STAT550 Fracture Group Project Report

Trajectory of Outcomes After A Distal Radius Fracture

Prepared for Dr. Roffey, Dr. Lefaivre and Dr. Larose

Li Zha, Denghuang Zhan, Zhipeng Zhu

Abstract

Distal radius fractures are one of the most common types of fractures that dramatically affect work productivity. Historically, studies evaluating the success of the treatment use arbitrary end-points, which are inadequate to inform surgeons. From a longitudinal study focusing on the evaluation of the patients' PRWE scores, we can further investigate how the patients' characteristics are associated with the "good outcome" at some specific time points and evaluate how the score changes over time. Specifically, by using logistic regression, we find that the "Removal of device" is negatively associated with the "good outcome" at 1 year; the early score difference is negatively associated with a "good outcome at 5 years. From the linear mixed effect modeling, we find that the score (in log-scale) on average has a 0.02-unit (in log-scale) decrement per month.

1. Introduction

Distal radius fracture, also known as wrist fracture, is a common orthopedic injury that has a dramatic effect on work productivity [1]. Traditional approaches to distal radius fractures included both surgical and non-surgical treatments. In recent years, surgical treatments of open reduction and internal fixation (ORIF) are more favored compared to other treatment options. A frequently used outcome measurement for the treatment of distal radius fracture is the Patient-rated Wrist Evaluation (PRWE) score, with a range from 0 to 100 (a higher score indicates a higher level of disability). To recognize a treatment effect expressed as the change in PRWE score, it is important to be conscious of the minimum clinically important difference (MCID) of the PRWE score. The MCID represents the smallest difference in scores, usually specified as 11.5, that would be perceived by the patient as beneficial [2]. Consequently, a score change within 11.5 difference from the baseline represents a "good outcome" of treatment effect.

Multiple studies have described outcomes after distal radius fracture treatment; however, researchers have historically used arbitrary end-points to evaluate the success or completion of the treatment. In such cases, information advising surgeons or patients about the final functional outcomes of these fractures after fixation, or the likely trajectory and timeline of that recovery is limited [3]. The expected recovery time after surgical treatment is valuable information for a patient's mental health. A longitudinal study has been conducted to study the recovery of distal radius fracture treatment. The hypothesis is that the patient pain and function will improve over time after operative treatment and clinically relevant changes will occur up to 5 years post-injury. Specifically, it is of interest to identify the important predictors to predict the "good outcome" at 1 year and 5 year, respectively, and evaluate the change of the score over time.

The main objective of this report is to provide statistical analyses on this longitudinal dataset to evaluate how the PRWE scores change over time. We start with explanatory data analysis and then further validate these results through confirmatory analysis. The logistic regression models were built to predict the "good outcome" at 1 year and 5 year, respectively. The linear mixed effect model and GEE were used to evaluate the score change over time. The details of the model selections and diagnostics can be found in Appendix.

2. Explanatory data analysis (EDA)

Firstly, data wrangling was performed to the dataset for the purpose of the following analysis. Pairwise relationships between the response variable PRWE scores and categorical explanatory variables were explored to examine the possible associations.

2.1 Data Overview

The data were collected by convenient sampling: a total of 448 patients with distal radius fracture surgically treated with open reduction and internal fixation (ORIF) at Vancouver General Hospital were enrolled between 2004 and 2015 into the orthopedic trauma prospective outcomes database. Basic demographic and diseases history information collected for each patient included patient age, sex, injury severity score (ISS), mental illness, substance abuse, etc. The outcome was evaluated by PRWE, a 15-item questionnaire designed to measure wrist pain and disability in activities of daily living with a score ranging from 0 to 100. The PRWE scores of the patients were evaluated at baseline time point (i.e., pre-injury), 3 months, 6 months, 1 year, and 5 years. Descriptions of variables were summarized in Table 1.

