for Mental Component Summary (MCS), an overall mental health score, was statistically significant, both in its linear form and quadratic transformation. While the linear MCS had a positive effect on satisfaction, quadratic was negative, and its magnitude was much smaller. This indicates that, at low MCS levels, an increase of MCS has a large effect of increasing probability of satisfaction. But at higher levels of MCS, the effect of increasing MCS is lessened. This is not surprising, as better mental health is expected to be positively associated with better outcomes and surgical satisfaction (Paredes et al., 2020). However, this effect is most pronounced in severe cases of mental illness risk (low MCS), with less of a difference between individuals with moderate vs high MCS scores, which helps to explain the results of the quadratic term.

4.3 Conclusion: Future Directions, Validation, and Public Health Implications

More research is necessary to better understand how measurement variability in the KOOS questionnaire impacts estimates of its effect on patient satisfaction after APM. Specifically, a validation study or a dataset with repeated measures of KOOS variables present for some or all observations would allow for the typical regression calibration or multiple imputation approaches to be performed and compared to the methods used here. In the present case, we lack an

understanding of the relationship between measurement errors on different KOOS variables. We could not use the covariance between measurement error model parameter estimates to generate new parameter estimates for imputation as is frequently done (Keogh & White, 2014), which is a limitation of this method. Nor could we perform regression calibration methods which often rely on some true values being present (Rosner, Spiegelman, &Willett, 1990).

To provide some validation for the present method, two approaches were tested. First, the amount of measurement error was adjusted by greatly increasing the simulated variance used in step 2 of the methods (equation 16). This made the predicted/imputed values for KOOS variables, Z*, much less accurate, resulting in the estimated (corrected) coefficients for those variables moving towards the null value (zero), becoming much less statistically significant, and the predictive performance of the model decreasing. Similarly, adjusting the amount of measurement error down to zero was also tested. This had the opposite effect: imputed values were very close to measured values, corrected coefficients were highly statistically significantly different from zero, and overall predictive performance of the model was high. Details are presented in appendix B. These results matched what was expected: the more uncertainty about the accuracy of measured values, the less can be said about the association between that variable and the outcome. This finding serves to additionally validate that MIME procedure and associated code was written and executed correctly.

The second approach tested was to use the outcome, PASS (or Y), as one of the covariates when building the linear imputation model in step 1 of the methods (equation 15). This was the initial plan for this project because it is done in other MIME methods (Cole et al., 2006) (Padilla et al., 2009) (Keogh & White, 2014). However, it was found that the final coefficient estimates for KOOS variables were implausibly large and their confidence intervals often did not overlap with

those from the uncorrected model. Since Y (PASS) was used to predict Z* (imputed/potential value for KOOS), and Z* was then used to model Y, the association between Z* and Y in step 4 (equation 20) appeared much higher than it should have. Details are presented in appendix A. It is speculated that related methods may suffer from similar problems, but to a much lesser extent, since they randomly draw potential coefficients for Y from its distribution in the measurement error model step, before making imputations.

A simulation study would be required and recommended for future research in dealing with this specific case of KOOS measurement error in predicting APM satisfaction, although this was beyond the scope of the present project. In such a study, "true" values for pain and function could be simulated, which are realistically impossible to attain. Then erroneous measurements could be simulated, drawn using the known error associated with each KOOS variable. This MIME procedure could be used on the erroneous measurements, and a set of corrected estimates and their confidence intervals could be attained. Then these could be compared to estimates from a model built on the "true" values. Ideally, the corrected estimates would be within the confidence limits of the true estimates and the width of the corrected confidence intervals would approximate the uncertainty associated with measurement error.

Findings from this project have useful applications for public health. Although APM is relatively non-invasive, it still requires substantial recovery time and rehabilitation efforts. This burdens both patients and their caregivers via potential financial expenses, loss of work, temporary handicap, and overall reduced quality of life. Although effort every effort is generally taken to minimize such downsides, the decision to undergo this procedure still requires careful consideration. It is the responsibility of physicians to both identify surgical candidates who they believe are likely to benefit, and to educate patients on the risks associated with their particular

case so they can make an informed decision. Often, a standardized improvement in KOOS has been a convenient determinant of surgical success. While findings from this project reveal that such improvement is a major determinant of surgical success, it is not the only factor affecting patients' perceived feelings of satisfaction, which involves baseline KOOS measurements and several personal demographic factors. The final model from this project is intended to better explain how KOOS can be used as a tool to make such clinical decisions, despite its measurement variability.

Appendix A Using Outcome in Measurement Error Model

All steps from the MIME procedure described in the methods were repeated, apart from adding Y to the covariates of the imputation model in step 1. Table 10 compares results from this MIME model to the reduced uncorrected model.

The MSE for imputations of baseline pain was 56.3, 51.9 for baseline function, 88.3 for change in pain, and 84.5 for change in function.

Table 10. Coefficients from Model using Y for Imputations

	Uncorrected	**Corrected**
Brier Score	0.098	0.110
AUC	0.930	0.928
Variable	**Coefficient (p-value)** **OR (95% CI)**	**Coefficient (p-value)** **OR (95% CI)**
Intercept	-7.899 (p < 0.001)	-10.047 (p = 0.004)
Baseline Pain	0.080 (p < 0.001) *2.22 (1.67, 2.96)	0.190 (p < 0.001) *6.69 (3.75, 11.94)
Baseline Function	-0.040 (p = 0.016) *0.67 (0.49, 0.93)	-0.114 (p < 0.001) *0.32 (0.20, 0.50)
Change in Pain	0.095 (p < 0.001) *2.58 (2.02, 3.31)	0.205 (p < 0.001) *7.79 (4.80, 12.64)
Change in Function	0.047 (p = 0.002) *1.60 (1.19, 2.13)	0.118 (p < 0.001) *3.26 (2.18, 4.88)
Age 50-70 (vs. <50 group)	0.086 (p = 0.761) 1.09 (0.63, 1.89)	-0.183 (p = 0.680) 0.83 (0.35, 1.99)
Age >70 (vs. <50 group)	0.619 (p = 0.256) 1.86 (0.64, 5.40)	1.782 (p = 0.027) 5.94 (1.22, 28.9)

Race non-white (vs. white)	-0.923 (p = 0.016) 0.40 (0.19, 0.84)	-2.023 (p < 0.001) 0.13 (0.04, 0.43)
Education Medium (vs. Low)	-0.668 (p = 0.012) 0.51 (0.30, 0.87)	-1.331 (p = 0.002) 0.26 (0.12, 0.60)
Education High (vs. Low)	-1.385 (p < 0.001) 0.25 (0.13, 0.47)	-2.309 (p < 0.001) 0.099 (0.03, 0.29)
Mental Health Score (MCS) MCS^2	0.215 (p = 0.005) 1.24 (1.07, 1.44) -0.0022 (p = 0.004) 0.997 (0.996, 0.999)	0.203 (p = 0.076) 1.23 (0.98, 1.53) -0.0026 (p = 0.027) 0.997 (0.995, 1.00)

*Odds ratios associated with a 10-point increase in KOOS score.

Appendix B Model Validation with Error Variance Simulations

All steps from the MIME procedure described in the methods were repeated, apart from changing the measurement variance in step 2. Table 11 shows abbreviated results from a model with a measurement error variance of zero, and a model with measurement error variance of 30 points for baseline pain and function variables and 60 points for change in pain and function variables.

For the no measurement error model, the MSE for imputations of baseline pain was 56.7, 49.7 for baseline function, 90.9 for change in pain, and 72.1 for change in function. For the large measurement error model, the MSE for imputations of baseline pain was 1664, 1645.1 for baseline function, 6803.2 for change in pain, and 6494.2 for change in function.

Table 11. Coefficients when Measurement Error is Adjusted

	No Measurement Error	**Large Measurement Error**
Brier Score	0.101	0.194
AUC	0.927	0.738
Variable	**Coefficient (p-value)**	**Coefficient (p-value)**
Baseline Pain	0.055 (p = 0.001)	0.0011 (p = 0.754)
Baseline Function	-0.106 (p < 0.001)	-0.0042 (p = 0.173)
Change in Pain	0.081 (p < 0.001)	0.0024 (p = 0.110)
Change in Function	0.096 (p < 0.001)	0.0027 (p = 0.074)

Appendix C Preliminary Full and Reduced Model Diagnostics

Model diagnostics were performed to assess both the full and reduced uncorrected models, shown in figures 5 and 6 respectively. A binned residual plot was used as an alternative to a standard residual plot to assess linearity on a logit scale. A severe violation of this assumption would be expected to result in points which show trends or patterns for certain predicted values, and/or far away from the zero line. Severe violations were not observed for either model. Standardized Pearson residuals were used to detect outliers. Such residuals would be expected to cluster around the zero, and points >2 could be considered potential outliers. In both models, there are approximately 20-30 points which fall slightly outside the normal range, but this was not thought to be a major concern for this project. Pregibon leverage was used to identify high leverage points, those which have predictor values which fall far from the mean. In this case, points with Pregibon leverage greater than approximately $2(25)/828 = 0.06$ for the full model and $2(11)/828 = 0.03$ for the reduced model were considered high. There were several points which fell into this high range for both models. Finally, the Pregibon delta-beta (dbeta) influence statistic was used to measure the degree of change in coefficient estimates which would result from deleting various covariate patterns. Here, points with dbeta statistics greater than approximately 1 were considered high. No points fell into this range for either model. Finally, variance inflation factors (VIF) were calculated to assess multicollinearity. These are shown in table 12. Variables with a VIF greater than approximately 10 were considered high. There were several variables in the full model which fell into this range. This was affected by the presence of both linear and squared transformations of some variables being included together, as well as KOOS variables being included in multiple interactions terms. However, no high VIFs were found in the reduced model.

