Spring 2023

The Application of Statistical Modeling to Identify Genetic Associations with Mild Traumatic Brain Injury Outcomes

Caroline Schott
Missouri University of Science and Technology

Department:

THE APPLICATION OF STATISTICAL MODELING TO IDENTIFY GENETIC ASSOCIATIONS WITH MILD TRAUMATIC BRAIN INJURY OUTCOMES

by

CAROLINE ELIZABETH CLAIRE SCHOTT

A THESIS

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

in

APPLIED MATHEMATICS

2023

Approved by:

Dr. Gayla R. Olbricht, Advisor
Dr. Tayo Obafemi-Ajayi
Dr. V.A. Samaranayake

PUBLICATION THESIS OPTION

This thesis contains the following article, formatted in the style used by the Missouri University of Science and Technology:

PAPER I. found on page 23-51, is intended for submission to the Frontiers in Bioinformatics journal.

ABSTRACT

Traumatic brain injury (TBI) is a growing health concern, with millions of TBI diagnoses in the United States each year. The vast majority of TBI diagnoses are mild traumatic brain injuries (mTBI), which can be challenging to manage due to variation in symptoms and outcomes. Most individuals with mTBI successfully recover quickly, but a small subset has a delayed recovery. Although the factors that contribute to this variation in recovery are not clearly understood, it is possible that genetic differences may play a role. Very few studies have investigated the association between single nucleotide polymorphisms (SNPs) with mTBI outcomes and this is an emerging area of research. In this study, we utilize data collected in the Transforming Research and Clinical Knowledge in Traumatic Brain Injury Pilot (TRACK-TBI Pilot) study to test the association between 10 different SNPs and 7 TBI outcomes measured at six- and twelve-months post injury. Linear mixed models are utilized to investigate the association between genotypes and mTBI outcome measurements over time. Previous studies have primarily focused on a single time point at six months for one or two SNPs. This study seeks to expand the existing literature by using the TRACK-TBI Pilot data to evaluate multiple SNPs and multiple outcome assessments to discover their connections over time. The findings in this study demonstrate the potential benefits of using linear mixed models to identify relationships between genotypes and TBI outcomes over time.

ACKNOWLEDGMENTS

I would like to thank Dr. Gayla Olbricht, my advisor, for her guidance in research, and my professional career. She brought passion to the project and deeply invested herself in my success in data science and statistics. My master's degree would not have been possible without her guidance and support throughout my undergraduate and graduate programs at Missouri S&T. I would like to thank Dr. Tayo Obafemi-Ajayi, Dr. Dan Hier, Hung Nguyen, and Dr. Robert Paige for bringing their expertise to the project and advising me on the data set. I would also like to thank Dr. V.A. Samaranayake for serving on my defense committee.

I am grateful for my family and friends, especially my parents Michael and Kimberly Schott and husband Nicholas Tompkins, for the support and encouragement throughout my studies.

TABLE OF CONTENTS

PAPER

LIST OF ILLUSTRATIONS

PAPER I Page

LIST OF TABLES

I. INTRODUCTION

1.1 MOTIVATION

Traumatic brain injuries (TBI) have resulted in over 56 thousand deaths in the United States and 50 million TBI diagnoses internationally in a single year, making it a leading cause of death and disability (Alan D. Kaplan, 2022). While TBI deaths have lowered dramatically since the early 1970s, TBIs are still a major global health issue and it is important to research the disabilities caused by them (Nelson LD, 2017). TBIs have a wide range of severity, with most injuries (70-90%) classified as mild (mTBI), defined by the Glasgow Coma Scale (GCS) score of 13 to 15 (Winkler EA, 2016). Patients with mTBIs, typically have full recovery within weeks to months, but effects can last up to a year after the initial accident (Cnossen MC & Investigators, 2017). Post-concussive syndrome, a clinical term, is used to explain cognitive, physical, and psychological functions that have been affected by a TBI (Cnossen MC & Investigators, 2017) (Nelson LD, 2017).

Within mTBI patients, there is a wide range of variation in outcomes and it is not well understood which factors may contribute to this variation (Lingsma HF & Investigators, 2015). Gaining a better understanding of this variation can aid in determining which patients are at higher risk for poor prognosis and could provide improved treatment plans. One factor that could be impacting this variation is genetic differences between individuals. Recent studies have begun to explore connections between single nucleotide polymorphisms (SNPs) and mTBI outcomes. A SNP is a common type of genetic variation in patients that represents a difference in a single DNA

building block (nucleotide) (National Library of Medicine, n.d.). Some prior studies have performed genetic association analysis to investigate the connection between a SNP and certain outcomes at the six-month post-injury time point (Winkler EA, 2016) (John K. Yue, 2015) (Yue JK W. E.-A., 2017). The results from these studies have shown that SNPs within certain genes (e.g., APOE and COMT) have a significant association with specific mTBI 6 month outcomes. Although this indicates that genetic association analysis is a promising avenue for better understanding mTBI outcome heterogeneity, the previous studies have been limited in analyzing one or two SNPs at a single time point on a small subset of mTBI outcomes.

In this study, a more comprehensive genetic association analysis is performed by utilizing data available from the TRACK-TBI Pilot study. The work aims to test the association between 10 different SNPs and 7 outcomes assessments measured for at least two post-injury time points in mTBI patients. A linear mixed model is utilized to determine if changes in post-injury outcomes over time differ significantly between genotypes. This is a unique aspect of this study, as prior works do not assess this type of relationship over time. If no such significant association over time is found for the SNP being tested, then an overall association test between the genotype and outcome will be conducted. Secondary analyses will also be conducted to test for overall changes in the outcomes over time and for associations between outcomes and potential covariates. Through this comprehensive investigation of multiple SNPS and mTBI outcomes together in one study, the chance of false associations can be better controlled and missing data can be handled in an effective way through the linear mixed modeling.

The thesis contains a paper intended for publication in a peer reviewed journal. The paper contains a description of the data and the genetic association analysis results, along with details of the overall statistical modeling framework that includes steps for data curation, demographic analysis, assumption checking, the linear mixed modeling and multiple testing corrections. However, additional details about the data set and statistical modeling methods that are not provided in the paper are given in this introduction to the thesis. Section 1.2 of the introduction provides a detailed description of the data set used in the research and Section 1.3 provides details about all of the statistical modeling methods that are utilized in the work. The overall goal of this thesis is to illustrate how linear mixed models can be utilized in genetic association studies connecting SNP data to longitudinal TBI outcomes.

1.2. TRACK-TBI PILOT DATA SET

The data used in this study is taken from the Transforming Research and Clinical Knowledge in Traumatic Brain Injury Pilot (TRACK-TBI Pilot) study (Yue JK & Investigators, 2013). The TRACK TBI Pilot data set is available through the Federal Interagency Traumatic Brain Injury Research (FITBIR) website (https://fitbir.nih.gov/) and this database provided the initial inspiration for this work. The data utilized for this study was obtained from the Principal Investigator (PI), Dr. Geoffrey Manley, which included data available on FITBIR as well as additional genotype data and 12-month neurocognitive assessments that were not available through FITBIR (Geoffrey Manley, n.d.).

TRACK-TBI Pilot is an observational study conducted at three level I trauma centers in the United States – San Francisco General Hospital, University of Pittsburgh Medical Center, and University Medical Center Brackenridge (Kreitzer NP, 2019). The pilot study inclusion criteria required adults to present at a level I trauma center with external force trauma to the head and complete a computed tomography (CT) scan in the first 24 hours after injury (Yue JK W. E.-A., 2017). Adults (age 18 or over) could not participate if they were pregnant, had a life-threatening disease, incarcerated, on psychiatric hold, or did not speak English (due to the assessments being given in English). Demographic, clinical, biomarker, and neuroimaging data were collected on patients in the acute injury phase. Outcome assessments were collected at multiple post-injury time points (3, 6, and 12 months) to monitor recovery, with the 6- and 12-month neurocognitive assessments being provided for most outcomes in the data set obtained from the TRACK-TBI Pilot study PI, Dr. Geoffrey Manley (Geoffrey Manley, n.d.). In addition, genotype data on a select set of candidate single nucleotide polymorphisms (SNPs) was obtained on a subset of patients, which was made available by Dr. Geoffrey Manley. The goal of this research is to test for associations between SNPs and post-injury outcomes collected at prolonged post-injury time points in a population of adult patients with mild traumatic brain injuries.

The initial data set had 650 participants in the study, but 51 of these patients were from a rehabilitation center instead of a level I trauma centers and these were excluded in this work (Yue JK & Investigators, 2013). Thus, a total of 599 patients were enrolled at the three sites and were considered as potential subjects for the current study. After limiting the inclusion criteria to mild TBIs (i.e., having a GCS score of 13 to 15 upon

arrival at the emergency department) and reducing the age range from 18 to 80, the total possible number of patients with available SNP and outcome assessment data is N=330.

Data were available on 11 SNPs and 6 different types of outcome assessments with 6- and 12-month measurements. One SNP (rs3219119) was removed due to not meeting Hardy-Weinberg Equilibrium (described in Section 1.3.3) and one type of assessment (Craig Handicap Assessment and Reporting Technique) was removed since the vast majority of patients attained the highest score (100) on all subscales at both time points, resulting in a lack of variation. Thus, a total of 10 SNPs and 5 different assessments (with 7 total outcomes since one assessment had 3 subscales) were considered for further analysis. Additionally, a set of 7 demographic and clinical variables were also selected to characterize the study population and to consider as potential covariates in the statistical modeling. The demographic variables, genotype/SNP data, and outcome assessments are briefly described in the following sections. Additional details can also be found in Paper I. Note that the sample size for each variable may differ from N=330 due to missing data on some variables.

1.2.1. Demographic and Clinical Data. The demographic variables include age, sex, race, and education in years. The race data was combined into 3 groups, consisting of 78.5% Caucasian, 9.5% African American/African, and 12% other races. The average age was 42, average number of education years was 14, and the study sample was comprised of 72% males and 28% females. In addition to the demographic variables, three clinical variables were also included in the analysis: GCS score upon arrival to the emergency department, Injury Severity Score (ISS) score (≤ 15 or >15), and whether the patient had an abnormal CT scan. Most patients (73.6%) had a GCS score of 15, 67.8% had an ISS

less than or equal to 15, and 57.1% did not have an abnormal CT scan. The GCS score helps in determining the level of consciousness a patient has post-injury and the ISS score assesses the severity of the injury, with a score over 15 indicating major trauma (Eric A Toschlog, 2003). Demographic summaries for individual SNPs can be found in the Appendix.

1.2.2. Genotype Data. For each patient included in this study, blood samples were collected within 24 hours of injury and genotyping was performed on a set of candidate SNPs (Yue JK & Investigators, 2013). It is first helpful to review some genetic definitions to clarify important concepts related to SNPs and genetic association studies. Alleles consist of alternate versions of DNA sequence at a given genetic locus, which could be a SNP that differs in one nucleotide base or a variation of a longer segment of nucleotide bases. Many loci are biallelic, in that there are two different forms of the allele (e.g., A and a), although some loci have more than two possible alleles. Alleles are inherited, one from each parent, with the genotype representing the set of two specific alleles that an individual has at the given genomic location (National Human Genome Research Institute, 2023). A homozygous individual has two alleles that are the same (e.g., A/A or a/a) and a heterozygous individual has two alleles that are different (e.g., A/a) (National Human Genome Research Institute, 2023). For biallelic SNPs, the allele that is the most frequent in the population is called the major allele, while the allele that is that is less common is called the minor or variant allele. A genotype is called a carrier of a minor allele if it contains at least one of the minor alleles (i.e., homozygous for the minor allele or heterozygous).

To perform the genetic association analysis, mTBI outcomes will be compared between genotypes to determine if there are any significant differences. In this study, it is of interest to compare presence of the minor allele (i.e., minor allele carriers) versus absence of the minor allele (i.e., non-carriers). The SNPs investigated in this work are described briefly below with an explanation of their importance based on previous studies and a description of the comparison that will be made. Note that when referring different SNP alleles, the letters A, C, T, G refer to the DNA nucleotide bases adenine, cytosine, thymine, and guanine.

- ***APOE.*** The Apolipoprotein E (APOE) gene is critical to maintenance, repair, and growth of neurons (Yue JK R. C.-A., 2017). APOE is a polymorphic gene with four allelic variants ($\varepsilon 1, \varepsilon 2, \varepsilon 3, \varepsilon 4$) that are defined by two SNPs (rs7412 and rs429358). The $\varepsilon 4$ allele is considered the high-risk variant with an association to Alzheimer's disease. APOE-$\varepsilon 4$, has also been shown to increase one's risk for unfavorable outcomes following a TBI (Yue JK R. C.-A., 2017). APOE plays an important role in neural response to brain injury, but the $\varepsilon 4$ allele causes reduced growth and branching of neurites in vitro (H Houlden, 2006). This study compares the presence and absence of the $\varepsilon 4$ allele. This is the only variant in the study where the allele is defined by two SNPs.
- ***rs1800497.*** The rs1800497 genotype is part of the Ankyrin Repeat and Kinase Domain Containing 1 (ANKK1) gene, that has been shown to play a part in the reduction of the Dopamine Receptor D2 (DRD2) density in the brain and possibly linked to neuropsychiatric disorders (McAllister TW, 2008). Previous studies compared the presence of the minor T allele (T/T, C/T) and absence of the T allele

(C/C), although some studies compare all three genotypes (John K. Yue, 2015) (McAllister TW, 2008). In this study, presence verses absence of the minor T allele is compared.

- ***rs4938016.*** The rs4938016 genotype is also part of the ANKK1 gene and has been shown to have an association with cognitive outcome measures after a brain injury. This study compared the presence (C/C, C/G) of the minor C allele against the absence (G/G) of it, as was done in previous studies (John K. Yue, 2015) (McAllister TW, 2008).
- ***rs11604671.*** The rs11604671 genotype is another SNP that is part of the ANKK1 gene, and it has similar associations with the TBI outcome measures and DRD2 that the rs1800497 and rs4938016 have. This study compared the presence (G/G; A/G) and absence (A/A) of the minor G allele based on the study of the ANKK1 gene affecting cognitive outcomes after a TBI (McAllister TW, 2008).
- ***rs17759659.*** The rs17759659 genotype is part of the BCL2 Apoptosis Regulator gene. It is associated with an increased risk of intracranial hypertension, cerebral edema, and the need for surgical intervention (Deng H, 2021). Previous studies compared the presence (A/G; G/G) and absence (A/A) of the minor G allele on neurobehavioral outcomes after a severe TBI (Nicole Zangrilli Hoh, 2010).
- ***rs6265.*** The rs6265 genotype is part of the Brain Derived Neurotrophic Factor (BDNF) gene and is associated with depression. This study compared the presence (A/A; A/G) and absence (G/G) of the minor A allele as was done in a study that investigated the association between suicidal thoughts and depression in the variant carriers (Sarchiapone M, 2008).

- ***rs4680.*** The rs4680 genotype is part of the Catechol-O-Methyltransferase (COMT) gene, which breaks down dopamine in the brain prefrontal cortex. This study compares the presence (A/A; A/G) and absence (G/G) of the minor A allele. Note that the A allele codes for the amino acid methionine and the G allele codes for valine. The G to A substitution occurs at codon 158. Thus, this SNP is often referred to as $Val^{158}Met$, with Met^{158} (A) being the minor allele and Val^{158} (G) being the major allele (Winkler EA, 2016).
- ***rs6277.*** The rs6277 genotype is part of the Dopamine Receptor D2 (DRD2) gene that is one of several SNPs of the dopamine receptors (Wiebke Bensmann, 2020). Previous studies compared the presence (T/T; C/T) and absence (C/C) of the minor T allele, including a study showing an association between DRD2 and six-month verbal learning following a traumatic brain injury (Yue JK W. E.-A., 2017).
- ***rs6311.*** The rs6311 genotype is part of the 5-Hydroxytryptamine Receptor 2A (HTR2A) gene that is important in human neuropsychiatric disorders (Ryan M. Smith, 2013). In this study, the presence (T/T; C/T) of the minor T allele was compared with the absence (C/C) of the T allele, as suggested by a study comparing anger- and aggression- related traits with the HTR2A gene (Giegling I, 2006).
- ***rs1799971.*** The rs1799971 genotype is part of the Opioid Receptor, Mu 1 (OPRM1) gene. When the amino acid at residue 40, asparagine (Asn), is replaced by aspartic acid (Asp), it can lead to drug and alcohol problems (Esther van den Wildenberg, 2007). Previous studies compared the presence (A/G; G/G) and absence (A/A) of the minor G allele (which codes for Asp) as suggested by a study done on adolescent alcohol misuse (Miranda R, 2010).