Table 1 Description of data

Variables	Type	Range (Levels)
Age at injury	Continuous	17 – 93
Injury Severity Score (ISS)	Continuous	0 – 27
Sex	Categorical	Male, Female
Removal of device	Categorical	Present, None
Mental Illness (Depression, Anxiety Disorder, Psychosis)	Categorical	Present, None
Cardiovascular Conditions (Dyslipidemia, Diabetes, etc(3))	Categorical	Present, None
Substance Abuse (Alcohol, Substance)	Categorical	Present, None
PRWE Scores		
Pain	Continuous	0 – 50
Function	Continuous	0 – 50
Total	Continuous	0 - 100

There are a total of 1582 observations recorded for 447 patients. Those records with all missing or zero scores, or with only 1 non-zero score are to be dropped. The missing rates are reported by periods in Table 2.

Table 2 The missing rate at each time point

Period	Baseline	3M	6M	1Y	5Y
Missing Rate	7.82%	63.57%	6.60%	24.21%	27.87%

For most of the periods, the missing rates are acceptable while the one of “3 Month” is quite high. However, by checking the histograms of scores in each period, the distribution of scores is similar

between periods. Besides, as confirmed with the client, these data are mostly missing in random. Hence, these missing values can be ignored in further analysis.

The main interest variables in the data are the scores recorded in each period. Due to the large variety of baseline scores of patients and the unbalanced outliers on a normal scale, the logarithms of score differences between other periods and the baseline will be used for the preview of the data. Figure 1 shows the box plots of the total score (original scale), indicating that the pain score and the function score change over time. The distributions of scores between periods are similar and almost symmetric. The similar plots with log transformed total score can be found in Figure S1 in Appendix. Figure 2 shows the individual trajectory of the total score (in log-scale) from some randomly selected patients, which shows that the measurements for each individual are likely to be correlated while there seem to be relatively large variations between individuals. This plot indicates the appropriateness of using the linear mixed effect model in the later section.

Figure 1 The boxplot of total, pain and function score (in original scale) between periods

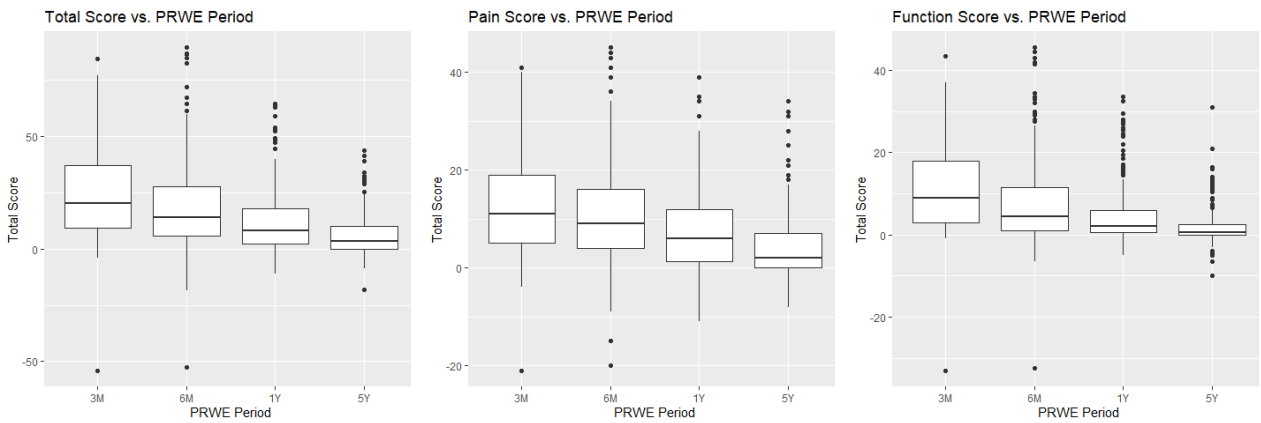
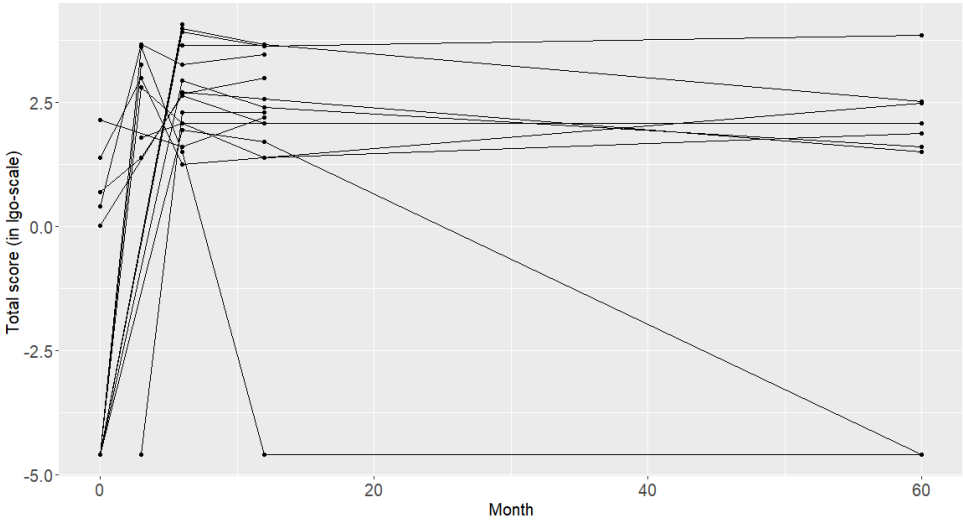