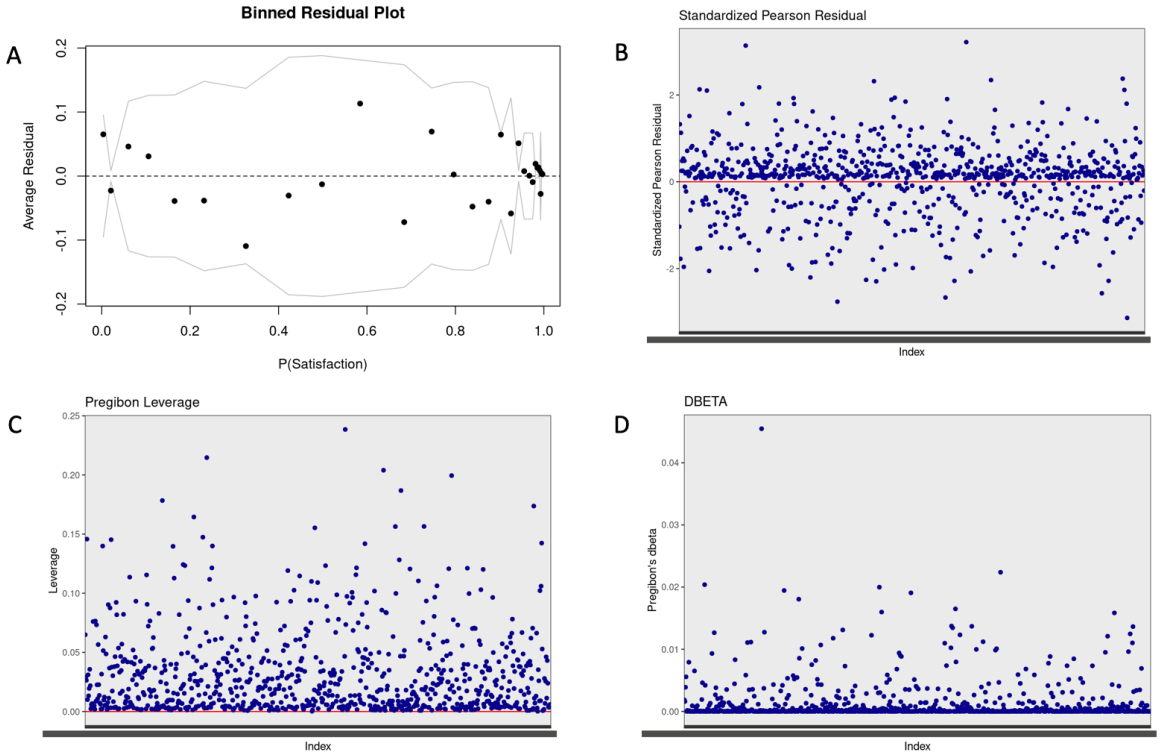

Figure 5. Diagnostic Plots for Full Model

(A) Binned Residual Plot: Data are divided into categories (or bins) based on fitted values. The average residual value is given for the average fitted value for each bin. Grey lines represent +/- 2 standard errors (SE), expected to contain approximately 95% of observations. (B) Standardized Pearson residuals. (C) Pregibon leverage. (D) Pregibon delta-beta (DBETA) influence statistics.

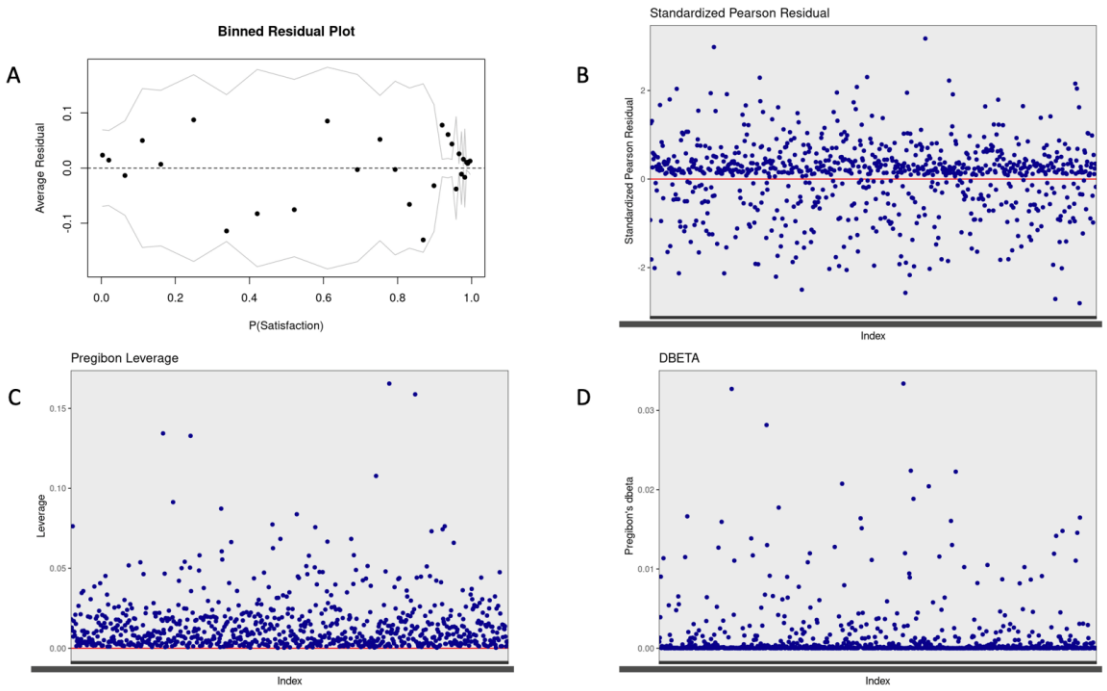

Figure 6. Diagnostic Plots for Reduced Model

(A) Binned Residual Plot: Data are divided into categories (or bins) based on fitted values. The average residual value is given for the average fitted value for each bin. Grey lines represent +/- 2 standard errors (SE), expected to contain approximately 95% of observations. (B) Standardized Pearson residuals. (C) Pregibon leverage. (D) Pregibon delta-beta (DBETA) influence statistics.

Table 12. Variance Inflation Factors from Full and Reduced Model

	Full Model	**Reduced Model**
Baseline Pain	14.62	4.57
Change in Pain	9.63	3.41
Baseline Function	18.15	5.58
Change in function	14.51	4.40
Age	1.23	1.06
Sex	1.15	
BMI BMI^2	64.78 64.59	

Race	1.41	1.08
Ethnicity	1.08	
Insurance	1.17	
Income	5.02	
Comorbidities	1.23	
Education	1.41	1.04
National Area Deprivation Index (ADI) ADI^2	37.19 25.88	
Mental Health Score (MCS) MCS^2	39.93 39.84	6.51 6.51
Smoking	1.08	
Baseline pain x change in pain	4.99	
Baseline function x change in function	21.08	
Baseline pain x baseline function	6.24	

Appendix D Receiver Operating Characteristic (ROC) Curves

ROC curves were built to evaluate models used in this project. Figure 10 shows ROC curves from the full, reduced, and corrected models, as well as the variations to the corrected model described in appendices A and B. AUCs (area under the curve) for each of these are presented elsewhere in the text and figures.

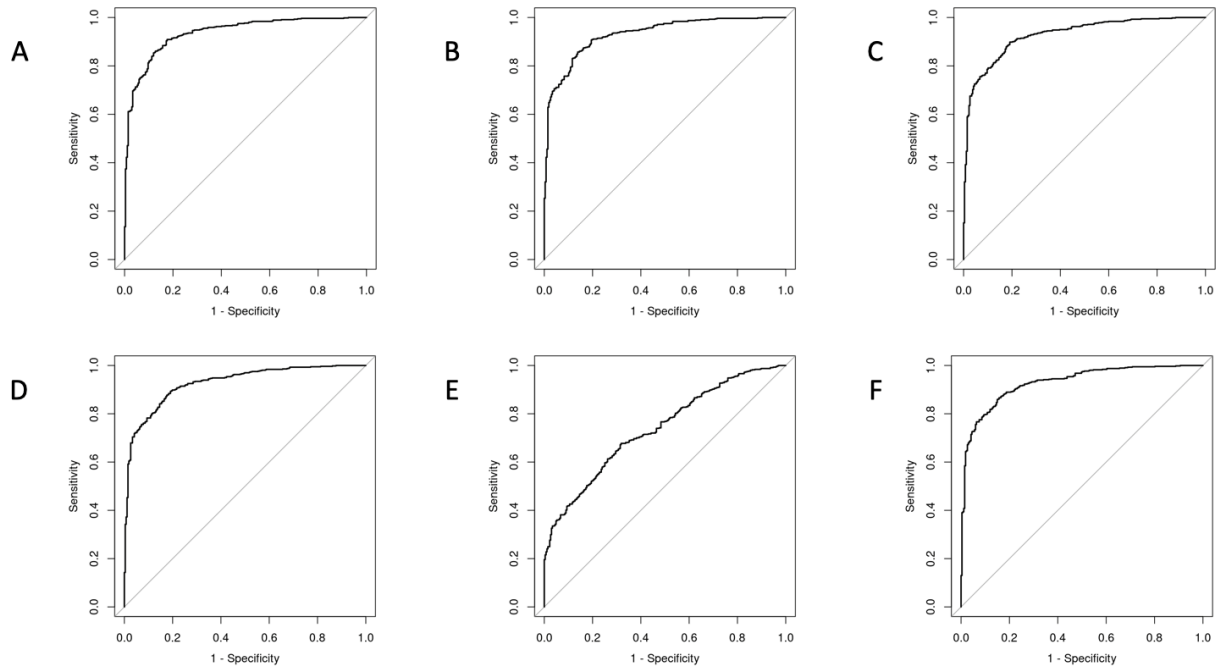

Figure 7. ROC Curves

ROC curves were built to evaluate the descrimination between PASS (outcome) categories for each model investigated. For each curve, the model sensitivity is plotted on the Y-axis and 1 minus the specificity on the X-axis. Diagonal line represents no descrimination (or random descrimination) between outcome categories. (A) Full uncorrected model. (B) Reduced uncorrected model. (C) Corrected model. (D) Corrected model with no error variance. (E) Corrected model with very large error variance. (F) Model using Y in imputation model step.