1.2.3. Outcome Assessments. The outcome assessment tools for traumatic brain injury patients that were used in this study include the Glasgow Outcome Scale-Extended (GOS-E), Trail Making Test (TMT), Weschler Adult Intelligence Scale 4th Edition (WAIS-IV), Brief Symptom Inventory 18 (BSI-18), and Satisfaction with Life scale (SWLS). For the WAIS-IV assessment, three metrics (composite, percentile, and sum of scaled) from the processing speed index (PSI) component of the test were utilized. These outcomes were selected since data were available for them at the 6- and 12-month post injury time points. The GOS-E also had data available at the 3-month time point. Each assessment is related to a specific domain, and these are summarized for each assessment along with the available time points in Table 1.1(Shirley Ryan, 2019) (Centre for Research Excellence in Brain Recovery) (Centre for Research Excellence in Brain Recovery) (Shirley Ryan, 2015) (Shirley Ryan, 2016). A more thorough explanation of each outcome assessment is provided in Paper I.

Table 1.1. Outcome assessment domains and available post-injury times in months.

Outcome Assessment	**Domain**	**Post-Injury Time (Months)**
GOS-E	Global Outcome	3, 6, 12
TMT	Neuropsychological Impairment	6, 12
WAIS-IV • Composite • Percentile • Sum of Scaled	Neuropsychological Impairment	6, 12
BSI-18	Psychological Status	6, 12
SWLS	Quality of Life	6, 12

1.3 STATISTICAL MODELING APPROACHES

In this section all the statistical methods that are utilized in this research are described in detail. The analysis is performed in a series of steps that are aimed at ensuring a robust result. First, it is important to note that this study is observational in nature, in that the main comparison of interest is between two genotype groups. Genotypes are inherent characteristics of individuals that must be observed and thus randomization is not possible. As such, there is a potential for confounding variables. Although this makes it difficult to establish causation, one attempt to reduce the impact of potential confounding variables is to test for differences on key demographic variables between genotype groups to determine if there is any major imbalance. A chi-squared test is used for categorical variables and a t-test is used for quantitative variables. Any variable that is found to differ significantly between groups can then be included as a covariate in the linear mixed model.

The next step in the analysis is to check assumptions and take action if they are not met. This involves testing SNPs to see if they meet Hardy-Weinberg Equilibrium (HWE) and checking the assumptions of the linear mixed model. Finally, the linear mixed model is performed to test for associations between each SNP and each mTBI outcome measured over time. Since multiple tests are conducted, the false discovery rate is controlled. In this section, the two-sample t-test, chi-squared test, HWE, linear mixed modeling approach, and multiple testing correction method are explained in detail.

1.3.1. Two-sample t-Test. A two-sample t-test was used to compare the means of quantitative demographic/clinical variables (age and education in years) between two genotype groups (variant allele carriers and non-carriers). The two samples should be

independent and normally distributed (de Winter, 2013). The hypotheses for the two-sample t-test are as follows: H_0: $\mu_1 - \mu_2$=0 vs. H_a: $\mu_1 - \mu_2 \neq 0$, where $\mu_1 - \mu_2$ represents the difference in the population means (μ_1, μ_2) for genotype groups 1 and 2, respectively. Under the within-group normality assumption, but allowing the variances to differ between the two groups, the t-statistic is calculated as shown in equation (1)

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t\left(\frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{1}{n_1 - 1}(\frac{s_1^2}{n_1})^2 + \frac{1}{n_2 - 1}(\frac{s_2^2}{n_2})^2}\right) \quad \text{under } H_0. \tag{1}$$

In equation (1), n_1 is the sample size for the first genotype group and n_2 is the sample size of the 2nd genotype group, s_1^2 and s_s^2 are the sample variances, and $\bar{X}_1$ and $\bar{X}_2$ are the sample means of the two genotype groups (Ruxton, 2006). Note the degrees of freedom for the t distribution under the null hypothesis is calculated using the Welch's approximation (Zach, 2019). When the p-value is below 0.05 the test was considered statistically significant. The two-sample t-tests were conducted using the t.test function in R. Results for all two-sample t-tests are found in the Appendix for each SNP. The average age was found to be significant different between carriers and non-carriers for APOE and rs17759659. No other significant differences between genotype groups were found for age or education years were found.

1.3.2. Chi-squared Test. A $\mathcal{X}^2$ test is used to compare the distribution of two categorical or nominal variables in a sample (Stat Trek, n.d.). In this study, a $\mathcal{X}^2$ test is used to test whether the distribution of patients on categorical demographic/clinical variables (sex, race, GCS on admission, ISS score ≤15 or not, and abnormal CT scan or not) differs between genotype group (variant allele carriers vs. non-carriers). The null

hypothesis for the $\mathcal{X}^2$ test is that both genotype groups have the same distribution/proportion of observations within each class of the demographic/clinical categorical variable (e.g., carriers and non-carriers have the same proportion of males and females). The alternative hypothesis is that there is some difference in the distributions/proportions between the genotype groups. The chi-square test-statistic is given in equation (2):

$$\chi^2_{obs} = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2)$$

where O_{ij} is the observed and E_{ij} is the expected number of patients in the i^{th} genotype group for the j^{th}level of the categorical variable. Note that $E_{ij} = \frac{n_i * n_j}{n}$ where n_i is the total number of patients in the i^{th} genotype group, n_j is the total number of patients in the j^{th} level of the categorical variable, and n is the total sample size (McHugh, 2013) (Stat Trek, n.d.). Under the null hypothesis, the approximate distribution of the test statistic (2) is a $\chi^2_{(r-1)*(c-1)}$, where r=2 (number of genotype groups) and c is the number of classes of the demographic/clinical variable being tested. The p-value is calculated from this distribution and if it is less than 0.05, then it was concluded that there was a significant difference in the distribution of the categorical demographic/clinical variable between the genotype groups. Note that a continuity correction was also used since the true distribution of the test statistic is discrete but is approximated by the continuous χ^2 distribution. Also, if the expected frequency for each cell (genotype by categorical variable combination) is less than 5 in more than 20% of the cells, the approximation may not work well. The chi-squared test is carried out by the chisq.test function in R for this study. Results for all chi-squared tests are found in the Appendix for each SNP. The

distribution of GCS values were found to differ significantly between carriers and non-carriers for rs6311. No other significant differences between genotype groups were found for any other SNP or demographic/clinical variable.

1.3.3 Hardy-Weinberg Principle. The Hardy-Weinberg Equilibrium (HWE) is a population genetic principle that is used to estimate genotype frequencies in a population for a genetic locus with two alleles (e.g., A and a) (Nikita Abramovs, 2020). The expected genotype frequencies are based off allele frequencies in the data. Under HWE, the genotype frequencies for A/A, A/a, and a/a should have expected frequencies p^2, $2pq$, and q^2, respectively, where p represents the allele frequency for A and $q = 1 - p$ is the allele frequency for a (Nikita Abramovs, 2020). The HWE principal states that these expected genotype frequencies will remain unaltered over time in the absence of outside factors such as non-random mating, natural selection, and other evolutionary forces (Nikita Abramovs, 2020). In genetic association studies it is important to assess whether each SNP meets the HWE expected genotype frequencies, since violations of HWE are indicative of potential issues such as genotyping errors and population substructure.

HWE is tested at each SNP by performing a chi-squared goodness-of-fit test, using the HWChisq function in the Hardy Weinberg package in R to find the p-value (Cran R-Project, 2022). Under the null hypothesis (H_0), the genotype frequencies at a particular SNP follow those expected under HWE. Under the alternative hypothesis (H_a) there is a deviation from HWE. Equation (3) provides the chi-squared test statistic:

$$\chi^2_{obs} = \frac{\left(n_{AA} - n\hat{p}_A^2\right)^2}{n\hat{p}_A^2} + \frac{(n_{Aa} - 2n\hat{p}_A(1-\hat{p}_A))^2}{2n\hat{p}_A(1-\hat{p}_A)} + \frac{(n_{aa} - n(1-\hat{p}_A)^2)^2}{n(1-\hat{p}_A)^2} \quad (3)$$

where n_{AA} and n_{aa} represent the sample genotype counts for the two homozygotes, n_{Aa}represents number of heterozygotes in the sample, and n is the total sample size (Hendrick, 2011). The estimated population allele frequency, $\hat{p}_A$, for allele A is found by equation (4)

$$\hat{p}_A = \frac{2n_{AA}+n_{Aa}}{2n} \quad (4)$$

(J Graffelman, 2016). The test statistic follows the general form for chi-squared test statistic, as described in Equation (2), where n_{AA}, n_{Aa}, n_{AA} are the observed genotype counts and $n\hat{p}_A^2$, $2n\hat{p}_A(1-\hat{p}_A)$, and $n(1-\hat{p}_A)^2$ are the respective expected genotype counts under HWE. Under the null hypothesis, the approximate distribution of the test statistic (3) is a χ_1^2 distribution, which is used to calculate p-values. Note that a continuity correction was also used since the true distribution of the test statistic is discrete but is approximated by the continuous χ^2 distribution. Any SNP with a p-value <0.05 indicates the null hypothesis can be rejected and that SNP violates HWE. Otherwise, the SNP is assumed to be in HWE.

The results of the HWE test for the 10 potential individual SNPs to be included in the study are shown in Table 1.2. After completion of the HWE test, one SNP, rs3219119, was removed because it had a significant deviation from HWE. The SNP rs1800497 was shown to deviate from HWE (p=0.044), but it was not removed to compare to work done in previous studies (John K. Yue, 2015). Note that for the APOE gene, this study only tested for presence/absence of the $\varepsilon 4$ allele, which is defined by two SNPs (rs429358 and rs7412). Those SNPs are not tested individually for genetic associations with TBI outcomes and are thus not included in the HWE testing. In total, there were 9 SNPs plus

the APOE-ε4 presence/absence that were included in the genetic association analysis in Paper I after the HWE testing.

Table 1.2. Hardy-Weinberg Equilibrium (HWE) p-value for each single nucleotide polymorphism (SNP) tested.

SNP	HW p-value
rs1800497	0.044
rs4938016	0.70
rs11604671	0.084
rs17759659	0.33
rs6265	0.36
rs4680	0.56
rs6277	0.079
rs6311	0.97
rs1799971	0.88
rs3219119	0.00016

1.3.4. Linear Model. The model used in this work to test whether there is a significant association between genotypes and mTBI outcomes is a linear mixed model. In this section, important concepts related to linear modeling are discussed prior to describing the linear mixed model. A linear model shows the relationship between an observable response and an observable design matrix of predictor variables (Peña EA, 2006), as given in equation (5):

$$Y = X\beta + \varepsilon \tag{5}$$

where Y is the response vector, X is the design matrix of predictor variables, β is the vector of unknown regression coefficients, ε is the unobservable error, and σ is the unknown error standard deviation, (Muller, 2004). The assumptions of the model include:

- Linearity: The assumption that the relationship between the predictor variables and response variable is linear. This can be assessed using a scatterplot and residual vs. fitted values plot (Deanna Schreiber-Gregory, 2018).
- Normality: This assumption requires that the distribution of the error terms follows a normal distribution (bell shape curve) (Deanna Schreiber-Gregory, 2018). This assumption can be checked with a normal quantile-quantile (Q-Q) plot of the residuals.
- Constant Variance: This assumption assumes the errors have constant variance σ^2 and this can be checked using the residuals vs. fitted values plot. This is also known as homoscedasticity (Nahhas, 2023).
- Independence: The error terms are also assumed to be independent. This assumption is typically assumed based on the nature of how the data were collected.

1.3.5. Linear Mixed Model. For longitudinal data where data are collected on the same subjects at multiple time points, the independence assumption of the previously described linear model is unlikely to be met since measurements collected on the same subject are likely to be correlated. Linear mixed models provide an ideal way to handle this type of data where subjects are considered to be random effects that can vary from the overall relationship of the response over time. The model is called mixed since it includes both fixed and random effects. The linear mixed model can be written as in equation (6):

$$Y = X\beta + Zb + \varepsilon \tag{6}$$

where Y is the response, β is the fixed-effect parameter vector, b is the random-effect parameter vector, ε is the random error, and X and Z are the design matrices of the fixed and random effects, respectively (Douglas Bates, 2015). The linearity between X and Y is still assumed in this model, as is the normality and constant variance of the errors ε. The b random effects are assumed to be independent of ε and are also assumed to be normally distributed with a variance-covariance matrix Σ. Scatterplots and residual plots can also be used to check assumptions similar to those described in Section 1.3.4 for the linear model.

For this study, a linear mixed model was fit for each SNP and outcome assessment (Y_{ij}) as shown in equation (7):

$$Y_{ij} = \beta_0 + \beta_1 Time_i + \beta_2 Genotype_j + \beta_3 Time_i * Genotype_j + \gamma_{0j} + \varepsilon_{ij}, \quad (7)$$

for $i = 1, \dots, I$ time points and $j = 1, \dots, n$ individuals (Douglas Bates, 2015) (Sven Hilbert, 2019). The fixed effects part of the model includes $Time$, $Genotype$, and their interaction ($Time * Genotype$), with their respective regression coefficients $\beta_1, \beta_2, \beta_3$, and the overall intercept β_0.The γ_{0j} term represents a random subject effect and the ε_{ij} term is the random error. It is assumed that $\gamma_{0j} \sim N(0, \tau^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$ are independent. The genotype groups were coded such that non-carriers of the variant allele were the reference group coded as 0 and the variant allele carriers were coded as 1. Thus, the model can be broken down by genotype group as follows.

When $Genotype = 0$, the model for non-carriers becomes, equation (8):

$$Y_{ij} = \beta_0 + \beta_1 Time_i + \gamma_{0j} + \varepsilon_{ij}. \quad (8)$$

When $Genotype = 1$, the model for the carriers of the variant allele becomes, equation (9):

$$Y_{ij} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)Time_i + \gamma_{0j} + \varepsilon_{ij}. \quad (9)$$

Thus, β_0 and β_1 are the intercept and slope for the linear model of the outcome over time for the non-carriers of the variant allele. The β_2 and β_3 coefficients represent the difference in the overall intercept and slope, respectively, between the variant allele carriers and non-carriers. By including the random subject effects, γ_{0j}, each subject is allowed to deviate from the overall intercept in both genotypes.

To determine if there is a significant association between the SNP and the mTBI outcome, a set of tests is conducted. First, the interaction term is tested ($H_0: \beta_3 = 0$ vs. $H_a: \beta_3 \neq 0$). This tests whether there is a difference in slopes between the carriers and non-carriers. That is, this tests whether there is a difference in the rate of change in the outcome over time between the genotype groups. Thus, a significant interaction effect would indicate that the rate of recovery on a specific TBI assessment is associated with what genotype the individuals has for that SNP. If the interaction is not significant and the slopes are not significantly different between the genotypes, then a test for the genotype effect ($H_0: \beta_2 = 0$ vs. $H_a: \beta_2 \neq 0$) is conducted. A significant genotype effect would indicate that there is a significant difference in the average outcome between carriers and non-carriers (irrespective of time). These are the primary tests of interest in this study focused on identifying genetic associations with mTBI outcomes. As a secondary analysis, the time effect is also tested ($H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$) when the interaction is

not significant to determine if there is a significant change in the outcome over time (irrespective of genotype).