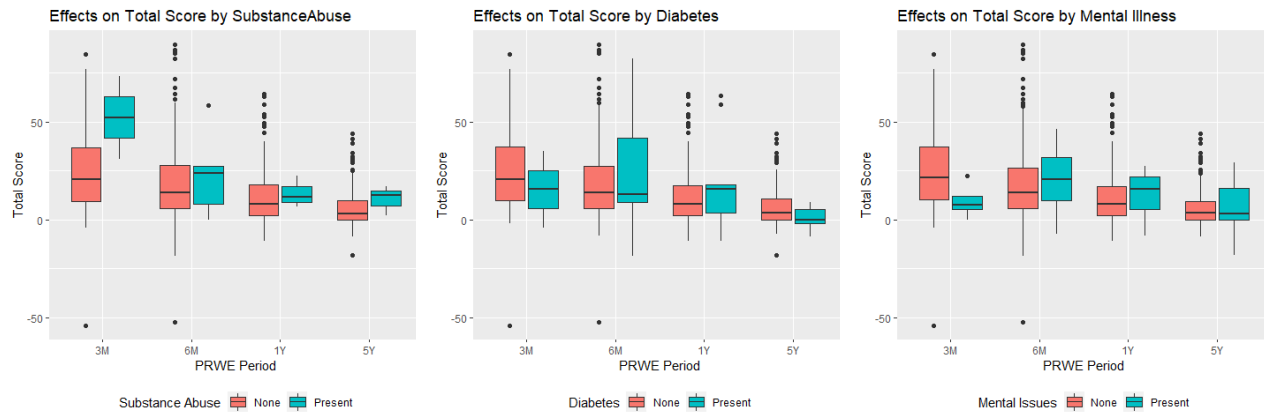
Figure 2 The individual trajectory of the total score (in log-scale) from some randomly selected patients