Appendix E R code for MIME Procedure

```
## 1. Regress error variable on others

```{r}
lm_baseline_pain<- summary(lm(baseline_pain ~ diff_pain + baseline_func + diff_func + age50 + age70 + race_cat + educMed + educHigh + score_pat1_vr12_mcs + MCS2, data = data))

lm_baseline_func<- summary(lm(baseline_func ~ diff_pain + baseline_pain + diff_func + age50 + age70 + race_cat + educMed + educHigh + score_pat1_vr12_mcs + MCS2, data = data))

lm_diff_pain<- summary(lm(diff_pain ~ baseline_pain + baseline_func + diff_func + age50 + age70 + race_cat + educMed + educHigh + score_pat1_vr12_mcs + MCS2, data = data))

lm_diff_func<- summary(lm(diff_func ~ baseline_pain + diff_pain + baseline_func + age50  + age70 + race_cat + educMed + educHigh + score_pat1_vr12_mcs + MCS2, data = data))
```

2-3. Predict the mis-measured variable with error variance
```{r}
for(i in 1:10) {
set.seed(i)
# Baseline Pain
e_baseline_pain <- rnorm(828,0,sqrt(4.8))
assign(paste0("baseline_pain_pred_", i), (
  coef(lm_baseline_pain)["(Intercept)","Estimate"] +
```
```

```
  data$baseline_func*coef(lm_baseline_pain)["baseline_func","Estimate"] +
  data$diff_pain*coef(lm_baseline_pain)["diff_pain","Estimate"] +
  data$diff_func*coef(lm_baseline_pain)["diff_func","Estimate"] +
  data$age50*coef(lm_baseline_pain)["age50","Estimate"] +
  data$age70*coef(lm_baseline_pain)["age70","Estimate"] +
  data$race_cat*coef(lm_baseline_pain)["race_cat","Estimate"] +
  data$educMed*coef(lm_baseline_pain)["educMed","Estimate"] +
  data$educHigh*coef(lm_baseline_pain)["educHigh","Estimate"] +
data$score_pat1_vr12_mcs*coef(lm_baseline_pain)["score_pat1_vr12_mcs","Estimate
"] +
  data$MCS2*coef(lm_baseline_pain)["MCS2","Estimate"] +
  e_baseline_pain))

### Baseline function
e_baseline_func <- rnorm(828,0,sqrt(8.3))
assign(paste0("baseline_func_pred_", i), (
  coef(lm_baseline_func)["(Intercept)","Estimate"] +
  data$baseline_pain*coef(lm_baseline_func)["baseline_pain","Estimate"] +
  data$diff_pain*coef(lm_baseline_func)["diff_pain","Estimate"] +
  data$diff_func*coef(lm_baseline_func)["diff_func","Estimate"] +
  data$age50*coef(lm_baseline_func)["age50","Estimate"] +
  data$age70*coef(lm_baseline_func)["age70","Estimate"] +
  data$race_cat*coef(lm_baseline_func)["race_cat","Estimate"] +
  data$educMed*coef(lm_baseline_func)["educMed","Estimate"] +
  data$educHigh*coef(lm_baseline_func)["educHigh","Estimate"] +
data$score_pat1_vr12_mcs*coef(lm_baseline_func)["score_pat1_vr12_mcs","Estimate
"] +
  data$MCS2*coef(lm_baseline_func)["MCS2","Estimate"] +
  e_baseline_func))

### Change in pain
e_diff_pain <- rnorm(828,0,sqrt(2*4.8))
```

```
      assign(paste0("diff_pain_pred_", i), (
        coef(lm_diff_pain)["(Intercept)","Estimate"] +
        data$baseline_pain*coef(lm_diff_pain)["baseline_pain","Estimate"] +
        data$baseline_func*coef(lm_diff_pain)["baseline_func","Estimate"] +
        data$diff_func*coef(lm_diff_pain)["diff_func","Estimate"] +
        data$age50*coef(lm_diff_pain)["age50","Estimate"] +
        data$age70*coef(lm_diff_pain)["age70","Estimate"] +
        data$race_cat*coef(lm_diff_pain)["race_cat","Estimate"] +
        data$educMed*coef(lm_diff_pain)["educMed","Estimate"] +
        data$educHigh*coef(lm_diff_pain)["educHigh","Estimate"] +
        data$score_pat1_vr12_mcs*coef(lm_diff_pain)["score_pat1_vr12_mcs","Estimate"]
+
        data$MCS2*coef(lm_diff_pain)["MCS2","Estimate"] +
        e_diff_pain))

      ### Change in pain
      e_diff_func <- rnorm(828,0,sqrt(2*8.3))
      assign(paste0("diff_func_pred_", i), (
        coef(lm_diff_func)["(Intercept)","Estimate"] +
        data$baseline_pain*coef(lm_diff_func)["baseline_pain","Estimate"] +
        data$baseline_func*coef(lm_diff_func)["baseline_func","Estimate"] +
        data$diff_pain*coef(lm_diff_func)["diff_pain","Estimate"] +
        data$age50*coef(lm_diff_func)["age50","Estimate"] +
        data$age70*coef(lm_diff_func)["age70","Estimate"] +
        data$race_cat*coef(lm_diff_func)["race_cat","Estimate"] +
        data$educMed*coef(lm_diff_func)["educMed","Estimate"] +
        data$educHigh*coef(lm_diff_func)["educHigh","Estimate"] +
        data$score_pat1_vr12_mcs*coef(lm_diff_func)["score_pat1_vr12_mcs","Estimate"]
+
        data$MCS2*coef(lm_diff_func)["MCS2","Estimate"] +
        e_diff_func))
      }
```

```

### Combine variables into data

```{r}
new.df <- data.frame(baseline_pain_pred_1, baseline_pain_pred_2,
baseline_pain_pred_3, baseline_pain_pred_4, baseline_pain_pred_5,
baseline_pain_pred_6, baseline_pain_pred_7, baseline_pain_pred_8,
baseline_pain_pred_9, baseline_pain_pred_10, baseline_func_pred_1,
baseline_func_pred_2, baseline_func_pred_3, baseline_func_pred_4,
baseline_func_pred_5, baseline_func_pred_6, baseline_func_pred_7,
baseline_func_pred_8, baseline_func_pred_9, baseline_func_pred_10, diff_pain_pred_1,
diff_pain_pred_2, diff_pain_pred_3, diff_pain_pred_4, diff_pain_pred_5,
diff_pain_pred_6, diff_pain_pred_7, diff_pain_pred_8, diff_pain_pred_9,
diff_pain_pred_10, diff_func_pred_1, diff_func_pred_2, diff_func_pred_3,
diff_func_pred_4, diff_func_pred_5, diff_func_pred_6, diff_func_pred_7,
diff_func_pred_8, diff_func_pred_9, diff_func_pred_10)
data <- cbind(data, new.df)
```

4. Regress PASS on all variables (including predicted KOOS)

```{r}
for(i in 1:10) {

assign(paste0("fit_corrected_", i),
summary(glm(as.formula(paste("as.factor(score_pat1_pass)", "~",
paste(colnames(data)[c((41+i),(51+i),(61+i),(71+i))], collapse="+"), paste(" + age50 +
age70 + race_cat + educMed + educHigh + score_pat1_vr12_mcs + MCS2"), sep = "")),
data=data, family = binomial)))

}
```
```

```
## 5. Construct corrected model

### Corrected Coefficients

```{r}
Beta_correct_intercept <- (1/10)*((coef(fit_corrected_1)["(Intercept)",
"Estimate"]) + (coef(fit_corrected_2)["(Intercept)", "Estimate"]) +
(coef(fit_corrected_3)["(Intercept)", "Estimate"]) +
(coef(fit_corrected_4)["(Intercept)", "Estimate"]) +
(coef(fit_corrected_5)["(Intercept)", "Estimate"]) +
(coef(fit_corrected_6)["(Intercept)", "Estimate"]) +
(coef(fit_corrected_7)["(Intercept)", "Estimate"]) +
(coef(fit_corrected_8)["(Intercept)", "Estimate"]) +
(coef(fit_corrected_9)["(Intercept)", "Estimate"]) +
(coef(fit_corrected_10)["(Intercept)", "Estimate"]))