Some SNPs may require the inclusion of a covariate, if there were any significant differences between demographic/clinical variables between the genotype groups. When a covariate is included, the model becomes, equation (10):

$$Y_{ij} = \beta_0 + \beta_1 Time_i + \beta_2 Genotype_j + \beta_3 Time_i * Genotype_j + \beta_4 Covariate_j + \gamma_{0j} + \varepsilon_{ij}. \quad (10)$$

The same tests for β_2 and β_3 as described for model (7) are performed as the primary analysis since the focus in this study is on the genetic associations. The tests for $Time$ (H_0: $\beta_1 = 0$) and $Covariate$ (H_0: $\beta_4 = 0$) are also performed as a secondary analysis.

The linear mixed model was fit using the lmer function from the lme4 library in R for the models with (10) and without a covariate (7) (Douglas Bates, 2015) (Wiley, 2020). The model is performed as shown below in equation (11)

$$lmer\left(Y \sim Time * Genotype + (1|GUID)\right) \quad (11)$$

for model (7) without a covariate and equation (12)

$$lmer\left(Y \sim Time * Genotype + Covariate + (1|GUID)\right) \quad (12)$$

for the model (10) with a covariate. The GUID represents an identifier for an individual subject. The estimation method used in lme4 is restricted maximum likelihood (REML), which is the standard method for estimating parameters in linear mixed models (Christopher M. Gotwalt). It has less bias for variance estimates by estimating the variances first, then estimating the fixed effects coefficients (Newsom, 2019). To test the

fixed effects of interest, a test statistic whose null distribution is approximated by an F-distribution with the denominator degrees of freedom approximated by the Kenward Roger method was utilized to obtain the p-values for the mixed model. This is obtained via the pbkrtest package in R (Cran R-Project, 2021). It is important to note that this estimation procedure for linear mixed models allows the use of all available data for a subject. Thus, if a subject has data available at one time but is missing data at another time point, the subject will still be included in the modeling. As long as the data are missing at random, the linear mixed model is robust to potential bias from missing data (G. L. Gadbury, 2003).

1.3.6. Multiple Testing Correction. The false discovery rate (FDR) estimation is used to combat the higher likelihood of a false positive occurring across multiple tests compared to a single test. The FDR is defined as the expected proportion of discoveries (rejected null hypotheses) that are false. FDR estimates are computed directly from the p-values, which are assumed to follow a uniform distribution under the null hypothesis (Noble, 2009). That is, the p-values follows equation (13) under the null hypothesis:

$$P \sim \mathcal{U}[0,1] \tag{13}$$

so multiple testing corrections can estimate the proportion of false positives (Maarten van Iterson, 2010). The FDR procedure works by sorting the p-values in ascending order, then dividing each observed p-value by its percentile rank to get an estimate (Noble, 2009). The ranked p-values for a total of m tests are denoted as $p_{(i)}$ where $i = 1, 2, \dots, m$ (Winkler, 2011). The Benjamini and Hochberg's FDR-controlling procedure works by letting k be the largest i such that equation (14) holds:

$$p_{(i)} < \frac{iq}{m}, \tag{14}$$

where q is the desired FDR level. The null hypotheses are then rejected for all $i = 1, \ldots, k$. Equation (15) provides the FDR-corrected p-values (Winkler, 2011):

$$q_{(i)} = \frac{p_{(i)}m}{i}. \tag{15}$$

One additional adjustment is made to ensure monotonicity, such that the corrected p-values for each i are the smallest $q_{(k)}$ where $k \geq i$. This formulation allows for a significance declaration and a new threshold of $q_{(i)}$ for significance determinations for when the FDR-corrected p-value is < 0.05 (Winkler, 2011) (Yekutieli, 2001). In this work, the FDR was utilized across the multiple tests for each SNP. The linear mixed model results with both the original and FDR-corrected p-values can be found in the Appendix with more details in Paper I.

PAPER
I. IDENTIFYING ASSOCIATIONS BETWEEN SINGLE NUCLEOTIDE POLYMORPHISMS AND MILD TRAUMATIC BRAIN INJURY OUTCOMES OVER TIME

ABSTRACT

Traumatic brain injury (TBI) is a growing health concern, with millions of TBI diagnoses in the United States each year. The vast majority of TBI diagnoses are mild traumatic brain injuries (mTBI), which can be challenging to manage due to variation in symptoms and outcomes. Most individuals with mTBI successfully recover quickly, but a small subset has a delayed recovery. Although the factors that contribute to this variation in recovery are not clearly understood, it is possible that genetic differences may play a role. Few studies have investigated the association between single nucleotide polymorphisms (SNPs) with mTBI outcomes and this is an emerging area of research. In this study, we utilize data collected in the Transforming Research and Clinical Knowledge in Traumatic Brain Injury Pilot (TRACK-TBI Pilot) study to test the association between 10 different SNPs and 7 TBI outcomes measured at six- and twelve-months post injury. Linear mixed models are utilized to investigate the association between genotypes and mTBI outcome measurements over time. Previous studies have primarily focused on a single time point at six months for one or two SNPs. The findings in this study demonstrate the potential benefits of using linear mixed models to identify relationships between genotypes and TBI outcomes over time.

1. INTRODUCTION

Traumatic brain injury (TBI) occurs when an external force causes a change in brain function or pathology (Menon DK & Health, 2010). Approximately 2.5 million people in the United States (US) sustain a TBI annually and up to 30% of injury related deaths in the US are related to TBIs (John K. Yue, 2015). It is estimated that long-term disability from TBIs currently affects up to 5.3 million people (Yue JK W. E.-A., 2017). TBIs have a wide range of severity, with most injuries (70-90%) classified as mild (mTBI) on an initial injury severity assessment called Glasgow Coma Scale (GCS) (Winkler EA, 2016). Most recovery from mTBI happens within 3 months if no other risk factors are present (Hilary Bertisch & Investigators, 2019). However, around 20% of mTBI patients have a prolonged experience with symptoms (e.g., headache and fatigue) and cognitive deficits (Winkler EA, 2016) (Hilary Bertisch & Investigators, 2019). Within mTBIs, a large amount of variability in outcomes appears across individuals (Winkler EA, 2016). Individuals with injuries that are nearly identical often have different symptoms resulting in variability in outcomes (Winkler EA, 2016). Furthermore, more studies are needed to identify factors that can help explain this variability in outcomes among individuals with mTBIs. Gaining a better understanding of which individuals are at risk for poor outcomes could assist in designing individual recovery plans.

One avenue that recent studies have explored as a potential factor in explaining some of the clinical variability is the association between genetic variations within individuals' genes and outcomes following a TBI (John K. Yue, 2015) (Yue JK W. E.-A.,

2017) (Winkler EA, 2016) (Yue JK R. C.-A., 2017) (Ethan A. Winkler, 2017). These genetic association analysis studies investigate a type of genetic variation called a single nucleotide polymorphism (SNP), which involves a single nucleotide substitution that occurs in the DNA. When SNPs arise within a gene's coding sequence or regulatory element, they can influence protein structures or abundance (Winkler EA, 2016). Several candidate SNPs have been identified that occur within genes that are known to be involved in processes important to cognitive function. Recent studies have tested for associations between some of these candidate SNPs in genes of interest (ANKK1, DRD2, APOE, and COMT) and TBI outcomes (John K. Yue, 2015) (Yue JK W. E.-A., 2017) (Winkler EA, 2016) (Yue JK R. C.-A., 2017) (Ethan A. Winkler, 2017).

One set of studies among TBI patients of all severities investigated SNPs in genes that are involved in the dopaminergic pathway, which is important to attention, memory, and executive function (John K. Yue, 2015) (Yue JK W. E.-A., 2017). The ankyrin repeat and kinase domain-containing 1 (ANKK1) gene, which is involved in dopamine transmission, has a C/T SNP (rs1800497, also known as Taq1A) (John K. Yue, 2015). Within TBI patients, the presence of the T allele of Taq1A has been shown to have a negative association with episodic memory and response latency 1-month post-injury (McAllister TW F. L., 2008). A similar association was found in another study on verbal learning ability and non-verbal processing 6-months after injury that identified poorer performance among the T/T homozygotes (John K. Yue, 2015). The dopamine D2 receptor (DRD2) gene is adjacent to ANKK1, and it also has a C/T SNP (rs6277, also known as C957T) (Yue JK W. E.-A., 2017). TBI patients with the T allele of C957T showed better verbal learning and recall at 6-months, even after adjusting for the ANKK1

genotype and other potential covariates. No significant associations were found for non-verbal processing speed or mental flexibility, indicating that the T allele may only exhibit a performance advantage in selective cognitive domains (Yue JK W. E.-A., 2017).

Some studies have focused on conducting genetic association analysis within mTBI patients. This can help in reducing the clinical heterogeneity that occurs between different TBI severities while elucidating genetic connections that may be influencing the variation within mTBI outcomes. One gene studied in mTBI patients is Apolipoprotein E (APOE), which is a highly polymorphic gene that plays a role in neuronal survival and execution of antioxidant effects (Yue JK R. C.-A., 2017). The APOE- $\varepsilon 4$ allele is a risk factor for neurogenerative disorders such as Alzheimer's disease (Yue JK R. C.-A., 2017). Within mTBI patients, the APOE-$\varepsilon 4$ allele has been associated with an increased risk of verbal memory impairment at the 6-month time point (Yue JK R. C.-A., 2017). The Catechol-O-Methyltransferase (COMT) gene has also been investigated in mTBI patients. COMT is an enzyme that inactivates catecholamine neurotransmitters and is involved in regulating dopamine transmission in brain regions important to cognition. COMT has a SNP (rs4680, also known as $Val^{158}Met$) in which valine (G; Val^{158}) is substituted with methionine (A; Met^{158}) at the codon 158. The Met^{158} allele is associated with higher catecholamine bioavailability with improved performance on memory and attention tasks (Winkler EA, 2016) (Ethan A. Winkler, 2017). In mTBI patients, the Met^{158} allele is associated with lower incidence in post-traumatic stress disorder and improved global function outcomes, as well as increased performance of non-verbal processing speed at the 6-month time point (Winkler EA, 2016) (Ethan A. Winkler, 2017).

These studies indicate that the impact of genetic effects on mTBI patients is an important component in disentangling the heterogeneity in post-injury outcomes and further research in this area is needed. The previous studies provide focused investigations on the association between a single SNP with a few TBI outcomes at one time point, typically at six months post-injury. However, a wealth of data is available through the Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) Pilot data that would enable a more thorough investigation across multiple SNPs, TBI outcomes, and post-injury time points. The TRACK-TBI Pilot study was conducted at three Level I trauma centers on patients who completed a computed tomography (CT) scan within 24 hours of sustaining a TBI (Winkler EA, 2016). Data were collected on demographics, clinical information, neuroimaging, and outcome assessments. Genotype information was also collected on a subset of the patients for a set of candidate SNPs that may be connected to cognitive outcomes (Yue JK W. E.-A., 2017). The pilot study had a total of ~600 subjects and was the foundation for the TRACK-TBI study that now has 3000 patients from 18 sites (University of California, 2014).

In this work, data from the TRACK-TBI Pilot study were used to test the association between genotypes of 10 different SNPs and 7 outcome assessments in mTBI patients. Each outcome is measured for at least two post-injury time points (6 and 12 months). A linear mixed model is employed to determine if the changes in post-injury outcomes over time differ significantly between genotypes. Since most previous studies focus on one post-injury time point, this work provides a unique perspective by exploring whether genetics plays a role in recovery over time. For SNPs where no such significant

association over time is found, an overall association test between genotype and outcome is conducted, irrespective of time. Through this comprehensive investigation of multiple SNPS and mTBI outcomes together in one study, the chance of false associations can be better controlled across the multiple tests conducted. Additionally, previously unexplored SNP and mTBI outcome associations have the potential to be revealed.

2. DESCRIPTION OF DATA

2.1. STUDY DESIGN

This study uses data from the Transforming Research and Clinical Knowledge in Traumatic Brain Injury Pilot (TRACK-TBI Pilot) study. TRACK-TBI Pilot data is available through the Federal Interagency Traumatic Brain Injury Research (FITBIR) website, which motivated this work (Health, n.d.). The data utilized in this work was obtained from the TRACK-TBI Pilot Study Principal Investigator (PI), Dr. Geoffrey Manley. This data set included data available on FITBIR as well as additional SNP data and 12-month neurocognitive assessment data that were not available through FITBIR.

As described in (Yue JK & Investigators, 2013), TRACK-TBI Pilot is a prospective observational study conducted at three acute care sites that are level 1 trauma centers (San Francisco General Hospital, University of Pittsburgh Medical Center, and University Medical Center Brackenridge in Austin, Texas). Patients who visited one of the trauma centers and received a computed tomography (CT) scan within 24 hours of a head injury were considered for inclusion in the study. Patients were excluded if they were less than 18 years old, pregnant, incarcerated, on a psychiatric hold, or diagnosed

with a life-threatening condition (Yue JK W. E.-A., 2017). Non-English speakers were also excluded due to challenges in obtaining accurate outcome assessments, which were administered in English (Yue JK W. E.-A., 2017). Information about the Institutional Review Board and informed consent can be found in the original paper describing the study (Yue JK & Investigators, 2013). A total of ~600 patients, recruited through convenience sampling, were enrolled in the study at the three sites (Health, n.d.). The initial study also included 51 patients from a rehabilitation center that were not included in this work (Yue JK & Investigators, 2013).

The TRACK-TBI Pilot study focused on collecting data in the set of Common Data Elements (CDEs) for TBI patients, as defined by the National Institutes of Health and National Institute of Neurological Disorders and Stroke (Winkler EA, 2016). A wealth of demographic, clinical, biomarker, and neuroimaging data were collected on patients in the acute stage after the injury. Outcome assessments were also collected at multiple post-injury time points to investigate patient recovery trajectories. The initial study targeted the 3- and 6-month time points, but data were also collected at 12 months for some outcome assessments. In a subset of the patients, genetic data were obtained from blood samples to obtain the genotypes for a set of candidate SNPs. The PI of the TRACK-TBI Pilot study, Dr. Geoffrey Manley, provided SNP data and 12 month neurocognitive assessment data that were not available from the FITBIR database.

2.2. PATIENT SELECTION

The goal of the current work is to test the association between candidate SNPs and prolonged post-injury outcomes in adult mild TBI patients. Thus, patients in the

current work were restricted to those with genetic data available on at least one SNP, had a GCS score of 13-15 indicating mild injury, and were between 18-80 years old. Additionally, outcomes with data collected at 6 and 12 months were selected to investigate recovery over a longer time period. Only patients with data on at least one of the selected outcomes in at least one time point were included in the current study.

2.3. DEMOGRAPHIC AND CLINICAL DATA

To understand the characteristics of patients included in the study, a certain set of demographic and clinical variables were selected. Variables that may have an impact on the outcome assessments were chosen and summarized between different genotypes for each SNP to determine if the groups were evenly balanced. The demographics include age (years), sex (male or female), race (Caucasian, African American/African or other races), and education (years). For the race data, the "other races" group was combined due to the small sample sizes within individual races that were not Caucasian or African American/African. The clinical data collected in the acute phase include GCS score (13, 14, 15) upon arrival at the emergency department, Injury Severity Score (ISS) (classified into two groups: ≤15 or >15), and whether or not the patient had an abnormal CT scan.