2.2 Univariate Analysis

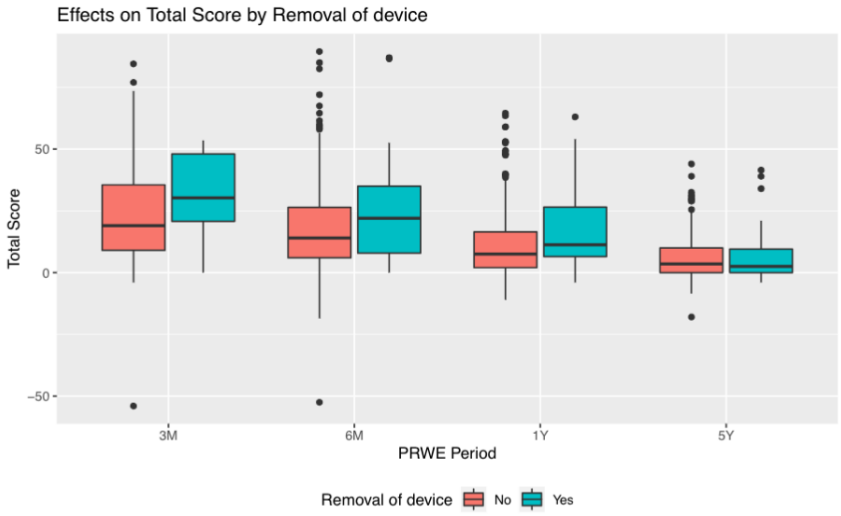
For each period, several conditions may have effects on the patients' scores. Three conditions with relatively vital effects are selected to show the differences, as shown in Figure 3 (all in original scale, the plots with log-scale can be found in Appendix). According to the boxplots grouped by the dummy variables such as "Substance Abuse", "Diabetes" and "Mental Illness", there seems to be some difference between two groups of patients who have these conditions present or not.

Figure 3 The effects of Substance Abuse, Diabetes and Mental illness on the total score



In addition to the basic condition variables, we are most interested in the effects of the records of device on patients' recovery. The types of records indicate whether the device has been removed or not for a patient after his/her one-year treatment. Figure 4 below is a boxplot showing the effect of the removal of device.

Figure 4 The effects of removal of device on the total score (in original scale)



It is observed that patients who would remove their device after one year of treatment may have higher scores in this year than others. However, after the removal of device, the two groups of patients seem to have similar scores.

To confirm the effects of removal of device, two sample t-tests and Wilcoxon tests were conducted for each period. By checking the normality of the scores in each level, it appears that the data are likely to be normally distributed after removing extreme values/ outliers. Additionally, variance was checked to be fairly constant in each period.

Table 3 The t-tests and Wilcoxon test results of the effects on the removal of device

Period	T statistic	P-Value	Wilcoxon Statistic	P-Value
3M	1.691	0.1050	690.0	0.0629
6M	1.763	0.0861	4458.0	0.0979
1Y	2.860	0.0073	2624.0	0.0019
5Y	0.728	0.4730	3088.5	0.8630

As shown in Table 3, though the p-values do not all strictly smaller than 0.05, which is the commonly used significance level, patients with device removed or not may still have a substantial difference in scores in the first year after treatment. However, similar scores may be observed in both groups of patients for the period of five years after treatment. In addition to the t tests, Wilcoxon test can also be applied to confirm the results. Under Wilcoxon test, p-values for the first year ("3M", "6M" and "1Y") are all fairly small, indicating the considerable difference between the two groups of patients with device removed or not. The scores at "5 Year" are highly similar between two groups. The test results appear to be consistent with the procedure of removal device that device would be removed for patients who were uncomfortable with the devices installed in surgeries and the removal would be conducted after one-year treatment. Besides, patients with the removal of device are likely to recover from the treatment similarly with patients who were initially comfortable with the device, in the long term, such as five years.

The score difference between the two groups of patients may also indicate the difference of recovery rates between periods. Table 4 summarizes the "good outcome" rates of two groups of patients with devices removed or not, based on the MCID standard. The tests on proportions of "good outcome", which are the Chi-Squared Test and the Fisher Exact Test, are also conducted and reported in Table 4.