Beta_correct_baseline_pain <-
(1/10)*((coef(fit_corrected_1)["baseline_pain_pred_1", "Estimate"]) +
(coef(fit_corrected_2)["baseline_pain_pred_2", "Estimate"]) +
(coef(fit_corrected_3)["baseline_pain_pred_3", "Estimate"]) +
(coef(fit_corrected_4)["baseline_pain_pred_4", "Estimate"]) +
(coef(fit_corrected_5)["baseline_pain_pred_5", "Estimate"]) +
(coef(fit_corrected_6)["baseline_pain_pred_6", "Estimate"]) +
(coef(fit_corrected_7)["baseline_pain_pred_7", "Estimate"]) +
(coef(fit_corrected_8)["baseline_pain_pred_8", "Estimate"]) +
(coef(fit_corrected_9)["baseline_pain_pred_9", "Estimate"]) +
(coef(fit_corrected_10)["baseline_pain_pred_10", "Estimate"]))

Beta_correct_baseline_func <-
(1/10)*((coef(fit_corrected_1)["baseline_func_pred_1", "Estimate"]) +
(coef(fit_corrected_2)["baseline_func_pred_2", "Estimate"]) +
(coef(fit_corrected_3)["baseline_func_pred_3", "Estimate"]) +
(coef(fit_corrected_4)["baseline_func_pred_4", "Estimate"]) +
```
```

```
(coef(fit_corrected_5)["baseline_func_pred_5", "Estimate"]) +
(coef(fit_corrected_6)["baseline_func_pred_6", "Estimate"]) +
(coef(fit_corrected_7)["baseline_func_pred_7", "Estimate"]) +
(coef(fit_corrected_8)["baseline_func_pred_8", "Estimate"]) +
(coef(fit_corrected_9)["baseline_func_pred_9", "Estimate"]) +
(coef(fit_corrected_10)["baseline_func_pred_10", "Estimate"]))

      Beta_correct_diff_pain <- (1/10)*((coef(fit_corrected_1)["diff_pain_pred_1",
"Estimate"]) + (coef(fit_corrected_2)["diff_pain_pred_2", "Estimate"]) +
(coef(fit_corrected_3)["diff_pain_pred_3", "Estimate"]) +
(coef(fit_corrected_4)["diff_pain_pred_4", "Estimate"]) +
(coef(fit_corrected_5)["diff_pain_pred_5", "Estimate"]) +
(coef(fit_corrected_6)["diff_pain_pred_6", "Estimate"]) +
(coef(fit_corrected_7)["diff_pain_pred_7", "Estimate"]) +
(coef(fit_corrected_8)["diff_pain_pred_8", "Estimate"]) +
(coef(fit_corrected_9)["diff_pain_pred_9", "Estimate"]) +
(coef(fit_corrected_10)["diff_pain_pred_10", "Estimate"]))

      Beta_correct_diff_func <- (1/10)*((coef(fit_corrected_1)["diff_func_pred_1",
"Estimate"]) + (coef(fit_corrected_2)["diff_func_pred_2", "Estimate"]) +
(coef(fit_corrected_3)["diff_func_pred_3", "Estimate"]) +
(coef(fit_corrected_4)["diff_func_pred_4", "Estimate"]) +
(coef(fit_corrected_5)["diff_func_pred_5", "Estimate"]) +
(coef(fit_corrected_6)["diff_func_pred_6", "Estimate"]) +
(coef(fit_corrected_7)["diff_func_pred_7", "Estimate"]) +
(coef(fit_corrected_8)["diff_func_pred_8", "Estimate"]) +
(coef(fit_corrected_9)["diff_func_pred_9", "Estimate"]) +
(coef(fit_corrected_10)["diff_func_pred_10", "Estimate"]))

      Beta_correct_age50 <- (1/10)*((coef(fit_corrected_1)["age50", "Estimate"]) +
(coef(fit_corrected_2)["age50", "Estimate"]) + (coef(fit_corrected_3)["age50",
"Estimate"]) + (coef(fit_corrected_4)["age50", "Estimate"]) +
```

```
(coef(fit_corrected_5)["age50", "Estimate"]) + (coef(fit_corrected_6)["age50",
"Estimate"]) + (coef(fit_corrected_7)["age50", "Estimate"]) +
(coef(fit_corrected_8)["age50", "Estimate"]) + (coef(fit_corrected_9)["age50",
"Estimate"]) + (coef(fit_corrected_10)["age50", "Estimate"]))

      Beta_correct_age70 <- (1/10)*((coef(fit_corrected_1)["age70", "Estimate"]) +
(coef(fit_corrected_2)["age70", "Estimate"]) + (coef(fit_corrected_3)["age70",
"Estimate"]) + (coef(fit_corrected_4)["age70", "Estimate"]) +
(coef(fit_corrected_5)["age70", "Estimate"]) + (coef(fit_corrected_6)["age70",
"Estimate"]) + (coef(fit_corrected_7)["age70", "Estimate"]) +
(coef(fit_corrected_8)["age70", "Estimate"]) + (coef(fit_corrected_9)["age70",
"Estimate"]) + (coef(fit_corrected_10)["age70", "Estimate"]))

      Beta_correct_race_cat <- (1/10)*((coef(fit_corrected_1)["race_cat",
"Estimate"]) + (coef(fit_corrected_2)["race_cat", "Estimate"]) +
(coef(fit_corrected_3)["race_cat", "Estimate"]) + (coef(fit_corrected_4)["race_cat",
"Estimate"]) + (coef(fit_corrected_5)["race_cat", "Estimate"]) +
(coef(fit_corrected_6)["race_cat", "Estimate"]) + (coef(fit_corrected_7)["race_cat",
"Estimate"]) + (coef(fit_corrected_8)["race_cat", "Estimate"]) +
(coef(fit_corrected_9)["race_cat", "Estimate"]) + (coef(fit_corrected_10)["race_cat",
"Estimate"]))

      Beta_correct_educMed <- (1/10)*((coef(fit_corrected_1)["educMed", "Estimate"])
+ (coef(fit_corrected_2)["educMed", "Estimate"]) + (coef(fit_corrected_3)["educMed",
"Estimate"]) + (coef(fit_corrected_4)["educMed", "Estimate"]) +
(coef(fit_corrected_5)["educMed", "Estimate"]) + (coef(fit_corrected_6)["educMed",
"Estimate"]) + (coef(fit_corrected_7)["educMed", "Estimate"]) +
(coef(fit_corrected_8)["educMed", "Estimate"]) + (coef(fit_corrected_9)["educMed",
"Estimate"]) + (coef(fit_corrected_10)["educMed", "Estimate"]))

      Beta_correct_educHigh <- (1/10)*((coef(fit_corrected_1)["educHigh",
"Estimate"]) + (coef(fit_corrected_2)["educHigh", "Estimate"]) +
```

```
(coef(fit_corrected_3)["educHigh", "Estimate"]) + (coef(fit_corrected_4)["educHigh",
"Estimate"]) + (coef(fit_corrected_5)["educHigh", "Estimate"]) +
(coef(fit_corrected_6)["educHigh", "Estimate"]) + (coef(fit_corrected_7)["educHigh",
"Estimate"]) + (coef(fit_corrected_8)["educHigh", "Estimate"]) +
(coef(fit_corrected_9)["educHigh", "Estimate"]) + (coef(fit_corrected_10)["educHigh",
"Estimate"]))

      Beta_correct_score_pat1_vr12_mcs <-
(1/10)*((coef(fit_corrected_1)["score_pat1_vr12_mcs", "Estimate"]) +
(coef(fit_corrected_2)["score_pat1_vr12_mcs", "Estimate"]) +
(coef(fit_corrected_3)["score_pat1_vr12_mcs", "Estimate"]) +
(coef(fit_corrected_4)["score_pat1_vr12_mcs", "Estimate"]) +
(coef(fit_corrected_5)["score_pat1_vr12_mcs", "Estimate"]) +
(coef(fit_corrected_6)["score_pat1_vr12_mcs", "Estimate"]) +
(coef(fit_corrected_7)["score_pat1_vr12_mcs", "Estimate"]) +
(coef(fit_corrected_8)["score_pat1_vr12_mcs", "Estimate"]) +
(coef(fit_corrected_9)["score_pat1_vr12_mcs", "Estimate"]) +
(coef(fit_corrected_10)["score_pat1_vr12_mcs", "Estimate"]))

      Beta_correct_MCS2 <- (1/10)*((coef(fit_corrected_1)["MCS2", "Estimate"]) +
(coef(fit_corrected_2)["MCS2", "Estimate"]) + (coef(fit_corrected_3)["MCS2",
"Estimate"]) + (coef(fit_corrected_4)["MCS2", "Estimate"]) +
(coef(fit_corrected_5)["MCS2", "Estimate"]) + (coef(fit_corrected_6)["MCS2",
"Estimate"]) + (coef(fit_corrected_7)["MCS2", "Estimate"]) +
(coef(fit_corrected_8)["MCS2", "Estimate"]) + (coef(fit_corrected_9)["MCS2",
"Estimate"]) + (coef(fit_corrected_10)["MCS2", "Estimate"]))

      ### Predicted Probabilities

      ```{r}
      log_odds_correct <- (Beta_correct_intercept) +
(data$baseline_pain*Beta_correct_baseline_pain) +
```
```