2.4. GENOTYPE DATA

Blood samples were collected within 24 hours of injury for genotyping on a set of candidate SNPs. A more detailed description of how samples were collected and processed, including the DNA extraction and genotyping methods, is described in (Yue JK & Investigators, 2013). In this work, 10 SNPs are tested for an association with

different outcome assessments in mTBI patients. For each SNP, it is important to determine which genotypes should be compared between the outcomes. In this study, the presence of the minor allele (i.e., minor allele carrier) is tested against the absence of the minor allele. Thus, there are two groups: absence (coded as 0) of minor allele (i.e., homozygous for the major allele) vs. presence (coded as 1) of the minor allele (i.e., homozygous for minor allele or heterozygous).

Table 1. Summary of single nucleotide polymorphisms (SNPs) and genotype comparisons.

SNP/variant	Gene Symbol	Full Gene Name	Variant Allele	Genotype Comparison
rs1800497	ANKK1	Ankyrin Repeat and Kinase Domain Containing 1	T	0: C/C | 1:C/T, T/T
rs4938016	ANKK1	Ankyrin Repeat and Kinase Domain Containing 1	C	0: G/G | 1: C/G, C/C
rs11604671	ANKK1	Ankyrin Repeat and Kinase Domain Containing 1	G	0: A/A | 1:A/G, G/G
APOE	APOE	Apolipoprotein E	$\varepsilon 4$	0: $\varepsilon 4$ Absent | 1: $\varepsilon 4$ Present
rs17759659	BCL2	BCL2 Apoptosis Regulator	G	0: A/A | 1: A/G, G/G
rs6265	BDNF	Brain Derived Neurotrophic Factor	A	0: G/G | 1: A/G, A/A
rs4680	COMT	Catechol-O-Methyltransferase	A (Met)	0: G(Val)/G(Val) | 1: A(Met)/G(Val), A(Met)/A(Met)
rs6277	DRD2	Dopamine Receptor D2	T	0: C/C | 1: C/T, T/T
rs6311	HTR2A	5-Hydroxytryptamine Receptor 2A	T	0: C/C | 1: C/T, T/T
rs1799971	OPRM1	Opioid Receptor, Mu 1	G	0:A/A | 1:A/G, G/G

The only exception is the comparison for APOE, since the allelic variants ($\varepsilon1, \varepsilon2, \varepsilon3, \varepsilon4$) are based on two SNPs (rs7412, rs429358), with the $\varepsilon4$ allele being the high-risk variant associated with Alzheimer's disease. Thus, the comparison of interest is presence vs. absence of the $\varepsilon4$ allele (Yue JK R. C.-A., 2017). Comparisons for all other SNPs were determined by reviewing previous studies (John K. Yue, 2015) (Yue JK W. E.-A., 2017) (Winkler EA, 2016) (Ethan A. Winkler, 2017) (Giegling I, 2006) (Nicole Zangrilli Hoh, 2010) (McAllister TW F. L., 2008) (Sarchiapone M, 2008). Table 1 provides a list of the 10 SNPs investigated in this work, along with a description of the gene in which they are located and the comparison of interest that will be tested. The letters A, C, G, T represent the nucleotide bases on the DNA sequence (A=adenine, C=cytosine, G=guanine, T=thymine). Val and Met are the amino acids valine and methionine, respectively. The polymorphism in the APOE gene is a combination of two SNPs (rs429358 and rs7412). The $\varepsilon4$ allele is the high-risk variant.

2.5. OUTCOME ASSESSMENTS

The outcome assessments chosen for this study were core CDEs with data available at the 6- and 12-month time points. All available assessments that met these criteria were utilized, apart from the Craig Handicap Assessment and Reporting Technique (CHART). CHART data were excluded due to a lack of variability in the scores, as most patients attained the highest score (100) on each subscale at both time points. A total of 7 outcomes (from 5 different assessments) were included in the genetic association analysis.

2.5.1 Glasgow Outcome Scale – Extended. The Glasgow Outcome Scale – Extended (GOS-E) is in the global outcome domain for TBI patients 18 and older

(Wilson L, 2021). The assessment is used as a performance measure to assess the global disability and recovery after a TBI (Wilson L, 2021) (Kreitzer NP, 2019). An overall evaluation for a patient is determined through a structured interview with questions related to consciousness, independence, employment, social and leisure activities, relationships, and return to daily life (Wilson L, 2021). The GOS-E has a score minimum of one and maximum of eight, with the following ordinal categories 1- Dead, 2-Vegetative State, 3-Lower Severe Disability, 4-Upper Severe Disability, 5-Lower Moderate Disability, 6 - Upper Moderate Disability, 7-Lower Good Recovery, 8-Upper Good Recovery (Wilson L, 2021). The data set has three-, six-, and twelve-month outcome time points for GOS-E. Note this is the only outcome with data at the three-month time point in this study.

2.5.2 Trail Making Test. The Trail Making Test (TMT) is in the neuropsychological impairment domain and has six- and twelve- month outcome time points. TMT is used to measure the neuropsychological functions, such as attention, memory, and executive function, that often affect everyday activities and social role participation (Salthouse, 2011). TMT is a timed test that is split into two parts (A and B), where the participant is given a task to perform in each part. TMT-A assesses visual attention and processing, while TMT-B is related to working memory and task switching ability (SÁNCHEZ-CUBILLO, 2009). The score on each part is the number of seconds required to complete the task, with lower scores indicating better performance. In order to attain a more general indicator of executive control function, the ratio (B/A) is used in this work as a commonly derived index of the TMT components where a lower value is preferred (Winkler EA, 2016) (Salthouse, 2011).

2.5.3 Weschler Adult Intelligence Scale, Fourth Edition. The Weschler Adult Intelligence Scale (WAIS) is in the neuropsychological domain with six- and twelve-month outcomes split into several sections, composite, percentile, and sum of scaled. The test was standardized for patients age 16 to 90 in the USA and are composed of 10 core subsets that produce scores for verbal comprehension, perceptual reasoning, working memory, and processing speed and an IQ level (Gomaa Said Mohamed Abdelhamid, 2021). This study used the processing speed index (PSI) consisting of symbol searching and coding tests for the composite and percentile scores (Injury, 2018).

The composite score ranges from 50 to 150 for correspondence across age groups with the 0.1st to 99.9th percentile in performance (Winkler EA, 2016). The percentile rank score is the percentile in which the patient's composite PSI falls in, with a range of 0.1 to 99.9 (The Washington Center for Cognitive Therapy, n.d.). The descriptive classifications are based on IQ level and percentile ranges are as follows by level and percentage, intellectual disability (level 69 and below, 0.01%-2%), borderline (level 70 to 79, 2%-8%), low average (level 80 to 89, 9%-23%), average (level 90 to 109, 25%-73%), high average (level 110 to 119, 75%-90%), superior (level 120 to 129, 91%-97%), and very superior (level 130 and above, 98%-99.9%) (The Washington Center for Cognitive Therapy, n.d.). The sum of scaled score is the proration for the verbal comprehension or the perceptual reasoning indices (Traci W. Olivier, 2013). All scores are preferred to be higher representing higher improvement in their specific category (verbal comprehension, perceptual reasoning, working memory, and non-verbal processing speed) (Gomaa Said Mohamed Abdelhamid, 2021) (Traci W. Olivier, 2013).

2.5.4 Brief Symptom Inventory. Brief Symptom Inventory-18 (BSI-18) is in the psychological domain with six- and twelve- month outcomes. BSI-18 measures psychological distress and psychiatric disorders associated with mTBIs that can affect life adjustments, personality changes, or mood disturbances (Recklitis CJ, 2007) (Injury, 2018). The self-reported questionnaire consists of six descriptions of psychical and emotional pain symptoms in the past seven days (Injury, 2018). The Likert scales ranges from 0 meaning "not at all" and 4 meaning "extremely" and cannot be administered to anyone under the age of 18 (15). The BSI-18 ranges from 0 to 24 and is a subset of the Global Severity Index that has a range of 0 to 72 which high scores measure a higher level of psychological distress (Injury, 2018) (Merport A, 2012).

2.5.5 Satisfaction with Life Scale. The Satisfaction with Life Scale (SWLS) is in the perceived generic and disease-specific health related quality of life domain with six- and twelve- month time points. SWLS is a self-reported measurement used to measure the life satisfaction component of subjective well-being and has been linked to mental health and predictive future behaviors (Kreitzer NP, 2019) (Injury, 2018). The seven-point Likert style response scale with eight sub-scales (physical functioning, role limitations-physical, bodily pain, general health, vitality, social functioning, role limitations-emotional, and menta health) range from 5 to 35, with a higher score indicating a higher satisfaction with life (Kreitzer NP, 2019) (Injury, 2018).

3. STATISTICAL METHODS

This study aims to examine the genetic associations on mTBI outcomes over time for multiple candidate SNPs. To accomplish this goal, the statistical modeling framework described in Figure 1 was employed to ensure a robust analysis. The modeling framework consists of five phases (data extraction, demographic analysis, check assumptions, statistical modeling, and multiple testing correction). Each of these phases is described in detail in the following subsections. R statistical software (version 4.2.2) was utilized for all statistical analyses.

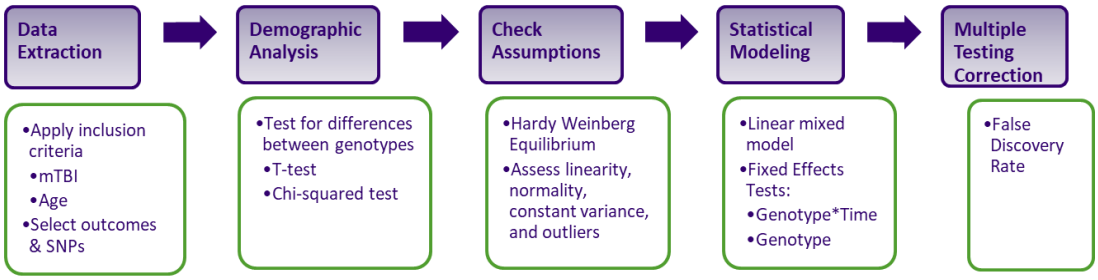

Figure 1. Statistical Modeling Framework.

3.1 DATA EXTRACTION

The first step in the analysis requires extracting data needed for this work from the larger TRACK-TBI Pilot data set. This involves discussion with domain experts to determine the study population, outcome assessments, and SNPs to target, so that the appropriate subset of the data can be obtained. For this study, the inclusion criteria described in Section 2.2 are applied to filter out any individuals with TBIs that were not

mild cases or were outside of the 18-80 age range. Among individuals that met the inclusion criteria, only those with data on at least one SNP and at least one selected outcome was extracted and utilized in the statistical analysis. A total sample size of n=330 patients met these criteria.

3.2. DEMOGRAPHIC ANALYSIS

An initial analysis was performed on the seven demographic and clinical variables described in Section 2.3 to summarize the characteristics of the study sample. Means and standard deviations were found for the quantitative variables (age, education in years), while counts and percentages per group were found for the categorical variables (race, sex, ISS score $\leq$15 or not, GCS score on admission, and abnormal CT scan status). It is important to compare key demographic and clinical variables that may influence the outcome assessment results between the genotypes being tested. Since this is an observational study, this helps determine if the genotype groups are evenly mixed. Any variables that are unbalanced can be addressed in the further modeling. To test for significant associations between the genotype and the seven demographic/clinical variables, *t*-tests and chi-squared tests were conducted for each SNP. A two-sample *t*-test, assuming unequal variance, compared the normally distributed quantitative demographic/clinical variables between the two genotype comparisons groups. A chi-squared test aimed to analyze group differences when the demographic/clinical variable is categorical (McHugh, 2013). Any demographic/clinical variable that was significantly different ($p<0.05$) between genotypes were then included as covariates in the linear mixed model for that SNP.

3.3. CHECK ASSUMPTIONS

Prior to making formal statistical inferences, it is important to check both genetic and modeling assumptions and apply appropriate filters/remedies when needed. The first aspect of checking assumptions involves testing each SNP to determine if it follows the Hardy-Weinberg Equilibrium (HWE). HWE is an important genetic principal that states that genotype frequencies should remain constant over generations unless disturbed by an outside factor (e.g., natural selection, genetic drift, non-random mating) (Nikita Abramovs A. B., 2020). In genetic association studies, deviations from HWE are potentially indicative of issues such as genotyping errors (Moreno, 2013). When testing HWE, the expected frequency of both homozygous genotypes (i.e., two major alleles or two variant alleles) and the heterozygous variant carriers (i.e., one major allele, one variant allele) are estimated from allele frequencies in the population. These proportions should remain constant even with increases and decreases in population size (Nikita Abramovs A. B., 2020). The chi-squared test with a continuity correction was used to test HWE (Jan Graffelman, 2023). Thus, if the SNP genotype frequencies deviates significantly ($p<0.05$) from those expected under HWE, the SNP was taken out of the study due to its instability.

There are several assumptions required for the linear mixed model, described in the next section (Section 3.4). These assumptions include normality and constant variance of the errors and the random subject effect. Linearity is also assumed between the outcome and time, as well as any continuous covariate. Prior to interpreting model results, it is important to check whether the assumptions are met and perform remedies if needed. Linearity was checked using scatterplots. Constant error variance was

investigated using residuals vs. fitted plots (Nahhas, 2023) (Deanna Schreiber-Gregory, 2018). Normality of the errors was checked to ensure the model residuals followed a normal distribution by using a Quantile-Quantile plot (Deanna Schreiber-Gregory, 2018). For any outcome assessment for which these assumptions were violated, a transformation of the outcome was considered.

3.4. LINEAR MIXED MODELING

Linear mixed models have several advantages for modeling the type of longitudinal data collected in this study. First, they allow the possibility of testing whether changes in mTBI outcomes over time differ between genotypes. In the case that the changes over time do not differ by genotype, the model also provides a test for overall differences in outcomes between the genotypes. Additionally, linear mixed models account for the correlation between data collected on the same subject over time, while allowing each subject to vary from the overall average within their genotype group. Finally, the mixed modeling process can also help combat the challenge of having missing data, since not all subjects may complete the outcome assessments at all the time points. Linear mixed models use all of the available data for a subject and are robust when data are missing at random (G. L. Gadbury, 2003). The specific model utilized for this study is described in detail below.

For each SNP, a linear mixed model was fit for each outcome assessment (Y_{ij}) shown in equation (1):

$$Y_{ij} = \beta_0 + \beta_1 Time_i + \beta_2 Genotype_j + \beta_3 Time_i * Genotype_j + \gamma_{0j} + \varepsilon_{ij}, \quad (1)$$

for $i = 1, \ldots, I$ time points and $j = 1, \ldots, n$ individuals (Douglas Bates, 2015) (Sven Hilbert, 2019). The number of time points is I =2 (6 and 12 months) for all outcomes except GOSE, which has I=3 outcomes (3, 6, and 12 months). The n varies slightly for each outcome and SNP, depending on the number of subjects with available data. The fixed effects include $Time$, $Genotype$, and their interaction ($Time * Genotype$). Note that the genotype groups were coded as described in Table 1, with non-carriers of the variant allele being the reference group coded as 0 and the variant allele carriers coded as 1. A random subject effect (γ_{0j}) was included due to the multiple measurements taken on the same subject over time. The ε_{ij} term represents the random error. In this model, β_0 represents the overall intercept for the non-carriers of the variant allele (reference genotype group), β_1 is the change in the average outcome for each additional month post-injury (i.e., slope) for the non-carriers, β_2 is the difference in the overall intercept between the variant allele carriers vs. non-carriers, and β_3 indicates how much the change in the outcome over time differs between the carriers and non-carriers (i.e., difference in slopes between genotype groups). The random subject effect (γ_{0j}) represent deviations from the overall intercept for individual subjects and are assumed to follow a normal distribution with a mean of 0 and a variance τ^2. The random errors (ε_{ij}) are also assumed to follow a normal distribution with a mean of 0 and a variance of σ^2. The ε_{ij} and γ_{0j} are assumed to be independent. If any of the demographic or clinical variables showed significant differences between genotype groups, it was also included as a fixed covariate for that SNP.