Table 4 The "good outcome" rates for removal and non-removal patients at each time point

Period	3M	6M	1Y	5Y
Non-Removal	36/121 (29.75%)	150/319 (47.02%)	165/262 (62.98%)	192/250 (76.80%)
Removal	2/16 (12.50%)	12/34 (35.29%)	15/31 (48.39%)	20/26 (76.92%)
Chi-Sq P-value	0.1475	0.192	0.1145	0.9887
Fisher P-Value	0.2341	0.2096	0.1226	1
Overall	27.74%	45.89%	61.43%	76.81%

According to the summarized good outcome rates, patients who had devices removed in the revision tend to have lower recovery rates in the first year after surgeries. However, in the long term such as five years, the results are consistent with the assumptions drawn from Table 3 that the two groups of patients with removal of device or not could have the same recovery rates. Though none of the periods has statistically significant difference in proportion of "good outcome", the p-values of periods in the first year are significantly lower than the "5 Year". This is corresponding to the imperfect, but meaningful p-values reported in Table 3. With the device removed, patients who had higher scores in the early periods might show recovery rates similar to those who did not remove the device, in the long term.

Based on the above findings, it is necessary to check the difference of score improvements between "1Y" and "5Y". By checking the normality of the score improvements, a t- test is used to check whether such difference is significant between two groups of patients, as shown in Table 5. Contradicted with the expected significance, the result of the t-test does not support a strong significant difference of score improvements between two groups of patients. One reason of such result may be the unbalanced sample size of both groups of patients. The number of patients who removed the device is much smaller than that of the others. Hence, in the t-test, the difference between the two groups of patients is harder to detect.

Table 5 The t-test and Wilcoxon test results of the removal and non-removal during 1 year to 5-year improvement

Mean Diff	SE of Diff	T Statistic	P-Value	Wilcoxon Statistic	P-Value
0.9286	0.9840	0.9595	0.34	20607	0.2556

3. Confirmatory data analysis

For the purpose of analysis, logistic regression and linear mixed effect models were applied and compared in this section to investigate the associations of explanatory variables and "good outcome". Assumptions for each model were checked to ensure the validity of the proposed models.

3.1 Logistic regression

To predict whether patients will have a "good outcome" or not at 1 year/5 year, we could dichotomize the score differences at 1 year/5 year based on if the score difference are larger than MCID or not. A full model was constructed by putting all the variables of interest such as age, ISS score, heart disease, etc. into the logistic model as covariates. Selections of variables were preformed based on the AIC values. The final selected model was compared with the full model by deviance test. Also, diagnostics of influential points and the over-dispersion issue were checked using Cook's distance and quasi-likelihood respectively.

For the predictive model at 1 year, the deviance test shows that the reduced model and the full model are not significantly different (p-value= 0.25). The reduced model is preferred because it is simpler. The Cook's distance (in Figure S2) shows that there is only one influential data point but after removing it, the model outputs do not change much. The dispersion parameter of this reduced

model is 1.02, suggesting slightly over-dispersion but it is ignorable. The final model outputs for the 1-year predictive model are summarized in Table 6.

Table 6 shows that both "Removal of device" and "Mental illness" are negatively associated with the "good outcome" at 1 year. This also confirms the results shown in EDA in Figure 3.

Table 6 The outputs of the final selected model for 1 year

	Log-odds ratio	SE	p-value
Removal of Device (Yes)	**-0.85**	**0.39**	**0.03**
Osteoporosis (Yes)	-0.99	0.60	0.10
Diabetes (Yes)	-1.01	0.64	0.12
Mental illness (Yes)	**-1.39**	**0.65**	**0.03**
Malignancy (Yes)	1.79	1.29	0.16

Similarly, to check if the early (i.e., the 6 month) score differences are associated with the "good outcome" at 5 year, we performed the same model selection and diagnostic procedures mentioned above for the 1 year by adding one extra covariate "The 6-month score difference." Again, the deviance test indicates that the reduced model and the full model are not significantly different (p-value= 0.18). Meanwhile, the Cook's distance shows that there is one influential data point but after removing it, the model outputs do not change much. The dispersion parameter of this reduced model is 0.99, suggesting slightly under-dispersion but it is ignorable. The final model outputs for the 5-year predictive model are summarized in Table 7.