```
(data$baseline_func*Beta_correct_baseline_func) +
(data$diff_pain*Beta_correct_diff_pain) + (data$diff_func*Beta_correct_diff_func) +
(data$age50*Beta_correct_age50) + (data$age70*Beta_correct_age70) +
(data$race_cat*Beta_correct_race_cat) + (data$educMed*Beta_correct_educMed) +
(data$educHigh*Beta_correct_educHigh) +
(data$score_pat1_vr12_mcs*Beta_correct_score_pat1_vr12_mcs) +
(data$MCS2*Beta_correct_MCS2)

      p_PASS <- exp(log_odds_correct) / (1 + (exp(log_odds_correct)))

      data$pred_PASS <- ifelse(p_PASS<0.5, 0, 1)
      ```


      ## 6. Variance of corrected betas

      ### Var (within)
      ```{r}
      Var_WI_correct_intercept <- ((1/10)*((coef(fit_corrected_1)["(Intercept)",
"Std. Error"]) + (coef(fit_corrected_2)["(Intercept)", "Std. Error"]) +
(coef(fit_corrected_3)["(Intercept)", "Std. Error"]) +
(coef(fit_corrected_4)["(Intercept)", "Std. Error"]) +
(coef(fit_corrected_5)["(Intercept)", "Std. Error"]) +
(coef(fit_corrected_6)["(Intercept)", "Std. Error"]) +
(coef(fit_corrected_7)["(Intercept)", "Std. Error"]) +
(coef(fit_corrected_8)["(Intercept)", "Std. Error"]) +
(coef(fit_corrected_9)["(Intercept)", "Std. Error"]) +
(coef(fit_corrected_10)["(Intercept)", "Std. Error"])))^2

      Var_WI_correct_baseline_pain <-
((1/10)*((coef(fit_corrected_1)["baseline_pain_pred_1", "Std. Error"]) +
(coef(fit_corrected_2)["baseline_pain_pred_2", "Std. Error"]) +
```

```
(coef(fit_corrected_3)["baseline_pain_pred_3", "Std. Error"]) +
(coef(fit_corrected_4)["baseline_pain_pred_4", "Std. Error"]) +
(coef(fit_corrected_5)["baseline_pain_pred_5", "Std. Error"]) +
(coef(fit_corrected_6)["baseline_pain_pred_6", "Std. Error"]) +
(coef(fit_corrected_7)["baseline_pain_pred_7", "Std. Error"]) +
(coef(fit_corrected_8)["baseline_pain_pred_8", "Std. Error"]) +
(coef(fit_corrected_9)["baseline_pain_pred_9", "Std. Error"]) +
(coef(fit_corrected_10)["baseline_pain_pred_10", "Std. Error"])))^2

      Var_WI_correct_baseline_func <-
((1/10)*((coef(fit_corrected_1)["baseline_func_pred_1", "Std. Error"]) +
(coef(fit_corrected_2)["baseline_func_pred_2", "Std. Error"]) +
(coef(fit_corrected_3)["baseline_func_pred_3", "Std. Error"]) +
(coef(fit_corrected_4)["baseline_func_pred_4", "Std. Error"]) +
(coef(fit_corrected_5)["baseline_func_pred_5", "Std. Error"]) +
(coef(fit_corrected_6)["baseline_func_pred_6", "Std. Error"]) +
(coef(fit_corrected_7)["baseline_func_pred_7", "Std. Error"]) +
(coef(fit_corrected_8)["baseline_func_pred_8", "Std. Error"]) +
(coef(fit_corrected_9)["baseline_func_pred_9", "Std. Error"]) +
(coef(fit_corrected_10)["baseline_func_pred_10", "Std. Error"])))^2

      Var_WI_correct_diff_pain <- ((1/10)*((coef(fit_corrected_1)["diff_pain_pred_1",
"Std. Error"]) + (coef(fit_corrected_2)["diff_pain_pred_2", "Std. Error"]) +
(coef(fit_corrected_3)["diff_pain_pred_3", "Std. Error"]) +
(coef(fit_corrected_4)["diff_pain_pred_4", "Std. Error"]) +
(coef(fit_corrected_5)["diff_pain_pred_5", "Std. Error"]) +
(coef(fit_corrected_6)["diff_pain_pred_6", "Std. Error"]) +
(coef(fit_corrected_7)["diff_pain_pred_7", "Std. Error"]) +
(coef(fit_corrected_8)["diff_pain_pred_8", "Std. Error"]) +
(coef(fit_corrected_9)["diff_pain_pred_9", "Std. Error"]) +
(coef(fit_corrected_10)["diff_pain_pred_10", "Std. Error"])))^2
```

```
      Var_WI_correct_diff_func <- ((1/10)*((coef(fit_corrected_1)["diff_func_pred_1",
"Std. Error"]) + (coef(fit_corrected_2)["diff_func_pred_2", "Std. Error"]) +
(coef(fit_corrected_3)["diff_func_pred_3", "Std. Error"]) +
(coef(fit_corrected_4)["diff_func_pred_4", "Std. Error"]) +
(coef(fit_corrected_5)["diff_func_pred_5", "Std. Error"]) +
(coef(fit_corrected_6)["diff_func_pred_6", "Std. Error"]) +
(coef(fit_corrected_7)["diff_func_pred_7", "Std. Error"]) +
(coef(fit_corrected_8)["diff_func_pred_8", "Std. Error"]) +
(coef(fit_corrected_9)["diff_func_pred_9", "Std. Error"]) +
(coef(fit_corrected_10)["diff_func_pred_10", "Std. Error"])))^2

      Var_WI_correct_age50 <- ((1/10)*((coef(fit_corrected_1)["age50", "Std. Error"])
+ (coef(fit_corrected_2)["age50", "Std. Error"]) + (coef(fit_corrected_3)["age50",
"Std. Error"]) + (coef(fit_corrected_4)["age50", "Std. Error"]) +
(coef(fit_corrected_5)["age50", "Std. Error"]) + (coef(fit_corrected_6)["age50", "Std.
Error"]) + (coef(fit_corrected_7)["age50", "Std. Error"]) +
(coef(fit_corrected_8)["age50", "Std. Error"]) + (coef(fit_corrected_9)["age50", "Std.
Error"]) + (coef(fit_corrected_10)["age50", "Std. Error"])))^2

      Var_WI_correct_age70 <- ((1/10)*((coef(fit_corrected_1)["age70", "Std. Error"])
+ (coef(fit_corrected_2)["age70", "Std. Error"]) + (coef(fit_corrected_3)["age70",
"Std. Error"]) + (coef(fit_corrected_4)["age70", "Std. Error"]) +
(coef(fit_corrected_5)["age70", "Std. Error"]) + (coef(fit_corrected_6)["age70", "Std.
Error"]) + (coef(fit_corrected_7)["age70", "Std. Error"]) +
(coef(fit_corrected_8)["age70", "Std. Error"]) + (coef(fit_corrected_9)["age70", "Std.
Error"]) + (coef(fit_corrected_10)["age70", "Std. Error"])))^2

      Var_WI_correct_race_cat <- ((1/10)*((coef(fit_corrected_1)["race_cat", "Std.
Error"]) + (coef(fit_corrected_2)["race_cat", "Std. Error"]) +
(coef(fit_corrected_3)["race_cat", "Std. Error"]) + (coef(fit_corrected_4)["race_cat",
"Std. Error"]) + (coef(fit_corrected_5)["race_cat", "Std. Error"]) +
(coef(fit_corrected_6)["race_cat", "Std. Error"]) + (coef(fit_corrected_7)["race_cat",
```

```
"Std. Error"]) + (coef(fit_corrected_8)["race_cat", "Std. Error"]) +
(coef(fit_corrected_9)["race_cat", "Std. Error"]) +
(coef(fit_corrected_10)["race_cat", "Std. Error"])))^2

      Var_WI_correct_educMed <- ((1/10)*((coef(fit_corrected_1)["educMed", "Std.
Error"]) + (coef(fit_corrected_2)["educMed", "Std. Error"]) +
(coef(fit_corrected_3)["educMed", "Std. Error"]) + (coef(fit_corrected_4)["educMed",
"Std. Error"]) + (coef(fit_corrected_5)["educMed", "Std. Error"]) +
(coef(fit_corrected_6)["educMed", "Std. Error"]) + (coef(fit_corrected_7)["educMed",
"Std. Error"]) + (coef(fit_corrected_8)["educMed", "Std. Error"]) +
(coef(fit_corrected_9)["educMed", "Std. Error"]) + (coef(fit_corrected_10)["educMed",
"Std. Error"])))^2

      Var_WI_correct_educHigh <- ((1/10)*((coef(fit_corrected_1)["educHigh", "Std.
Error"]) + (coef(fit_corrected_2)["educHigh", "Std. Error"]) +
(coef(fit_corrected_3)["educHigh", "Std. Error"]) + (coef(fit_corrected_4)["educHigh",
"Std. Error"]) + (coef(fit_corrected_5)["educHigh", "Std. Error"]) +
(coef(fit_corrected_6)["educHigh", "Std. Error"]) + (coef(fit_corrected_7)["educHigh",
"Std. Error"]) + (coef(fit_corrected_8)["educHigh", "Std. Error"]) +
(coef(fit_corrected_9)["educHigh", "Std. Error"]) +
(coef(fit_corrected_10)["educHigh", "Std. Error"])))^2

      Var_WI_correct_score_pat1_vr12_mcs <-
((1/10)*((coef(fit_corrected_1)["score_pat1_vr12_mcs", "Std. Error"]) +
(coef(fit_corrected_2)["score_pat1_vr12_mcs", "Std. Error"]) +
(coef(fit_corrected_3)["score_pat1_vr12_mcs", "Std. Error"]) +
(coef(fit_corrected_4)["score_pat1_vr12_mcs", "Std. Error"]) +
(coef(fit_corrected_5)["score_pat1_vr12_mcs", "Std. Error"]) +
(coef(fit_corrected_6)["score_pat1_vr12_mcs", "Std. Error"]) +
(coef(fit_corrected_7)["score_pat1_vr12_mcs", "Std. Error"]) +
(coef(fit_corrected_8)["score_pat1_vr12_mcs", "Std. Error"]) +
```