The restricted maximum likelihood method was used for estimating the model parameters with the 'lme4' package in R (Douglas Bates, 2015). To obtain p-values for

the testing significance of the fixed effects, the Kenward-Roger approach was utilized to approximate the denominator degrees of freedom for the F-tests (Luke, 2016). The primary tests of interest are the fixed effects that involve the genotype, as these indicate whether or not there is a genetic association with the outcome. The $Time * Genotype$ interaction was tested first to determine if the changes in outcome assessments over time differed by genotype. If the interaction term was not significant, then the $Genotype$ effect was tested. This determines if there is a significant difference in the average outcome between carriers and non-carriers (regardless of time). As a secondary analysis, when the interaction was not significant, the $Time$ effect was also tested to determine if there were any overall changes in the average outcomes over time (regardless of genotype). For SNPs with a covariate, the $Covariate$ effect was also tested.

3.5. MULTIPLE TESTING CORRECTIONS

For each SNP, a linear mixed model was performed for each outcome assessment, resulting in multiple tests for the $Time * Genotype$, $Genotype$, $Time$, and $Covariate$ effects. Performing multiple tests increases the probability of making a false association (Type I error) across the set of tests (Center, 2019). Thus, a multiple testing correction procedure was used to control the false discovery rate (FDR) across the multiple tests for each SNP. The FDR is defined to be the expected proportion of discoveries (statistically significant tests) that are false. That is, among all tests where the null hypothesis was rejected and an association was detected, the FDR is defined as the proportion of those that are incorrect and there is not a true underlying association (Maarten van Iterson,

2010). The raw p-values from the multiple tests within a SNP are adjusted using the Benjamini and Hochberg (BH) procedure to control the FDR at 5% (Yoav Benjamini, 2018). Thus, any adjusted p-value less than $\alpha = 0.05$ is considered statistically significant.

4. RESULTS

After the data extraction step was complete, a total sample size of n=330 subjects were potentially available for each analysis. Among these subjects, the average age was 42 and average education years was 14. There were 72.1% males and 27.9% females. For race, the data consisted of 78.5% Caucasian, 9.5% African American/African, and 12% other races. Most subjects (73.6%) had a GCS score of 15 upon arrival to the emergency department, while 22.7% had a score of 14 and 3.7% had a score of 13. The majority of subjects (67.8%) had an Injury Severity Score less than or equal to 15 and most (57.1%) did not have an abnormal CT scan.

Not all subjects were included in all analyses due to incomplete genotype and/or outcome data. Sample sizes and summary statistics for each SNP (overall and by genotype) are provided in the demographic analysis results in the Supplemental Data (Appendix) along with p-values corresponding to the tests for differences between genotypes. Only three SNPs exhibited a statistically significant difference ($p<0.05$) on a demographic/clinical variable between genotype groups. Age was significantly different between the variant allele carriers and non-carriers for APOE and rs17759659 (BCL2

gene), while arrival GCS differed between the genotype groups for rs6311(HTR2A gene). These variables were included as covariates in the linear mixed models for those SNPs.

All SNPs in Table 1 met the HWE assumption with the exception of rs1800497 in the ANKK1 gene (p=0.044). However, this SNP was included in the analysis to compare against the results in (John K. Yue, 2015). After checking the linear mixed model assumptions, it was determined that the TMT (B/A ratio) needed a natural logarithmic transformation. Assumptions were met for this outcome after applying the transformation. The model assumptions for all other outcome assessments were met and no extreme outliers were identified.

After running the linear mixed models, the FDR adjusted p-values using the BH method were first examined (controlling the FDR at 5%) for the $Time * Genotype$ and $Genotype$ effects from model (1). No outcomes for any SNP had significant results for the $Time * Genotype$ interaction. This indicates that there were no significant associations between the genotypes tested and the rate of change in mTBI outcomes over time. One SNP, rs11604671 (in the ANNK1 gene), had significant results for the $Genotype$ effect on some outcomes. Results for all SNPs, outcomes, and model effects are available in the Supplemental Data (Appendix).

The rs11604671 genotype was significant for three outcomes, WAIS Sum of Scale, WAIS Percentile, and WAIS Composite. Table 2 shows the averages, standard deviations, and total number of participants per genotype group of the rs11604671 SNP for all three WAIS measures. Summaries are provided overall and at the six- and twelve-month post-injury time points.

Table 2. Significant WAIS results for G allele absence/presence in rs11604671.

WAIS Sum Scaled	G allele absent	G allele present	P-value of genotype
Sample Size	43	190	0.033
Mean (SD)	22.33 (5.47)	19.84 (6.10)	
	6 months (n=31)	6 months (n=157)	
Mean (SD)	22.45 (5.58)	19.75 (5.74)	
	12 months (n=21)	12 months (n=99)	
Mean (SD)	22.14 (5.42)	19.99 (5.58)	
WAIS Percentile	G allele absent	G allele present	P-value of genotype
Sample Size	42	188	0.033
Mean (SD)	61.50 (26.74)	48.06 (29.81)	
	6 months (n=31)	6 months (n=157)	
Mean (SD)	62.35 (25.82)	47.21 (28.64)	
	12 months (n=21)	12 months (n=99)	
Mean (SD)	60.24 (28.64)	49.41 (31.67)	
WAIS Composite	G allele absent	G allele present	P-value of genotype
Sample Size	42	188	0.033
Mean (SD)	106.37 (14.97)	99.64 (16.40)	
	6 months (n=31)	6 months (n=157)	
Mean (SD)	106.22 (15.44)	99.22 (15.44)	
	12 months (n=21)	12 months (n=99)	
Mean (SD)	106 (14.84)	100.29 (17.88)	

To visualize the difference between the G allele presence and absence in the rs11604671 SNP, side-by-side boxplots were created for WAIS Sum Scaled, WAIS Percentile, and WAIS Composite (Figure 2). These results show lower average (and median) WAIS scores across the board when the G allele is present for the rs11604671 SNP (in the ANKK1 gene). Higher scores indicate better cognitive ability post injury in adults. This indicates that the presence of the variant G allele is significantly associated with poorer WAIS series outcomes after mTBI.

Six different SNPs (rs1800497, rs4938016, APOE, rs17759659, rs6265, and rs1799971) exhibited a significant time effect for at least one outcome. There were

significant decreases over time on BSI in all six SNPs. The APOE and rs17759659 SNPs also showed significant increases over time on GOSE and SWLS, while rs4938016 showed a significant increase on GOSE only. In all cases, the direction of the change was indicative of improvement over time. The APOE and rs17759659 SNPs also showed significant age effects on GOSE and WAIS outcomes. Older individuals had decreased performance on these assessments. Although not the primary tests of interest, these results highlight the ability of the linear mixed model to yield results that are consistent with trends that would be expected based on previous literature.

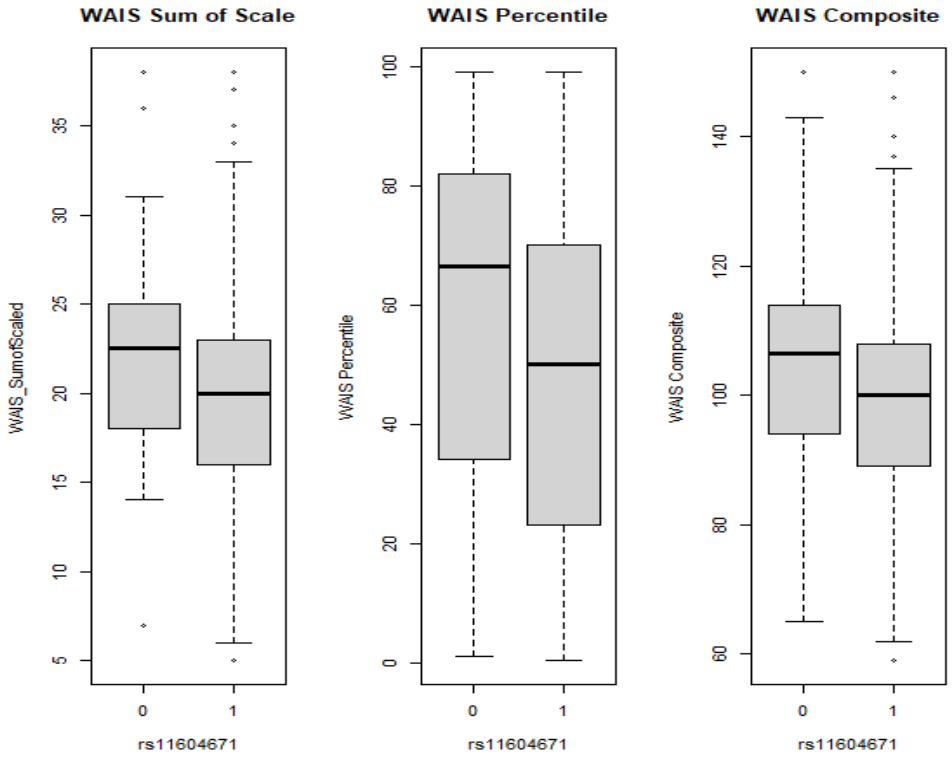

Figure 2. Side-by Side boxplots for WAIS Sum of Scale (left), Percentile (middle), and Composite (right) by G allele presence (group 1)/absence (group 0) for rs11604671.

5. CONCLUSION

5.1. DISCUSSION AND CONLUSIONS

This study evaluated the relationship between 10 SNPs on 7 mTBI outcomes at 2-3 post-injury time points using a robust modeling framework (Figure 1) that employed a linear mixed model approach. Little to no studies have completed modeling on multiple time points for mTBI outcomes for several SNPs. This study showed that it is possible to use linear mixed models to identify genotypes associated with differences in a patient's recovery over time. Although no significant time by genotype effects were found, the model is a promising starting point for mTBI clinical longitudinal data. Since the time points investigated for most outcomes in this study indicate prolonged recovery at 6- and 12-months post-injury, it is possible that most patients will not have large changes so long after a mild injury. Investigating changes from an earlier time point (1-3 months post-injury) to the 6-month time point may be more informative.

When testing for overall genotype effects with the mixed model, one SNP (rs11604671) that is located in the ANKK1 gene showed significant associations with certain mTBI outcomes (WAIS series). In this case, the variant allele carrier exhibited poorer outcomes. Previous studies have also found associations between SNPs located in the ANKK1 gene (which is involved in dopamine transmission) and TBI certain outcomes (John K. Yue, 2015), (Yue JK W. E.-A., 2017). These findings suggest that further investigations on SNPs located in the ANKK1 gene provide a promising path forward in better understanding genetic associations with mTBI outcomes.

5.2. LIMITATIONS AND FUTURE WORK

A limitation of the study was that the study sample consisted of mostly Caucasian males, so it is difficult to draw conclusions to a broader diverse population. It should also be noted that due to the observational nature of the study, any significant associations are not considered causal. Although the demographic/clinical analysis attempts to identify and address any potential imbalance between genotypes on key variables, there could be additional confounding variables that were not considered. A further limitation is the relatively small sample size and the large number of missing data for the twelve-month follow up outcome measures. Although the linear mixed model does utilize all the available data and works well when data are missing at random, using a larger data set with more complete data and more diverse patients could improve the generalizability of the conclusions. Additionally, this study was limited to the available SNPs and outcomes in the TRACK-TBI Pilot data. The modeling framework presented in this study could be used as a motivation for applying linear mixed models in future research on a wider range of SNPs and outcome assessments collected at strategic post-injury time points.

REFERENCES

Agtarap SD, C.-S. L., & Investigators, T.-T. (2021). Satisfaction with Life after Mild Traumatic Brain Injury: A TRACK-TBI Study . *J Neurotrauma , 38*(5), 546-554.

Alan D. Kaplan, . C.-E. (2022). Mixture Model Framework for Traumatic Brain Injury Prognosis Using Heterogeneous Clinical and Outcome Data. *IEEE Journal of Biomedical and Health Informatics , 3*(1), 1285-1296.

Center, C. U. (2019). *False Discovery Rate*. (Columbia University Irving Medical Center) Retrieved 2023, from https://www.publichealth.columbia.edu/research/population-health-methods/false-discovery-rate

Deanna Schreiber-Gregory, H. M. (2018). Logistic and Linear Regression Assumptions: Violation Recognition and Control. *SESUG, WUSS, SCSUG, PharmaSUG, MWSUG.*

Douglas Bates, M. M. (2015). Fitting Linear Mixed-Effects Models Using lme4 . *Journal of Statistical Software* .

Ethan A. Winkler, J. K. (2017). COMT Val158Met polymorphism is associated with post-traumatic stress disorder and functional outcome following mild traumatic brain injury . *J Clin Neurosci , 5*, 109-116.

G. L. Gadbury, C. S. (2003). Modern statistical methods for handling missing repeated measurements in obesity trial data:beyond LOCF. *Obes Rev.*, 175-184.

Giegling I, H. A. (2006). Anger- and aggression-related traits are associated with polymorphisms in the 5-HT-2A gene. *J Affect Disord, 96*(1-2), 75-81.

Gomaa Said Mohamed Abdelhamid, M. G.-B. (2021). Assessing Cognitive Abilities Using the WAIS-IV: An Item. *Environmental Research and Public Health, 18*.

Health, N. I. (n.d.). *Federal Interagency Traumatic Brain Injury Research Informatics System*. (National Institutes of Health Center for Information Technology) Retrieved 2023, from https://fitbir.nih.gov/

Hilary Bertisch, G. S., & Investigators, T. R.-T. (2019). Rehabilitation Trajectories and Outcomes in Individuals With Mild Traumatic Brain Injury and Psychiatric Histories: A TRACK-TBI Pilot Study . *J Head Trauma Rehabil. , 34*(1), 36-44.

Injury, T. R. (2018). Standard Oporating Procedures for Outcome Assessment. *Outcome Assessment SOP, 10*.

Jain, S., & Iverson, L. M. (2022). Glasgow Coma Scale. *StatPearls Publishing*.

Jan Graffelman, C. C. (2023). *Package 'HardyWeinberg'.* Retrieved from The Comprehensive R Archive Network: https://cran.r-project.org/web/packages/HardyWeinberg/HardyWeinberg.pdf

John K. Yue, A. M. (2015). Association of common genetic variant within ANKK1 with six-month cognitive performance after traumatic brain injury . *Nuerogenetics , 16*(3), 169-180.

Kreitzer NP, H. K. (2019). A Comparison of Satisfaction With Life and the Glasgow Outcome Scale-Extended After Traumatic Brain Injury: An Analysis of the TRACK-TBI Pilot Study . *J Head Trauma Rehabil , 34*(3), E10-E17.

Luke, S. G. (2016). Evaluating significance in linear mixed-effects models in R. *Bahavior Research Methods*, 1494-1502.

Maarten van Iterson, J. M. (2010). Filtering, FDR and power . *BMC Bioinformatics , 11*(450).

McAllister TW, F. L. (2008). Single nucleotide polymorphisms in ANKK1 and the dopamine D2 receptor gene affect cognitive outcome shortly after traumatic brain injury: a replication and extension study . *Brain Inj , 22*(9), 705-714.

McAllister TW, T. A. (2012). Polymorphisms in the brain-derived neurotrophic factor gene influence memory and processing speed one month after brain injury . *J Neurotrauma , 29*(6), 1111-1118.

McHugh, M. L. (2013). The Chi-square test of independence . *Biochemia Medica , 23*(2), 143-149.

Menon DK, S. K., & Health, D. a. (2010). Position statement: definition of traumatic brain injury . *Arch Phys Med Rehabil , 91*(11), 1637-1640.

Merport A, R. C. (2012). Does the Brief Symptom Inventory-18 Case Rule Apply inAdult Survivors of Childhood Cancer?: Comparison With theSymptom Checklist-90 . *J Pediatr Psychol , 37*(6), 650-659.

Moreno, J. G. (2013). The mid p-value in exact tests for Hardy-Weinberg equilibrium. *Statistical Applications in Genetics and Molecular Biology*, 433-448.

Nahhas, R. W. (2023). Checking the constant variance assumption. In *Introduction to Regression Methods for Public Health Using R.* CRC Press.

Nicole Zangrilli Hoh, . K. (2010). BCL2 Genotypes: Functional and Neurobehavioral Outcomes after Severe Traumatic Brain Injury . *J Neurotrauma. , 27*(8), 1413-1427.

Nikita Abramovs, A. B. (2020). Hardy-Weinberg Equilibrium in the Large Scale Genomic Sequencing Era. *Front. Genet., 11*.

Recklitis CJ, R. P. (2007). Screening childhood cancer survivors with the brief symptom inventory-18: classification agreement with the symptom checklist-90-revised . *Psychooncology , 16*(5), 429-436.

Routledge, R. (2005). Fisher's Exact Test. In *Encyclopedia of Biostatistics.*

Salthouse, T. A. (2011). What cognitive abilities are involved in trail-makingperformance? . *Intelligence , 39*(4), 222-232.

SÁNCHEZ-CUBILLO, I. P.-R.-S.-L. (2009). Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society, 15*(3), 438-450.

Sarchiapone M, C. V. (2008). Association of polymorphism (Val66Met) of brain-derived neurotrophic factor with suicide attempts in depressed patients . *Neuropsychobiology , 57*(3), 139-145.

Shirley Ryan. (2019, June 1). *Shirley Ryan Ability Lab*. (Ability Lab) Retrieved 2023, from https://www.sralab.org/rehabilitation-measures/glasgow-outcome-scale-extended#:~:text=Key%20Descriptions%208%20levels%20are%20in,the%20scale%3A%20A%29%20Minimum%20Score%20%3D%201

Sven Hilbert, M. S. (2019). ANALYZING LONGITUDINAL INTERVENTION STUDIES WITH LINEAR MIXED MODELS. *TPM*.

The Washington Center for Cognitive Therapy. (n.d.). *Wechsler Adult Intelligence Scale-IV*. (The Washington Center for Cognitive Therapy) Retrieved 2023, from http://washingtoncenterforcognitivetherapy.com/wp-content/uploads/2015/01/greenwood_description-wais-1.pdf

Traci W. Olivier, M. C. (2013). WA I S-IV Index and Full Scale Intelligence Quotient Score Differences between Standard and Prorated Scoring Methods . *Archives of Assessment Psychology , 3*(1), 57-77.

University of California, S. F. (2014). *TRACK-TBI*. (The Regents of the University of California) Retrieved 2023, from https://tracktbi.ucsf.edu/

Wilson L, B. K. (2021). A Manual for the Glasgow Outcome Scale-Extended Interview . *J Neurotrauma , 38*(17), 2435-2446.

Winkler EA, Y. J. (2016). COMT Val158Met polymorphism is associated with nonverbal cognition following mild traumatic brain injury . *Neurogenetics , 17*(1), 31-41.

Yamal JM, H. H. (2019). Glasgow Outcome Scale Measures and Impact on Analysis and Results of a Randomized Clinical Trial of Severe Traumatic Brain Injury . *J Neurotrauma , 36*(17), 2484-2492.

Yoav Benjamini, Y. H. (2018). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Associationg. Series B (Methodological)*, 289-300.

Yue JK, R. C.-A. (2017). Apolipoprotein E epsilon 4 (APOE-ε4) genotype is associated with decreased 6- month verbal memory performance after mild traumatic brain injury . *Brain Behav , 7*(9).

Yue JK, V. M., & Investigators, T.-T. (2013). Transforming research and clinical knowledge in traumatic brain injury pilot: multicenter implementation of the common data elements for traumatic brain injury. *J Neurotrauma, 30*(22), 1831-1844.

Yue JK, W. E.-A. (2017). DRD2 C957T polymorphism is associated with improved 6-month verbal learning following traumatic brain injury . *Neurogenetics , 18*(1), 29-38.

SECTION

2. CONCLUSION

This thesis was motivated by the potential of utilizing linear mixed models to support a genetic association analysis with longitudinal data collected from mild traumatic brain injury (mTBI) outcomes. To accomplish this, a multi-step modeling framework was employed to ensure a robust analysis. First, demographic and clinical variables were evaluated for differences between genotype groups using *t*-tests and chi-squared tests to determine if any variable should be considered a covariate in the linear mixed model. It was determined that age and GCS score influenced genotypes for three of the candidate single nucleotide polymorphisms (SNPs) being analyzed. The assumption of Hardy Weinberg equilibrium (HWE) was checked for each SNP and linearity, constant variance, and normality were assessed for the linear mixed models. It was determined that one SNP should be removed due to violating HWE and one outcome, the trail making test (TMT) ratio (B/A), should have a natural logarithmic transformation applied to meet the model assumptions. Linear mixed models with multiple testing corrections for the tests of interest were then run for each SNP and outcome to determine the significance of genotypic associations with longitudinal data on mTBI outcomes.

For this study, the focus of the linear mixed modeling was on testing the $Genotype$ and $Genotype * Time$ interaction effects to determine the association between the outcome and the genotype. No interaction terms were significant, indicating that no significant differences between genotypes were found with respect to the rate of recovery on different outcomes. However, one SNP (rs11604671), located in the ANKK1

gene, was found to have a significant association with all three-assessment metrics from the Weschler Adult Intelligence Scale (WAIS) fourth edition Processing Speed Index (PSI), with the variant allele carrier showing poorer outcomes. These results suggest that not all SNPs have the same effects on outcome measurements of an mTBI and not all outcomes have the same importance for clinical studies. It was shown that the WAIS series for PSI could have potential to be more impactful when looking at specific genetic variants located in the ANKK1 gene.

Though not the primary tests of interest, the linear mixed model also allows testing for a significant time and covariate effect (when a covariate is present). There are six SNPs with at least one significant time effect. GOSE and SWLS increase over time and BSI decreases over time, all moving in the preferred direction showing improvement over time. Two SNPs whose models included the age covariate showed significance for GOSE, WAIS sum of scaled, and WAIS composite. An additional outcome, WAIS percentile, had a significant age effect for rs17759659. In all cases where age was significant, older individuals had poorer performance. Although this was not the focus of the study, these significant results highlight that the linear mixed models yield results that are consistent with previous literature (outcomes improve over time and are worse for older individuals). The p-values and coefficients for all time and covariate effects can be found in the Appendix.

The key findings in the paper section of this thesis demonstrate the possible benefits for using a linear mixed model approach to analyze mTBI longitudinal data. The proposed approach shows that linear mixed models could help future research for genetic evaluation on outcome measurements collected over time and have the potential to detect

differing rates of recovery between genotypes. Although the results only showed a small amount of significance, future research can explore the implementation of linear mixed modeling on larger datasets with fewer missing values at earlier time points post-injury. Furthermore, using data collected on mTBIs will help the performance of neurological studies when comparing genotypes in other diseases, leading to relationships with TBIs and other diseases.

APPENDIX

The demographic analysis results for each SNP are provided in Tables A.1-A.10 below. For each table, results are shown for all individuals in the study sample (overall) as well as by genotype group (variant allele absent vs. present). Means and standard deviations are given for quantitative variables (age and education years). Counts and percentages are given for all other variables, which are categorical. The p-value is also provided for testing differences between the genotype groups. The sample sizes for each group are provided in the column heading. Note that there are differing numbers of missing observations for each variable, so the sample sizes for each variable will differ slightly from the overall sample size in each table. There was only a small amount of missingness, with at most 7 missing observations for any individual variable and SNP.

The linear mixed model results for each SNP are provided in Figures A.1-A.3 below. Figures A.1 and A.2 provide results for 3 SNPs and Figure A.3 gives results for 4 SNPS. Results for each of the 10 SNPs include coefficient (Coef) estimates with their standard errors (SE), the original (raw) p-value, and the adjusted FDR p-value (adj) for the Time*Genotype (interaction), Genotype, Time, and Covariate (if one was included in the model) effects for the linear mixed model of each of the 7 outcomes. The coefficients and standard errors were also provided for the intercepts of each model. The sections are colored by the model effect: light yellow for interaction (Genotype*Time), blue for Genotype, green for Time, and orange for Covariate. The bold yellow highlighted values represent all significant effects where the FDR-corrected p-value is <0.05.

Table A.1. Demographic analysis results for rs1800497.

	rs1800497			
	Overall ($n = 329$)	T Absent (n_0 =182)	T Present (n_1 =147)	p-value
Age (years)				0.1586
Mean, SD	42.41, 6.34	41.13, 17.36	43.69, 15.32	
Sex				0.2767
Male	234 (72.00%)	124 (69.27%)	110 (75.34%)	
Female	91 (28.00%)	55 (30.73%)	36 (24.66%)	
Education (years)				0.6896
Mean, SD	13.97, 2.84	14.03, 2.89	13.90, 2.79	
GCS				0.9294
13	12 (3.69%)	7 (3.93%)	5 (3.40%)	
14	73 (22.46%)	41 (23.03%)	32 (21.77%)	
15	240 (73.85%)	130 (73.03%)	110 (74.83%)	
ISS Score				0.8336
≤15	221 (67.79%)	122 (67.03%)	99 (68.75%)	
>15	105 (32.21%)	60 (32.96%)	45 (31.25%)	
Race / Ethnicity				0.9024
White	255 (78.46%)	142 (79.33%)	113 (77.40%)	
Black / African American	31 (9.54%)	16 (8.94%)	15 (10.27%)	
Other	39 (12.00%)	21 (11.73%)	18 (12.33%)	
CT				0.1804
Abnormal	139 (42.77%)	83 (46.37%)	56 (38.36%)	
Normal	186 (57.06%)	96 (53.63%)	90 (61.64%)	

Table A.2. Demographic analysis results for rs4938016.

	rs4938016			
	Overall (n =232)	C Absent (n_0 = 100)	C Present (n_1 =132)	p-value
Age (years)				0.8035
Mean, SD	43.09, 17.24	42.77, 16.75	43.34, 17.72	
Sex				0.3822
Male	163 (71.18%)	67 (67.68%)	96 (73.85%)	
Female	66 (28.82%)	32 (32.32%)	34 (26.15%)	
Education (years)				0.4842
Mean, SD	13.92, 2.89	14.07, 3.00	13.8, 2.77	
GCS				0.2516
13	8 (3.49%)	1 (1.01%)	7 (5.38%)	
14	49 (21.39%)	22 (22.22%)	27 (20.77%)	
15	172 (75.11%)	76 (76.77%)	96 (73.85%)	
ISS Score				0.6731
≤15	151 (65.65%)	67 (67.68%)	84 (64.12%)	
>15	79 (34.35%)	32 (32.32%)	47 (35.88%)	
Race / Ethnicity				0.0949
White	183 (79.91%)	79 (79.80%)	104 (80.00%)	
Black / African American	20 (8.73%)	5 (5.05%)	15 (11.54%)	
Other	26 (11.35%)	15 (15.15%)	11 (8.46%)	
CT				0.1521
Abnormal	98 (42.61%)	48 (48.48%)	50 (38.17%)	
Normal	132 (57.39%)	51 (51.52%)	81 (61.83%)	

Table A.3. Demographic analysis results for rs11604671.

rs11604671				
	Overall (n =233)	G Absent (n_0 =43)	G Present (n_1 =190)	p-value
Age (years)				0.7908
Mean, SD	43.09, 17.39	43.74, 17.58	42.94, 17.20	
Sex				1
Male	164 (71.30%)	30 (71.43%)	134 (71.28%)	
Female	66 (28.70%)	12 (28.57%)	54 (28.72%)	
Education (years)				0.1044
Mean, SD	13.93, 2.89	14.60, 2.93	13.77, 2.84	
GCS				0.6956
13	8 (3.48%)	1 (2.38%)	7 (3.72%)	
14	49 (21.30%)	11 (26.19%)	38 (20.21%)	
15	173 (75.22%)	30 (71.43%)	143 (76.06%)	
ISS Score				0.7771
≤ 15	152 (65.80%)	27 (62.79%)	125 (66.49%)	
> 15	79 (34.20%)	16 (37.21%)	63 (33.53%)	
Race / Ethnicity				0.9516
White	184 (80.00%)	35 (83.33%)	149 (79.26%)	
Black / African American	20 (8.70%)	3 (7.14%)	17 (9.04%)	
Other	26 (11.30%)	4 (9.52%)	22 (11.70%)	
CT				0.5616
Abnormal	98 (42.42%)	20 (47.62%)	78 (41.27%)	
Normal	133 (57.58%)	22 (52.38%)	111 (58.73%)	

Table A.4. Demographic analysis results for APOE.

	APOE			
	Overall (n =329)	$\varepsilon 4$ Absent (n_0 =247)	$\varepsilon 4$ Present (n_1 =82)	p-value
Age (years)				0.0045
Mean, SD	42.28, 15.98	40.83, 16.67	46.59, 15.28	
Sex				0.1204
Male	234 (72.00%)	169 (69.55%)	65 (79.27%)	
Female	91 (28.00%)	74 (30.45%)	17 (20.73%)	
Education (years)				0.3616
Mean, SD	13.98, 2.83	14.06, 2.86	13.73, 2.79	
GCS				0.3427
13	12 (3.69%)	11 (4.53%)	1 (1.22%)	
14	74 (22.77%)	57 (23.46%)	17 (20.73%)	
15	239 (73.54%)	175 (72.02%)	64 (78.05%)	
ISS Score				0.7301
≤ 15	221 (67.79%)	168 (68.02%)	53 (67.09%)	
> 15	105 (32.21%)	79 (31.98%)	26 (32.91%)	
Race / Ethnicity				0.8958
White	255 (78.46%)	189 (77.78%)	66 (80.49%)	
Black / African American	31 (9.54%)	23 (9.47%)	8 (9.76%)	
Other	39 (12.00%)	31 (12.76%)	8 (9.76%)	
CT				0.767
Abnormal	105 (32.21%)	79 (31.98%)	26 (32.91%)	
Normal	221 (67.79%)	168 (68.02%)	53 (67.09%)	

Table A.5. Demographic analysis results for rs17759659.

	rs17759659			
	Overall (n =231)	G Absent (n_0 =113)	G Present (n_1 =118)	p-value
Age (years)				0.0207
Mean, SD	43, 17.18	40.32, 17.41	45.63, 16.95	
Sex				0.5747
Male	159 (70.04%)	81 (72.97%)	79 (68.7%)	
Female	68 (29.96%)	30 (27.03%)	36 (31.3%)	
Education (years)				
Mean, SD	13.92, 2.87	13.72, 3.04	14.17, 2.69	
GCS				0.5586
13	8 (3.52%)	5 (4.51%)	3 (2.61%)	
14	49 (21.59%)	26 (23.42%)	23 (20%)	
15	171 (75.33%)	80 (72.07%)	89 (77.39%)	
ISS Score				0.954
≤15	149 (65.64%)	72 (65.45%)	77 (65.81%)	
>15	78 (34.36%)	38 (34.55%)	40 (34.19%)	
Race / Ethnicity				0.7581
White	181 (79.74%)	86 (77.48%)	94 (81.74%)	
Black / African American	22 (9.69%)	11 (9.91%)	9 (7.83%)	
Other	24 (10.57%)	14 (12.62%)	12 (10.43%)	
CT				0.3551
Abnormal	96 (41.92%)	43 (38.39%)	53 (45.30%)	
Normal	133 (58.08%)	69 (61.61%)	64 (54.70%)	

Table A.6. Demographic analysis results for rs6265.