```
(coef(fit_corrected_9)["score_pat1_vr12_mcs", "Std. Error"]) +
(coef(fit_corrected_10)["score_pat1_vr12_mcs", "Std. Error"])))^2

      Var_WI_correct_MCS2 <- ((1/10)*((coef(fit_corrected_1)["MCS2", "Std. Error"]) +
(coef(fit_corrected_2)["MCS2", "Std. Error"]) + (coef(fit_corrected_3)["MCS2", "Std.
Error"]) + (coef(fit_corrected_4)["MCS2", "Std. Error"]) +
(coef(fit_corrected_5)["MCS2", "Std. Error"]) + (coef(fit_corrected_6)["MCS2", "Std.
Error"]) + (coef(fit_corrected_7)["MCS2", "Std. Error"]) +
(coef(fit_corrected_8)["MCS2", "Std. Error"]) + (coef(fit_corrected_9)["MCS2", "Std.
Error"]) + (coef(fit_corrected_10)["MCS2", "Std. Error"])))^2

      ```

      ### Var (between)

      ```{r}
      Var_BTW_correct_intercept <- (1/9)*((((coef(fit_corrected_1)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2) + (((coef(fit_corrected_2)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2) + (((coef(fit_corrected_3)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2) + (((coef(fit_corrected_4)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2) + (((coef(fit_corrected_5)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2) + (((coef(fit_corrected_6)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2) + (((coef(fit_corrected_7)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2) + (((coef(fit_corrected_8)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2) + (((coef(fit_corrected_9)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2) + (((coef(fit_corrected_10)["(Intercept)",
"Estimate"]) - Beta_correct_intercept)^2))

      Var_BTW_correct_baseline_pain <-
(1/9)*((((coef(fit_corrected_1)["baseline_pain_pred_1", "Estimate"]) -
Beta_correct_baseline_pain)^2) + (((coef(fit_corrected_2)["baseline_pain_pred_2",
"Estimate"]) - Beta_correct_baseline_pain)^2) +
```

```
(((coef(fit_corrected_3)["baseline_pain_pred_3", "Estimate"]) -
Beta_correct_baseline_pain)^2) + (((coef(fit_corrected_4)["baseline_pain_pred_4",
"Estimate"]) - Beta_correct_baseline_pain)^2) +
(((coef(fit_corrected_5)["baseline_pain_pred_5", "Estimate"]) -
Beta_correct_baseline_pain)^2) + (((coef(fit_corrected_6)["baseline_pain_pred_6",
"Estimate"]) - Beta_correct_baseline_pain)^2) +
(((coef(fit_corrected_7)["baseline_pain_pred_7", "Estimate"]) -
Beta_correct_baseline_pain)^2) + (((coef(fit_corrected_8)["baseline_pain_pred_8",
"Estimate"]) - Beta_correct_baseline_pain)^2) +
(((coef(fit_corrected_9)["baseline_pain_pred_9", "Estimate"]) -
Beta_correct_baseline_pain)^2) + (((coef(fit_corrected_10)["baseline_pain_pred_10",
"Estimate"]) - Beta_correct_baseline_pain)^2))

      Var_BTW_correct_baseline_func <-
(1/9)*((((coef(fit_corrected_1)["baseline_func_pred_1", "Estimate"]) -
Beta_correct_baseline_func)^2) + (((coef(fit_corrected_2)["baseline_func_pred_2",
"Estimate"]) - Beta_correct_baseline_func)^2) +
(((coef(fit_corrected_3)["baseline_func_pred_3", "Estimate"]) -
Beta_correct_baseline_func)^2) + (((coef(fit_corrected_4)["baseline_func_pred_4",
"Estimate"]) - Beta_correct_baseline_func)^2) +
(((coef(fit_corrected_5)["baseline_func_pred_5", "Estimate"]) -
Beta_correct_baseline_func)^2) + (((coef(fit_corrected_6)["baseline_func_pred_6",
"Estimate"]) - Beta_correct_baseline_func)^2) +
(((coef(fit_corrected_7)["baseline_func_pred_7", "Estimate"]) -
Beta_correct_baseline_func)^2) + (((coef(fit_corrected_8)["baseline_func_pred_8",
"Estimate"]) - Beta_correct_baseline_func)^2) +
(((coef(fit_corrected_9)["baseline_func_pred_9", "Estimate"]) -
Beta_correct_baseline_func)^2) + (((coef(fit_corrected_10)["baseline_func_pred_10",
"Estimate"]) - Beta_correct_baseline_func)^2))

      Var_BTW_correct_diff_pain <-
(1/9)*((((coef(fit_corrected_1)["diff_pain_pred_1", "Estimate"]) -
```

```
Beta_correct_diff_pain)^2) + (((coef(fit_corrected_2)["diff_pain_pred_2", "Estimate"])
- Beta_correct_diff_pain)^2) + (((coef(fit_corrected_3)["diff_pain_pred_3",
"Estimate"]) - Beta_correct_diff_pain)^2) +
(((coef(fit_corrected_4)["diff_pain_pred_4", "Estimate"]) - Beta_correct_diff_pain)^2)
+ (((coef(fit_corrected_5)["diff_pain_pred_5", "Estimate"]) -
Beta_correct_diff_pain)^2) + (((coef(fit_corrected_6)["diff_pain_pred_6", "Estimate"])
- Beta_correct_diff_pain)^2) + (((coef(fit_corrected_7)["diff_pain_pred_7",
"Estimate"]) - Beta_correct_diff_pain)^2) +
(((coef(fit_corrected_8)["diff_pain_pred_8", "Estimate"]) - Beta_correct_diff_pain)^2)
+ (((coef(fit_corrected_9)["diff_pain_pred_9", "Estimate"]) -
Beta_correct_diff_pain)^2) + (((coef(fit_corrected_10)["diff_pain_pred_10",
"Estimate"]) - Beta_correct_diff_pain)^2))

      Var_BTW_correct_diff_func <-
(1/9)*((((coef(fit_corrected_1)["diff_func_pred_1", "Estimate"]) -
Beta_correct_diff_func)^2) + (((coef(fit_corrected_2)["diff_func_pred_2", "Estimate"])
- Beta_correct_diff_func)^2) + (((coef(fit_corrected_3)["diff_func_pred_3",
"Estimate"]) - Beta_correct_diff_func)^2) +
(((coef(fit_corrected_4)["diff_func_pred_4", "Estimate"]) - Beta_correct_diff_func)^2)
+ (((coef(fit_corrected_5)["diff_func_pred_5", "Estimate"]) -
Beta_correct_diff_func)^2) + (((coef(fit_corrected_6)["diff_func_pred_6", "Estimate"])
- Beta_correct_diff_func)^2) + (((coef(fit_corrected_7)["diff_func_pred_7",
"Estimate"]) - Beta_correct_diff_func)^2) +
(((coef(fit_corrected_8)["diff_func_pred_8", "Estimate"]) - Beta_correct_diff_func)^2)
+ (((coef(fit_corrected_9)["diff_func_pred_9", "Estimate"]) -
Beta_correct_diff_func)^2) + (((coef(fit_corrected_10)["diff_func_pred_10",
"Estimate"]) - Beta_correct_diff_func)^2))

      Var_BTW_correct_age50 <- (1/9)*((((coef(fit_corrected_1)["age50", "Estimate"])
- Beta_correct_age50)^2) + (((coef(fit_corrected_2)["age50", "Estimate"]) -
Beta_correct_age50)^2) + (((coef(fit_corrected_3)["age50", "Estimate"]) -
Beta_correct_age50)^2) + (((coef(fit_corrected_4)["age50", "Estimate"]) -
```

```
Beta_correct_age50)^2) + (((coef(fit_corrected_5)["age50", "Estimate"]) -
Beta_correct_age50)^2) + (((coef(fit_corrected_6)["age50", "Estimate"]) -
Beta_correct_age50)^2) + (((coef(fit_corrected_7)["age50", "Estimate"]) -
Beta_correct_age50)^2) + (((coef(fit_corrected_8)["age50", "Estimate"]) -
Beta_correct_age50)^2) + (((coef(fit_corrected_9)["age50", "Estimate"]) -
Beta_correct_age50)^2) + (((coef(fit_corrected_10)["age50", "Estimate"]) -
Beta_correct_age50)^2))

      Var_BTW_correct_age70 <- (1/9)*((((coef(fit_corrected_1)["age70", "Estimate"])
- Beta_correct_age70)^2) + (((coef(fit_corrected_2)["age70", "Estimate"]) -
Beta_correct_age70)^2) + (((coef(fit_corrected_3)["age70", "Estimate"]) -
Beta_correct_age70)^2) + (((coef(fit_corrected_4)["age70", "Estimate"]) -
Beta_correct_age70)^2) + (((coef(fit_corrected_5)["age70", "Estimate"]) -
Beta_correct_age70)^2) + (((coef(fit_corrected_6)["age70", "Estimate"]) -
Beta_correct_age70)^2) + (((coef(fit_corrected_7)["age70", "Estimate"]) -
Beta_correct_age70)^2) + (((coef(fit_corrected_8)["age70", "Estimate"]) -
Beta_correct_age70)^2) + (((coef(fit_corrected_9)["age70", "Estimate"]) -
Beta_correct_age70)^2) + (((coef(fit_corrected_10)["age70", "Estimate"]) -
Beta_correct_age70)^2))