rs6265				
	Overall (n =231)	A Absent (n_0 =148)	A Present (n_1 =83)	p-value
Age (years)				0.7389
Mean, SD	43, 17.49	42.71, 16.88	43.53, 18.10	
Sex				0.7481
Male	162 (71.05%)	106 (72.11%)	56 (69.14%)	
Female	66 (28.95%)	41 (27.89%)	25 (30.86%)	
Education (years)				0.3785
Mean, SD	13.92, 2.81	14.04, 3.01	13.7, 2.61	
GCS				0.7594
13	8 (3.51%)	6 (4.08%)	2 (2.47%)	
14	49 (21.49%)	33 (22.45%)	16 (19.75%)	
15	171 (75%)	108 (73.47%)	63 (77.78%)	
ISS Score				0.947
≤15	151 (65.94%)	97 (66.44%)	54 (65.06%)	
>15	78 (34.06%)	49 (33.56%)	29 (34.94%)	
Race / Ethnicity				0.4929
White	182 (79.82%)	117 (79.59%)	65 (80.25%)	
Black / African American	20 (8.77%)	15 (10.2%)	5 (6.17%)	
Other	26 (11.40%)	15 (10.2%)	11 (13.58%)	
CT				0.3524
Abnormal	97 (42.36%)	58 (39.73%)	39 (46.99%)	
Normal	132 (57.64%)	88 (60.27%)	44 (53.01%)	

Table A.7. Demographic analysis results for rs4680.

rs4680				
	Overall (n =327)	Met Absent (n_0 =94)	Met Present (n_1 =233)	p-value
Age (years)				0.7328
Mean, SD	42.38, 16.12	42.71, 15.17	42.04, 17.07	
Sex				0.755
Male	233 (72.14%)	68 (73.91%)	165 (71.43%)	
Female	90 (27.86%)	24 (26.08%)	66 (28.57%)	
Education (years)				0.29
Mean, SD	14.04, 2.95	14.24, 3.16	13.84, 2.74	
GCS				0.58
13	12 (3.69%)	5 (5.43%)	7 (3.03%)	
14	73 (22.46%)	20 (21.74%)	53 (22.94%)	
15	240 (73.85%)	67 (72.83%)	171 (74.03%)	
ISS Score				0.9743
≤ 15	219 (67.80%)	63 (68.48%)	156 (67.53%)	
> 15	104 (32.20%)	29 (31.52%)	75 (32.47%)	
Race / Ethnicity				0.4943
White	253 (78.33%)	68 (73.91%)	185 (80.09%)	
Black / African American	31 (9.60%)	11 (11.96%)	20 (8.66%)	
Other	39 (12.07%)	13 (14.13%)	26 (11.26%)	
CT				0.9055
Abnormal	139 (43.03%)	41 (44.09%)	98 (42.61%)	
Normal	184 (56.97%)	52 (55.91%)	132 (57.39%)	

Table A.8. Demographic analysis results for rs6277.

rs6277				
	Overall (n =233)	T Absent (n_0 =88)	T Present (n_1 =145)	p-value
Age (years)				0.9645
Mean, SD	43.09, 17.11	43.02, 16.51	43.13, 17.71	
Sex				0.2976
Male	164 (71.30%)	66 (75.86%)	98 (68.53%)	
Female	66 (28.70%)	21 (24.14%)	45 (31.47%)	
Education (years)				0.6924
Mean, SD	13.93, 2.90	13.83, 3.02	13.99, 2.7	
GCS				0.3842
13	8 (3.48%)	4 (4.60%)	4 (2.80%)	
14	49 (21.30%)	22 (25.29%)	27 (18.88%)	
15	173 (75.22%)	61 (70.11%)	112 (78.32%)	
ISS Score				1
≤15	152 (65.80%)	58 (65.91%)	94 (65.73%)	
>15	79 (34.20%)	30 (34.09%)	49 (34.27%)	
Race / Ethnicity				0.7971
White	184 (80.00%)	68 (78.16%)	116 (81.12%)	
Black / African American	20 (8.70%)	9 (10.34%)	11 (7.69%)	
Other	26 (11.30%)	10 (11.49%)	16 (11.19%)	
CT				0.3489
Abnormal	98 (42.43%)	33 (37.93%)	65 (45.14%)	
Normal	133 (57.58%)	54 (62.07%)	79 (54.86%)	

Table A.9. Demographic analysis results for rs6311.

	rs6311			
	Overall (n =230)	T Absent (n_0 =79)	T Present (n_1 =151)	p-value
Age (years)				0.237
Mean, SD	43.05, 16.98	41.28, 15.99	44.04, 17.97	
Sex				0.5632
Male	164 (72.25%)	52 (68.42%)	109 (73.65%)	
Female	63 (27.75%)	24 (31.58%)	39 (26.35%)	
Education (years)				0.2072
Mean, SD	13.93, 2.87	13.84, 2.81	13.95, 2.93	
GCS				0.0145
13	7 (3.08%)	6 (7.89%)	1 (0.68%)	
14	48 (21.15%)	14 (18.42%)	34 (22.97%)	
15	172 (75.77%)	56 (73.68%)	113 (76.35%)	
ISS Score				0.5338
≤15	150 (65.79%)	47 (61.04%)	100 (67.57%)	
>15	78 (34.21%)	30 (38.96%)	48 (32.43%)	
Race / Ethnicity				0.1977
White	182 (80.18%)	61 (79.22%)	119 (80.41%)	
Black / African American	20 (8.81%)	10 (12.99%)	10 (6.76%)	
Other	25 (11.01%)	6 (7.79%)	19 (12.84%)	
CT				0.3615
Abnormal	96 (42.10%)	37 (46.84%)	59 (39.60%)	
Normal	132 (57.89%)	42 (53.16%)	90 (60.40%)	

Table A.10. Demographic analysis results for rs1799971.

	rs1799971			
	Overall (n =330)	G Absent (n_0 =246)	G Present (n_1 =84)	p-value
Age (years)				0.6915
Mean, SD	42.52, 16.73	42.09, 16.27	42.95, 17.18	
Sex				0.6363
Male	235 (72.09%)	176 (72.73%)	59 (70.24%)	
Female	91 (27.91%)	66 (27.27%)	25 (29.76%)	
Education (years)				0.4153
Mean, SD	13.96, 2.84	13.90, 2.89	14.13, 2.78	
GCS				1
13	12 (3.68%)	9 (3.72%)	3 (3.57%)	
14	74 (22.70%)	55 (22.73%)	19 (22.62%)	
15	240 (73.62%)	178 (73.55%)	62 (73.81%)	
ISS Score				0.7372
≤15	221 (67.79%)	166 (68.03%)	55 (67.07%)	
>15	105 (32.21%)	78 (31.97%)	27 (32.93%)	
Race / Ethnicity				0.9225
White	256 (78.53%)	189 (77.78%)	67 (80.72%)	
Black / African American	31 (9.51%)	24 (9.88%)	7 (8.43%)	
Other	39 (11.96%)	30 (12.35%)	9 (10.84%)	
CT				0.6568
Abnormal	140 (42.94%)	107 (43.85%)	33 (40.24%)	
Normal	186 (57.06%)	137 (56.15%)	49 (59.76%)	

	rs1800497											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	6.68 (0.13)	0.04 (0.02)	0.016	0.329	-0.32 (0.19)	0.093	0.551	0.01 (0.01)	0.269	0.628			
TMT_log	0.38 (0.03)	0.01 (0.01)	0.17	0.158	-0.06 (0.05)	0.19	0.551	0.00 (0.00)	0.56	0.98			
SWLSTS	18.91 (1.30)	0.12 (0.20)	0.565	0.809	-0.15 (1.92)	0.939	0.915	0.22 (0.13)	0.314	0.35			
WAIS_SumofScaled	20.27 (0.77)	0.03 (0.10)	0.734	0.809	-0.63 (1.15)	585	0.833	-0.02 (0.06)	0.743	0.991			
WAIS_Percentile	49.80 (4.01)	0.18 (0.54)	0.737	0.988	-4.27 (0.97)	0.475	0.915	0.00 (0.35)	0.991	0.991			
WAIS_PSI	100.43 (2.09)	0.16 (0.27)	0.56	0.809	-2.12 (3.11)	0.495	0.833	-0.02 (0.18)	0.889	0.991			
BSI18GSIScoreT	59.18 (1.62)	0.15 (0.24)	0.52	0.329	-0.67 (2.41)	0.782	0.551	-0.59 (0.16)	0.001	0.007			
	rs4938016											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	6.32 (0.16)	-0.02 (0.02)	0.218	0.417	0.32 (0.21)	0.128	0.179	0.05 (0.01)	0.001	0.007			
TMT_log	0.31 (0.04)	-0.00 (0.00)	0.493	0.553	0.09 (0.05)	0.059	0.103	0.00 (0.00)	0.557	0.65			
SWLSTS	20.33 (1.48)	0.24 (0.20)	0.238	0.417	-2.58 (1.94)	0.185	0.216	0.16 (0.15)	0.314	0.612			
WAIS_SumofScaled	21.47 (0.88)	0.13 (0.10)	0.187	0.417	-2.44 (1.17)	0.036	0.103	-0.07 (0.07)	0.376	0.612			
WAIS_Percentile	56.41 (4.56)	0.87 (0.54)	0.107	0.417	-14.40 (6.01)	0.017	0.103	-0.32 (0.41)	0.437	0.612			
WAIS_PSI	103.27 (2.39)	0.26 (0.27)	0.326	0.456	-6.20 (3.15)	0.05	0.103	-0.06 (0.20)	0.75	0.75			
BSI18GSIScoreT	57.70 (1.85)	-0.14 (0.24)	0.553	0.553	2.04 (2.44)	0.404	0.404	-0.49 (0.18)	0.008	0.028			
	rs11604671											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	6.49 (0.24)	0.00 (0.02)	0.845	0.845	0.01 (0.27)	0.956	0.856	0.03 (0.02)	0.142	0.433			
TMT_log	0.33 (0.06)	0.00 (0.01)	0.711	0.83	0.04 (0.06)	0.555	0.648	-0.00 (0.01)	0.771	0.778			
SWLSTS	22.30 (2.30)	0.44 (0.26)	0.093	0.317	-4.14 (2.53)	0.104	0.182	-0.07 (0.23)	0.778	0.778			
WAIS_SumofScaled	23.23 (1.39)	0.17 (0.13)	0.189	0.331	-3.78 (1.52)	0.014	0.033	-0.13 (0.11)	0.26	0.433			
WAIS_Percentile	66.48 (7.20)	1.05 (0.70)	0.136	0.317	-21.90 (7.89)	0.006	0.033	-0.69 (0.63)	0.279	0.433			
WAIS_PSI	108.53 (3.77)	0.49 (0.34)	0.016	0.111	-10.56 (4.13)	0.011	0.033	-0.32 (0.31)	0.309	0.433			
BSI18GSIScoreT	55.89 (2.89)	-0.12 (0.31)	0.703	0.83	3.46 (3.18)	0.277	0.388	-0.47 (0.28)	0.099	0.433			

Figure A.1. P-values and coefficients for each SNP by outcome and model effect. Part 1 – results for 3 SNPs.

	APOE Presence/Absence											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	7.32 (022)	0.03 (0.02)	0.153	0.536	0.04 (0.22)	0.839	0.839	0.02 (0.01)	0.016	0.037	-0.02 (0.00)	0.001	0.0007
TMT_log	0.30 (0.3)	-0.01 (0.01)	0.112	0.536	0.089 (0.05)	0.091	0.637	0.00 (0.00)	0.227	0.397	0.00 (0.00)	0.216	0.303
SWLSTS	19.63 (1.68)	-0.19 (0.23)	0.421	0.589	2.26 (2.24)	0.314	0.714	0.32 (0.11)	0.006	0.021	-0.03 (0.03)	0.334	0.39
WAIS_SumofScaled	22.41 (1.20)	-0.06 (0.11)	0.571	0.666	0.67 (1.32)	0.612	0.714	0.01 (0.06)	0.869	0.869	-0.06 (0.03)	0.015	0.037
WAIS_Percentile	58.52 (5.98)	-0.53 (0.62)	0.401	0.589	2.80 (6.85)	0.583	0.714	0.21 (0.30)	0.493	0.575	-0.27 (0.12)	0.03	0.053
WAIS_PSI	105.86 (3.23)	-0.25 (0.31)	0.413	0.589	2.25 (3.57)	0.528	0.714	0.10 (0.15)	0.489	0.575	-0.17 (0.07)	0.016	0.037
BSI18GSIScoreT	60.38 (2.27)	0.12 (0.28)	0.675	0.675	-2.93 (2.83)	0.303	0.714	-0.55 (0.13)	0.001	0.007	-0.02 (0.05)	0.665	0.665
	rs17759659											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	7.12 (0.26)	-0.02 (0.02)	0.357	0.648	0.19 (0.21)	0.373	0.653	0.04 (0.01)	0.001	0.004	-0.02 (0.01)	0.002	0.007
TMT_log	0.30 (0.04)	-0.01 (0.01)	0.291	0.648	0.03 (0.05)	0.537	0.686	0.00 (0.00)	0.421	0.502	0.00 (0.00)	0.079	0.111
SWLSTS	19.99 (1.84)	-0.18 (0.20)	0.37	0.648	-0.78 (1.93)	0.686	0.686	0.37 (0.14)	0.011	0.026	-0.01 (0.03)	0.645	0.645
WAIS_SumofScaled	22.48 (1.27)	-0.01 (0.10)	0.897	0.897	1.63 (1.16)	0.16	0.373	0.02 (0.07)	0.828	0.828	-0.08 (0.03)	0.003	0.007
WAIS_Percentile	57.45 (6.38)	-0.35 (0.54)	0.517	0.724	12.27 (6.00)	0.042	0.294	0.37 (0.39)	0.354	0.502	-0.37 (0.13)	0.004	0.007
WAIS_PSI	105.87 (3.44)	-0.13 (0.27)	0.639	0.746	5.02 (3.14)	0.111	0.373	0.1 (0.19)	0.43	0.502	-0.21 (0.07)	0.003	0.007
BSI18GSIScoreT	60.37 (2.49)	0.30 (0.24)	0.213	0.648	-0.72 (0.17)	0.634	0.686	-1.16 (2.44)	0.001	0.004	-0.03 (0.05)	0.577	0.645
	rs6265											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	6.60 (0.13)	0.02 (0.02)	0.333	0.963	-0.27 (0.22)	0.212	0.742	0.03 (0.01)	0.021	0.058			
TMT_log	0.35 (0.03)	-0.00 (0.01)	0.621	0.963	0.03 (0.05)	0.566	0.77	0.00 (0.00)	0.689	0.876			
SWLSTS	18.92 (1.21)	-0.01 (0.21)	0.963	0.963	0.17 (2.00)	0.93	0.93	0.29 (0.13)	0.025	0.058			
WAIS_SumofScaled	19.78 (0.72)	-0.01 (0.10)	0.926	0.963	0.96 (1.21)	0.428	0.77	0.01 (0.06)	0.876	0.876			
WAIS_Percentile	47.48 (3.72)	0.16 (0.55)	0.78	0.963	2.77 (6.28)	0.66	0.77	0.11 (0.34)	0.757	0.876			
WAIS_PSI	99.18 (1.94)	0.09 (0.27)	0.736	0.963	1.88 (3.28)	0.567	0.77	0.04 (0.17)	0.796	0.876			
BSI18GSIScoreT	59.95 (1.52)	0.25 (0.24)	0.304	0.963	-3.43 (2.49)	0.17	0.742	-0.66 (0.15)	0.001	0.007			

Figure A.2. P-values and coefficients for each SNP by outcome and model effect. Part 2 – results for 3 SNPs.

	rs4680 ValVal											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	6.61 (0.11)	0.00 (0.02)	0.893	0.893	-0.24 (0.21)	0.263	0.697	0.03 (0.01)	0.003	0.105			
TMT_log	0.33 (0.26)	-0.01 (0.01)	0.131	0.392	0.10 (0.05)	0.059	0.413	0.00 (0.00)	0.227	0.397			
SWLSTS	19.20 (1.12)	0.31 (0.22)	0.168	0.392	-1.23 (2.13)	0.564	0.697	0.19 (0.11)	0.101	0.236			
WAIS_SumofScaled	20.19 (0.66)	-0.06 (.011)	0.594	0.832	-0.68 (1.28)	0.597	0.697	0.01 (0.06)	0.886	0.886			
WAIS_Percentile	49.19 (3.45)	-0.19 (0.61)	0.751	0.876	-4.56 (6.67)	0.495	0.697	0.13 (0.31)	0.673	0.785			
WAIS_PSI	99.89 (1.80)	-0.22 (0.30)	0.461	0.807	-1.32 (3.47)	0.703	0.703	0.10 (0.15)	0.518	0.725			
BSI18GSIScoreT	58.37 (1.40)	-0.37 (0.27)	0.168	0.392	1.77 (2.68)	0.509	0.697	-0.42 (0.14)	0.002	0.105			
	rs6277											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	6.61 (1.11)	0.00 (0.02)	0.745	0.745	-0.24 (0.21)	0.901	0.901	0.03 (0.01)	0.012	0.06			
TMT_log	0.33 (0.03)	-0.01 (0.01)	0.121	0.212	0.10 (0.05)	0.022	0.095	0.00 (0.00)	0.252	0.252			
SWLSTS	19.20 (1.12)	0.31 (0.22)	0.623	0.727	-1.23 (2.13)	0.155	0.271	0.19 (0.11)	0.036	0.084			
WAIS_SumofScaled	20.19 (0.66)	-0.06 (0.11)	0.111	0.212	-0.68 (1.28)	0.048	0.112	0.01 (0.06)	0.174	0.203			
WAIS_Percentile	49.19 (3.46)	-0.19 (0.61)	0.061	0.212	-4.57 (6.67)	0.027	0.095	0.13 (0.31)	0.06	0.084			
WAIS_PSI	99.89 (1.80)	-0.22 (0.30)	0.047	0.212	-1.32 (3.47)	0.3	0.35	0.10 (0.15)	0.05	0.084			
BSI18GSIScoreT	58.37 (1.40)	-0.37 (0.27)	0.577	0.727	1.77 (2.68)	0.274	0.35	-0.42 (0.13)	0.017	0.06			
	rs6311											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	1.12 (2.58)	0.02 (0.02)	0.417	0.487	0.18 (0.22)	0.423	0.592	0.02 (0.02)	0.185	0.46	0.36 (0.18)	0.042	0.252
TMT_log	0.56 (0.27)	0.01 (0.01)	0.306	0.428	-0.06 (0.05)	0.22	0.385	-0.00 (0.00)	0.435	0.46	-0.01 (0.02)	0.558	0.651
SWLSTS	14.90 (15.53)	0.23 (0.22)	0.294	0.428	-0.55 (2.04)	0.789	0.789	0.13 (0.18)	0.46	0.46	0.30 (1.05)	0.772	0.772
WAIS_SumofScaled	2.76 (12.41)	0.17 (0.10)	0.102	0.252	-1.55 (1.22)	0.203	0.385	-0.10 (0.09)	0.252	0.46	1.24 (0.84)	0.144	0.252
WAIS_Percentile	-45.01 (60.62)	0.93 (0.57)	0.108	0.252	-8.74 (6.31)	0.167	0.385	-0.39 (0.47)	0.414	0.46	6.68 (4.12)	0.107	0.252
WAIS_PSI	52.37 (33.40)	0.52 (0.28)	0.068	0.252	-4.53 (3.29)	0.171	0.385	-0.24 (0.23)	0.304	0.46	3.40 (2.27)	0.136	0.252
BSI18GSIScoreT	88.08 (22.35)	-0.12 (0.25)	0.627	0.627	-0.75 (2.50)	0.765	0.789	-0.45 (0.20)	0.029	0.203	-1.98 (1.52)	0.193	0.27
	rs1799971											
	Intercept	Genotype x Time			Genotype			Time			Covariate		
Response	Coef (SE)	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value	Coef (SE)	raw p-value	adj p-value
GOSE	6.58 (0.11)	0.02 (0.02)	0.222	0.482	-0.14 (0.22)	0.516	0.602	0.02 (0.01)	0.017	0.06			
TMT_log	0.36 (0.03)	-0.00 (0.01)	0.805	0.805	-0.01 (0.05)	0.803	0.803	0.00 (0.00)	0.57	0.695			
SWLSTS	19.40 (1.10)	0.35 (0.22)	0.113	0.424	-1.81 (2.17)	0.407	0.584	0.17 (0.12)	0.136	0.317			
WAIS_SumofScaled	19.30 (0.66)	-0.11 (0.11)	0.309	0.482	2.64 (1.28)	0.04	0.182	0.02 (0.06)	0.695	0.695			
WAIS_Percentile	44.52 (3.44)	-0.56 (0.59)	0.344	0.482	12.87 (6.68)	0.055	0.182	0.23 (0.31)	0.464	0.695			
WAIS_PSI	97.89 (1.79)	-0.16 (0.29)	0.588	0.686	6.15 (3.48)	0.078	0.182	0.08 (0.16)	0.596	0.695			
BSI18GSIScoreT	58.19 (1.38)	-0.40 (0.26)	0.121	0.424	2.21 (2.72)	0.417	0.584	-0.41 (0.14)	0.004	0.028			

Figure A.3. P-values and coefficients for each SNP by outcome and model effect. Part 3 – results for 4 SNPs.

BIBLIOGRAPHY

Alan D. Kaplan, . C.-E. (2022). Mixture Model Framework for Traumatic Brain Injury Prognosis Using Heterogeneous Clinical and Outcome Data . *IEEE Journal of Biomedical and Health Informatics* , 1285-1296.

Centre for Research Excellence in Brain Recovery. (n.d.). *Comprehensive Trail-Making Test.* Retrieved from Moving Ahead: http://movingahead.psy.unsw.edu.au/documents/research/outcome%20measures/paediatric/Neuropsychological%20Impairment/Website%20Comprehensive%20Trail%20Making%20Test%20-%20Paediatrics.pdf

Centre for Research Excellence in Brain Recovery. (n.d.). *Wechsler Adult Intelligence Scale - IV.* Retrieved from Moving Ahead: http://movingahead.psy.unsw.edu.au/documents/research/outcome%20measures/adult/Neuropsychological%20Impairment/Website%20WAIS-IV.pdf

Christopher M. Gotwalt, E. A. (n.d.). What is REML? Why Does it Work? And How Do. Lincoln, Nebraska.

Cnossen MC, W. E., & Investigators, T.-T. (2017). Development of a Prediction Model for Post-Concussive Symptoms following Mild Traumatic Brain Injury: A TRACK-TBI Pilot Study . *J Neurotrauma* , 2396-2409.

Cran R-Project. (2021). *pbkrtest: Parametric Bootstrap, Kenward-Roger and Satterthwaite Based Methods for Test in Mixed Models*. Retrieved from Cran R-Project: https://cran.r-project.org/web/packages/pbkrtest/index.html

Cran R-Project. (2022). *Package 'HardyWeinberg'*. Retrieved from cran.r-project: https://cran.r-project.org/web/packages/HardyWeinberg/HardyWeinberg.pdf

D. Yeboah, H. N.-A. (2021). A deep learning model to predict traumatic brain injury severity and outcome from MR images . *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* , 1-6.

de Winter, J. (2013). Using the Student's t-test with extremely small sample size. *Practical Assessment, Research, and Evaluation* .

Deanna Schreiber-Gregory, H. M. (2018). Logistic and Linear Regression Assumptions: Violation Recognition and Control. *SESUG, WUSS, SCSUG, PharmaSUG, MWSUG* .

Deng H, Z. B. (2021). B-Cell Lymphoma 2 (Bcl-2) Gene Is Associated with Intracranial Hypertension after Severe Traumatic Brain Injury . *J Neurotrauma* , 291-299.

Diener, E. E. (1985). The Satisfaction with Life . *Journal of Personality Assessment*, 71-75.

Douglas Bates, M. M. (2015). Fitting Linear Mixed-Effects Models Using lme4 . *Journal of Statistical Software* .

Eric A Toschlog, J. M. (2003). The relationship of Injury Severity Score and Glasgow Coma Score to rehabilitative potential in patients suffering traumatic brain injury. *Am Surg*, 497-498.

Esther van den Wildenberg, R. W. (2007). A functional polymorphism of the mu-opioid receptor gene (OPRM1) influences cue-induced craving for alcohol in male heavy drinkers. *Alcohol Clin Exp Res*, 1-10.

G. L. Gadbury, C. S. (2003). Modern statistical methods for handling missing repeated . *Obes Rev*, 175-184.

Giegling I, H. A. (2006). Anger- and aggression-related traits are associated with polymorphisms in the 5-HT-2A gene . *J Affect Disord* , 75-81.

Geoffrey Manley (n.d.). *TRACK-TBI Pilot Study Initiative.* Retrieved from Brain and Spinal Injury Center: https://brainandspinalinjury.org/research/track-tbi-pilot-initiative/

H Houlden, R. G. (2006). Apolipoprotein E4 and traumatic brain injury. *Journal of Neurology, Neurosurgery & Psychiatry*, 1106-1107.

Hendrick, P. W. (2011). *Genetic Populations.* Sudbury, Massachusetts: Jones and Bartlett Publishers.

J Graffelman, B. S. (2016). Testing for Hardy–Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Heredity*, 558-568.

John K. Yue, A. M. (2015). Association of common genetic variant within ANKK1 with six-month cognitive performance after traumatic brain injury . *Nuerogenetics , 16*(3), 169-180.

Kreitzer NP, H. K. (2019). Comparison of Satisfaction With Life and the Glasgow Outcome Scale-Extended After Traumatic Brain Injury: An Analysis of the TRACK-TBI Pilot Study . *J Head Trauma Rehabil* , E10-E17.

Lingsma HF, Y. J., & Investigators, T.-T. (2015). Outcome prediction after mild and complicated mild traumatic brain injury: external validation of existing models and identification of new predictors using the TRACK-TBI pilot study . *J Neurotrauma* , 83-94.

Maarten van Iterson, J. M. (2010). Filtering, FDR and power . *BMC Bioinformatics* .

McAllister TW, F. L. (2008). Single nucleotide polymorphisms in ANKK1 and the dopamine D2 receptor gene affect cognitive outcome shortly after traumatic brain injury: a replication and extension study . *Brain Inj* , 705-714.

McHugh, M. L. (2013). The Chi-square test of independence . *Biochemia Medica* , 143-149.

Miranda R, R. L. (2010). Initial evidence of an association between OPRM1 and adolescent alcohol misuse . *Alcohol Clin Exp Res* , 112-122.

Muller, M. (2004). Generalized Linear Models. *Fraunhofer Institute for Industrial Mathematics*.

Nahhas, R. W. (2023). Introduction to Regression Methods for Public Health Using R. CRC Press.

National Human Genome Research Institute. (2023). *Allele*. Retrieved from Genome: https://www.genome.gov/genetics-glossary/Allele

National Library of Medicine. (n.d.). *What are single nucleotide polymorphisms (SNPs)?* Retrieved from Medline Plus Trusted Health Information for You: https://medlineplus.gov/genetics/understanding/genomicresearch/snp/

Nelson LD, R. J. (2017). Validating Multidimensional Outcome Assessment Using the TBI Common Data Elements: An Analysis of the TRACK-TBI Pilot Sample . *J Neurotrauma* , 3158-3172.

Newsom. (2019). Multilevel Regression Estimation Methods for Continuous Dependent Variables.

Nicole Zangrilli Hoh, . K. (2010). BCL2 Genotypes: Functional and Neurobehavioral Outcomes after Severe Traumatic Brain Injury . *J Neurotrauma.* , 1413-1427.

Nikita Abramovs, A. B. (2020). Hardy-Weinberg Equilibrium in the Large Scale Genomic Sequencing Er. *Front. Genet.* .

Noble, W. (2009). How does multiple testing correction work? *Nat Biotechnol*, 1135-1137.

Peña EA, S. E. (2006). Global Validation of Linear Model Assumptions . *J Am Stat Assoc* , 341.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, 688-690.

Ryan M. Smith, A. C. (2013). Multiple Regulatory Variants Modulate Expression of 5-Hydroxytryptamine 2A Receptors in Human Cortex . *Biological Psychiatry* , 546-554.

Sarchiapone M, C. V. (2008). Association of polymorphism (Val66Met) of brain-derived neurotrophic factor with suicide attempts in depressed patients . *Neuropsychobiology* , 139-145.

Shirley Ryan. (2015, September 5). *Brief Symptom Inventory.* Retrieved from Ability Lab: https://www.sralab.org/rehabilitation-measures/brief-symptom-inventory

Shirley Ryan. (2016, April 11). *Satisfaction With Life Scale*. Retrieved from Ability Lab: https://www.sralab.org/rehabilitation-measures/satisfaction-life-scale

Shirley Ryan. (2019, June 1). *Glasgow Outcome Scale - Extended*. Retrieved from Ability Lab: https://www.sralab.org/rehabilitation-measures/glasgow-outcome-scale-extended#:~:text=The%20GOS-E%20classifies%20global%20outcomes%20in%20TBI%20survivors.,purposes.%20The%20utility%20for%20individual%20assessment%20is%20limited.

Stat Trek. (n.d.). *Chi-Square Test of Homogeneity*. Retrieved from Stat Trek : https://stattrek.com/chi-square-test/homogeneity

Sven Hilbert, M. S. (2019). ANALYZING LONGITUDINAL INTERVENTION STUDIES WITH LINEAR MIXED MODELS. *TPM*.

TK, K. (2015). T test as a parametric statistic . *Korean J Anesthesiol* , 540-546.

Wiebke Bensmann, N. Z.-K. (2020). Dopamine D1, but not D2, signaling protects mental representations from distracting bottom-up influences. *Neuroimage*.

Wiley, J. F. (2020). *Multilevel Models using lmer.* Retrieved from Cran R-Project: https://cran.r-project.org/web/packages/multilevelTools/vignettes/lmer-vignette.html

Winkler EA, Y. J. (2016). COMT Val158Met polymorphism is associated with nonverbal cognition following mild traumatic brain injury . *Neurogenetics* , 31-41.

Winkler, A. M. (2011). *False Discovery Rate: Corrected & Adjusted P-values*. Retrieved from Brainder: https://brainder.org/2011/09/05/fdr-corrected-fdr-adjusted-p-values/

Yekutieli, Y. B. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics*, 1165-1188.

Yue JK, R. C.-A. (2017). Apolipoprotein E epsilon 4 (APOE-ε4) genotype is associated with decreased 6- month verbal memory performance after mild traumatic brain injury . *Brain Behav* .

Yue JK, V. M., & Investigators, T.-T. (2013). Transforming research and clinical knowledge in traumatic brain injury pilot: multicenter implementation of the common data elements for traumatic brain injury. *J Neurotrauma, 30*(22), 1831-1844.

Yue JK, W. E.-A. (2017). DRD2 C957T polymorphism is associated with improved 6-month verbal learning following traumatic brain injury . *Neurogenetics* , 29-38.

Zach. (2019, May 29). *Welch's t-test: When to Use it + Examples*. Retrieved from Statology: https://www.statology.org/welchs-t-test/#:~:text=Welch%E2%80%99s%20t-test%20Test%20statistic%3A%20%28x1%20%E2%80%93%20x2%29%20%2F,account%20the%20difference%20between%20the%20two%20standard%20deviations.

VITA

Caroline Elizabeth Claire Schott received her B.S. in Mathematics with an emphasis in Statistics from Missouri University of Science and Technology in May 2020, and joined the graduate program at Missouri University of Science and Technology in the Department Mathematics and Statistics in August 2020. She received her M.S. degree in Applied Mathematics with a Statistics Emphasis in May 2023.