      Var_BTW_correct_race_cat <- (1/9)*((((coef(fit_corrected_1)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2) + (((coef(fit_corrected_2)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2) + (((coef(fit_corrected_3)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2) + (((coef(fit_corrected_4)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2) + (((coef(fit_corrected_5)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2) + (((coef(fit_corrected_6)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2) + (((coef(fit_corrected_7)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2) + (((coef(fit_corrected_8)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2) + (((coef(fit_corrected_9)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2) + (((coef(fit_corrected_10)["race_cat",
"Estimate"]) - Beta_correct_race_cat)^2))
```

```
      Var_BTW_correct_educMed <- (1/9)*((((coef(fit_corrected_1)["educMed",
"Estimate"]) - Beta_correct_educMed)^2) + (((coef(fit_corrected_2)["educMed",
"Estimate"]) - Beta_correct_educMed)^2) + (((coef(fit_corrected_3)["educMed",
"Estimate"]) - Beta_correct_educMed)^2) + (((coef(fit_corrected_4)["educMed",
"Estimate"]) - Beta_correct_educMed)^2) + (((coef(fit_corrected_5)["educMed",
"Estimate"]) - Beta_correct_educMed)^2) + (((coef(fit_corrected_6)["educMed",
"Estimate"]) - Beta_correct_educMed)^2) + (((coef(fit_corrected_7)["educMed",
"Estimate"]) - Beta_correct_educMed)^2) + (((coef(fit_corrected_8)["educMed",
"Estimate"]) - Beta_correct_educMed)^2) + (((coef(fit_corrected_9)["educMed",
"Estimate"]) - Beta_correct_educMed)^2) + (((coef(fit_corrected_10)["educMed",
"Estimate"]) - Beta_correct_educMed)^2))

      Var_BTW_correct_educHigh <- (1/9)*((((coef(fit_corrected_1)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2) + (((coef(fit_corrected_2)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2) + (((coef(fit_corrected_3)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2) + (((coef(fit_corrected_4)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2) + (((coef(fit_corrected_5)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2) + (((coef(fit_corrected_6)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2) + (((coef(fit_corrected_7)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2) + (((coef(fit_corrected_8)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2) + (((coef(fit_corrected_9)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2) + (((coef(fit_corrected_10)["educHigh",
"Estimate"]) - Beta_correct_educHigh)^2))

      Var_BTW_correct_score_pat1_vr12_mcs <-
(1/9)*((((coef(fit_corrected_1)["score_pat1_vr12_mcs", "Estimate"]) -
Beta_correct_score_pat1_vr12_mcs)^2) + (((coef(fit_corrected_2)["score_pat1_vr12_mcs",
"Estimate"]) - Beta_correct_score_pat1_vr12_mcs)^2) +
(((coef(fit_corrected_3)["score_pat1_vr12_mcs", "Estimate"]) -
Beta_correct_score_pat1_vr12_mcs)^2) + (((coef(fit_corrected_4)["score_pat1_vr12_mcs",
"Estimate"]) - Beta_correct_score_pat1_vr12_mcs)^2) +
(((coef(fit_corrected_5)["score_pat1_vr12_mcs", "Estimate"]) -
```

```
Beta_correct_score_pat1_vr12_mcs)^2) + (((coef(fit_corrected_6)["score_pat1_vr12_mcs",
"Estimate"]) - Beta_correct_score_pat1_vr12_mcs)^2) +
(((coef(fit_corrected_7)["score_pat1_vr12_mcs", "Estimate"]) -
Beta_correct_score_pat1_vr12_mcs)^2) + (((coef(fit_corrected_8)["score_pat1_vr12_mcs",
"Estimate"]) - Beta_correct_score_pat1_vr12_mcs)^2) +
(((coef(fit_corrected_9)["score_pat1_vr12_mcs", "Estimate"]) -
Beta_correct_score_pat1_vr12_mcs)^2) +
(((coef(fit_corrected_10)["score_pat1_vr12_mcs", "Estimate"]) -
Beta_correct_score_pat1_vr12_mcs)^2))

      Var_BTW_correct_MCS2 <- (1/9)*((((coef(fit_corrected_1)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2) + (((coef(fit_corrected_2)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2) + (((coef(fit_corrected_3)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2) + (((coef(fit_corrected_4)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2) + (((coef(fit_corrected_5)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2) + (((coef(fit_corrected_6)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2) + (((coef(fit_corrected_7)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2) + (((coef(fit_corrected_8)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2) + (((coef(fit_corrected_9)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2) + (((coef(fit_corrected_10)["MCS2", "Estimate"]) -
Beta_correct_MCS2)^2))
      ```


      ### Var (Total)


      ```{r}
      Var_T_correct_intercept <- (Var_WI_correct_intercept + (1 +
(1/10))*Var_BTW_correct_intercept)

      Var_T_correct_baseline_pain <- (Var_WI_correct_baseline_pain + (1 +
(1/10))*Var_BTW_correct_baseline_pain)
```

```
      Var_T_correct_baseline_func <- (Var_WI_correct_baseline_func + (1 +
(1/10))*Var_BTW_correct_baseline_func)

      Var_T_correct_diff_pain <- (Var_WI_correct_diff_pain + (1 +
(1/10))*Var_BTW_correct_diff_pain)

      Var_T_correct_diff_func <- (Var_WI_correct_diff_func + (1 +
(1/10))*Var_BTW_correct_diff_func)

      Var_T_correct_age50 <- (Var_WI_correct_age50 + (1 +
(1/10))*Var_BTW_correct_age50)

      Var_T_correct_age70 <- (Var_WI_correct_age70 + (1 +
(1/10))*Var_BTW_correct_age70)

      Var_T_correct_race_cat <- (Var_WI_correct_race_cat + (1 +
(1/10))*Var_BTW_correct_race_cat)

      Var_T_correct_educMed <- (Var_WI_correct_educMed + (1 +
(1/10))*Var_BTW_correct_educMed)

      Var_T_correct_educHigh <- (Var_WI_correct_educHigh + (1 +
(1/10))*Var_BTW_correct_educHigh)

      Var_T_correct_score_pat1_vr12_mcs <- (Var_WI_correct_score_pat1_vr12_mcs + (1 +
(1/10))*Var_BTW_correct_score_pat1_vr12_mcs)

      Var_T_correct_MCS2 <- (Var_WI_correct_MCS2 + (1 + (1/10))*Var_BTW_correct_MCS2)
```

7. Test

Statistic

```{r}
B0 <- 0

T_intercept <- (abs(B0 - Beta_correct_intercept)) /
(sqrt(Var_T_correct_intercept))
T_baseline_pain <- (abs(B0 - Beta_correct_baseline_pain)) /
(sqrt(Var_T_correct_baseline_pain))
T_baseline_func <- (abs(B0 - Beta_correct_baseline_func)) /
(sqrt(Var_T_correct_baseline_func))
T_diff_pain <- (abs(B0 - Beta_correct_diff_pain)) /
(sqrt(Var_T_correct_diff_pain))
T_diff_func <- (abs(B0 - Beta_correct_diff_func)) /
(sqrt(Var_T_correct_diff_func))
T_age50 <- (abs(B0 - Beta_correct_age50)) / (sqrt(Var_T_correct_age50))
T_age70 <- (abs(B0 - Beta_correct_age70)) / (sqrt(Var_T_correct_age70))
T_race_cat <- (abs(B0 - Beta_correct_race_cat)) /
(sqrt(Var_T_correct_race_cat))
T_educMed <- (abs(B0 - Beta_correct_educMed)) / (sqrt(Var_T_correct_educMed))
T_educHigh <- (abs(B0 - Beta_correct_educHigh)) /
(sqrt(Var_T_correct_educHigh))
T_score_pat1_vr12_mcs <- (abs(B0 - Beta_correct_score_pat1_vr12_mcs)) /
(sqrt(Var_T_correct_score_pat1_vr12_mcs))
T_MCS2 <- (abs(B0 - Beta_correct_MCS2)) / (sqrt(Var_T_correct_MCS2))
```

DF

```{r}
df_intercept <- (9*((1+(Var_WI_correct_intercept/((1-
(1/10))*Var_BTW_correct_intercept)))^2))
df_baseline_pain <- (9*((1+(Var_WI_correct_baseline_pain/((1-
(1/10))*Var_BTW_correct_baseline_pain)))^2))
df_baseline_func <- (9*((1+(Var_WI_correct_baseline_func/((1-
(1/10))*Var_BTW_correct_baseline_func)))^2))
df_diff_pain <- (9*((1+(Var_WI_correct_diff_pain/((1-
(1/10))*Var_BTW_correct_diff_pain)))^2))
df_diff_func <- (9*((1+(Var_WI_correct_diff_func/((1-
(1/10))*Var_BTW_correct_diff_func)))^2))
df_age50 <- (9*((1+(Var_WI_correct_age50/((1-
(1/10))*Var_BTW_correct_age50)))^2))
df_age70 <- (9*((1+(Var_WI_correct_age70/((1-
(1/10))*Var_BTW_correct_age70)))^2))
df_race_cat <- (9*((1+(Var_WI_correct_race_cat/((1-
(1/10))*Var_BTW_correct_race_cat)))^2))
df_educMed <- (9*((1+(Var_WI_correct_educMed/((1-
(1/10))*Var_BTW_correct_educMed)))^2))
df_educHigh <- (9*((1+(Var_WI_correct_educHigh/((1-
(1/10))*Var_BTW_correct_educHigh)))^2))
df_score_pat1_vr12_mcs <- (9*((1+(Var_WI_correct_score_pat1_vr12_mcs/((1-
(1/10))*Var_BTW_correct_score_pat1_vr12_mcs)))^2))
df_MCS2 <- (9*((1+(Var_WI_correct_MCS2/((1-(1/10))*Var_BTW_correct_MCS2)))^2))
```

P-values

```{r}
p_intercept <- 2*pt(q=T_intercept, df=df_intercept, lower.tail=F)
p_baseline_pain <- 2*pt(q=T_baseline_pain, df=df_baseline_pain, lower.tail=F)
p_baseline_func <- 2*pt(q=T_baseline_func, df=df_baseline_func, lower.tail=F)
```

```
p_diff_pain <- 2*pt(q=T_diff_pain, df=df_diff_pain, lower.tail=F)
p_diff_func <- 2*pt(q=T_diff_func, df=df_diff_func, lower.tail=F)
p_age50 <- 2*pt(q=T_age50, df=df_age50, lower.tail=F)
p_age70 <- 2*pt(q=T_age70, df=df_age70, lower.tail=F)
p_race_cat <- 2*pt(q=T_race_cat, df=df_race_cat, lower.tail=F)
p_educMed <- 2*pt(q=T_educMed, df=df_educMed, lower.tail=F)
p_educHigh <- 2*pt(q=T_educHigh, df=df_educHigh, lower.tail=F)
p_score_pat1_vr12_mcs <- 2*pt(q=T_score_pat1_vr12_mcs,
df=df_score_pat1_vr12_mcs, lower.tail=F)
p_MCS2 <- 2*pt(q=T_MCS2, df=df_MCS2, lower.tail=F)
```

Corrected Model Performance

Brier Score

```{r}
Brier(p_PASS, data$score_pat1_pass, 0, 1)
```

ROC

```{r}
par(pty="s")
roc(data$score_pat1_pass, p_PASS, plot=T, legacy.axes=T)
```

Bibliography

Abram, S.G., Judge, A., Beard, D.J., Price, A.J. (2018). Adverse Outcomes after Arthroscopic Partial Meniscectomy: A Study of 700,000 Procedures in the National Hospital Episode Statistics Database for England. *The Lancet, 392,* 2194–2202. https://doi.org/10.1016/s0140-6736(18)31771-9

Armstrong, B. Exposure measurement error: consequences and design issues. In Nieuwenhuijsen MJ ed. Exposure assessment in occupational and environmental epidemiology. Oxford University Press, 2003.

Baldwin, J.N., McKay, M.J., Simic, M., Hiller, C.E., Moloney, N., Nightingale, E.J., Burns, J. (2017). Self-Reported Knee Pain and Disability among Healthy Individuals: Reference Data and Factors Associated with the Knee Injury and Osteoarthritis Outcome Score (Koos) and Koos-Child. *Osteoarthritis and Cartilage, 25 (8),* 1282–90. https://doi.org/10.1016/j.joca.2017.03.007

Bashir, S.A., Duffy, S.W. (1997). The Correction of Risk Estimates for Measurement Error. *Annals of Epidemiology, 7*, 154-164.

Bursac, Z., Gauss, C.H., Williams, D.K., Hosmer, D.W., (2008). Purposeful selection of variables in logistic regression. *Source Code Biol Med, 3(17).* https://doi.org/10.1186/1751-0473-3-17

Clark, D., Royer, H. (2013). The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review, 103(6),* 2087-2120.

Cole, S.R, Chu, H., Greenland, S. (2006). Multiple-Imputation for Measurement-Error Correction. *International Journal of Epidemiology, 35(4),* 1074–81. https://doi.org/10.1093/ije/dyl097

Collins, N.J., Misra, D., Felson, D.T., Crossley, K.M., Roos, E.M. (2011). Measures of Knee Function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (Koos), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (Koos-PS), Knee Ou. *Arthritis Care & Research 63(11).* https://doi.org/10.1002/acr.20632

Egede, L.E. (2006). Race, ethnicity, culture, and disparities in health care. *Journal of general internal medicine, 21*(6), 667–669. https://doi.org/10.1111/j.1525-1497.2006.0512.x

Englund, M., Roos, E.M., Lohmander, L.S. (2003). Impact of Type of Meniscal Tear on Radiographic and Symptomatic Knee Osteoarthritis: A Sixteen-Year Followup of Meniscectomy with Matched Controls. *Arthritis & Rheumatism, 48(8),* 2178–87. https://doi.org/10.1002/art.11088

Esnaola, N.F., Hall, B.L., Hosokawa, P.W., Ayanian, J.Z., Henderson, W.G., Khuri, S.F., Zinner, M.J., Rogers, S.O. (2008). Race and Surgical Outcomes. *Annals of Surgery, 248(4),* 647-655. doi: 10.1097/SLA.0b013e31818a159a

Freedman, L.S., Fainberg, V., Kipnis, V., Midthune, D., Carroll, R.J. (2004). A New Method for Dealing with Measurement Error in Explanatory Variables of Regression Models. *Biometrics, 60(1),* 172–81. https://doi.org/10.1111/j.0006-341x.2004.00164.x

Guolo, A. (2008). Robust Techniques for Measurement Error Correction: A Review. *Statistical Methods in Medical Research*, 17 (6), 555-580. https://doi-org.pitt.idm.oclc.org/10.1177/0962280207081318

James, G., Witten, D., Hastie, T., Tibshirani, R. (2017). *An introduction to statistical learning: With applications in R*. Springer New York Heidelberg Dordrecht London.

Katz, J.N., Brophy, R.H., Chaisson, C.E., de Chaves, L., Cole, B.J., Dahm, D.L., Donnell-Fink, L.A. (2013). Surgery versus Physical Therapy for a Meniscal Tear and Osteoarthritis. *New England Journal of Medicine*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3690119/

Keogh, R. H., & White, I. R. (2014). A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Statistics in medicine*, *33*(12), 2137–2155. https://doi.org/10.1002/sim.6095

Kromhout, D. (2009). Correcting for Multivariate Measurement Error by Regression Calibration in Meta-Analyses of Epidemiological Studies. *Statistics in Medicine 28(7),* 1067–92. https://doi.org/10.1002/sim.3530

Padilla, M. A., Divers, J., Vaughan, L. K., Allison, D. B., & Tiwari, H. K. (2009). Multiple imputation to correct for measurement error in admixture estimates in genetic structured association testing. *Human heredity, 68(1),* 65–72. https://doi.org/10.1159/000210450

Paredes, A. Z., Hyer, J. M., Diaz, A., Tsilimigras, D. I., & Pawlik, T. M. (2020). The Impact of Mental Illness on Postoperative Outcomes Among Medicare Beneficiaries: A Missed Opportunity to Help Surgical Patients?. *Annals of surgery*, *272*(3), 419–425. https://doi.org/10.1097/SLA.0000000000004118

Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Logistic regression. *Perspectives in clinical research*, *8*(3), 148–151. http://europepmc.org/abstract/MED/28828311

Roos, E.M., Lohmander, L.S. (2003). Knee Injury and Osteoarthritis Outcome Score (KOOS): from Joint Injury to Osteoarthritis. *Health Qual Life Outcomes, 1(64).*

Rosner, B., Willett, W.C., Spiegelman, D. (1989). Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Systemic Within-Person Measurement Error. *Statistics in Medicine, 8(9),* 1051-1069. https://doi.org/10.1002/sim.4780080905

Rosner, B., Spiegelman, D., Willett, W.C. (1990). Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Measurement Error: The Case of Multiple Covariates Measured with Error. *American Journal of Epidemiology, 132(4),* 734–45. https://doi.org/10.1093/oxfordjournals.aje.a115715

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.

Thürigen D., Spiegelman, D., Blettner, M., Heuer, C., Brenner, H. (2000). Measurement error correction using validation data: a review of methods and their applicability in case-control studies. *Statistical Methods in Medical Research, 9,* 447-74.

Samson, D.J., Grant, M.D., Ratko, T.A., Bonnell, C.J., Ziegler, K.M., Aronson, N. Treatment of Primary and Secondary Osteoarthritis of the Knee, Evidence Report/Technology Assessment No. 157 (Prepared by Blue Cross and Blue Shield Association Technology Evaluation Center Evidence- based Practice Center under Contract No. 290-02-0026), AHRQ Publication No. 07-E012, 2007.

Wong, M.Y., Day, N.E., Bashir, S.A., Duffy, S.W. (1999). Measurement Error in Epidemiology: The Design of Validation Studies I: Univariate Situation. *Statistics in Medicine, 18,* 2815-2829.

Wright, A. A., Hensley, C. P., Gilbertson, J., Leland, J. M., Jackson, S. (2015). Defining patient acceptable symptom state thresholds for commonly used patient reported outcomes measures in general orthopedic practice. *Manual therapy*, *20*(6), 814–819. https://doi.org/10.1016/j.math.2015.03.011

Wu, Y.C., Lee, W.C. (2014). Alternative performance measures for prediction models. *PLoS One, 9(3),* e91249. doi: 10.1371/journal.pone.0091249. PMID: 24608868; PMCID: PMC3946724

Yuan, Y. C. (2010). *Multiple imputation for missing data: Concepts and new ...* Retrieved January 28, 2022, from https://support.sas.com/rnd/app/stat/papers/multipleimputation.pdf