Table 7 The outputs of the final selected model for 5 year

	Log-odds ratio	SE	p-value
Sex (Male)	0.61	0.39	0.12
ISS	6.69	339.19	0.98
Heart disease (Yes)	-1.49	1.01	0.14
Osteoporosis (Yes)	-1.12	0.77	0.15
Diabetes (Yes)	18.43	1865.11	0.99
Substance Abuse (Yes)	-2.05	1.40	0.14
Previous orthopedic trauma (Yes)	16.43	1025.01	0.98
The 6-month score difference	**-0.06**	**0.01**	**3.49e-09**

Table 7 above shows that the early score difference is negatively associated with a "good outcome", which means a patient with a higher early score difference will be less likely to have a "good outcome" at 5 year. Other variables, such as the "Substance Abuse," are negatively associated with

the “good outcome”, which is also consistent with the EDA results in Figure 3, though it is not significant in the model. However, according to the model outputs, some variables, such as the “Diabetes” and the “Previous orthopedic trauma”, are positively associated with the “good outcome”, which seems to be counter intuitive. This is probably due to the fact that most of the patients in this dataset only have one level, thus the SE of the estimates are really high (and the p-values are high as well) and the results of these estimates are not reliable. The bootstrap was also used to estimate the SE of the coefficients but since most of the times (over 60%) the ‘sampled' datasets only contained one level of the covariate (e.g. Previous orthopedic trauma present), and simply ignoring all of these datasets may lead to biased results. Perhaps the best way to get more accurate estimates of the two coefficients is to recruit more patients with (or without) these symptoms to balance the two levels.

3.2 Linear mixed effect (LME) model

Figure 2 in the EDA indicates that there are large variations between individuals while the observations within each patient are correlated. This suggests the linear mixed effect (LME) model will be quite useful here. Also, an LME model uses all data simultaneously, so it is more efficient.

For LME model, as mentioned previously, we transformed the score at each time period into a log-scale, as the normality assumption of the original score does not hold that well. Since the baseline score is the “before injury score”, it is thus much lower than the score afterwards. We thus considered to treat the baseline score as an individual covariate rather than the score at time “0”.

We first fitted a linear model to each individual and checked how the estimates vary across individuals based on a 95% confidence interval (Figure S3). It suggests that the random effects may be needed for both the intercept and the slope in a single mixed effect model as some of these confidence intervals do not overlap with each other. We then fitted a random intercept model as well as a random intercept and random slope model and compared the two mixed effects model using AIC and BIC values. The random intercept and random slope model turned out to have a better fit as it had the smaller AIC and BIC values, and the likelihood ratio test showed significant results ($p<0.0001$). The normality of the random effects and the residues were checked by normal Q-Q plots (Figure S4), which support the assumption of normality.

GEE models could be alternative methods here for the analysis of longitudinal data. Noticeably, GEE models do not need distributional assumptions, so they are especially useful if distributional assumptions are hard to verify, or if the normality assumptions for the residues and the random effects do not hold that well. Unlike LME, the GEE models do not have random effects, thus they do not allow individual-specific inference. Also, the GEE estimates may be less efficient than LME estimates if the distributional assumptions hold.

Moreover, we conducted the sensitive analysis by looking at the model fit through GEE with different working correlation structures. The results from the GEE models with different working correlation matrix seem to be good as their Naïve SE (standard errors) are close to the Robust SE.

The results from the GEE models are also quite comparable with the LME, as shown in Table 8. This sensitivity analysis confirms the reliability of the results from the LME. Detailed results from LME and GEE with different correlation structures are summarized in Table 8.

Table 8 The outputs from LME and GEE with different correlation structures

Model		Estimates	SE/Naïve S.E.	Robust S.E.	p-value
LME	Intercept	2.69	0.06	-	0
	Month	-0.02	0.001	-	0
	Baseline score	0.18	0.04	-	0
GEE with unstructured correlation structure	Intercept	2.68	0.06	0.06	0
	Month	-0.02	0.001	0.001	0
	Baseline score	0.19	0.05	0.05	0
GEE with independent correlation structure	Intercept	2.65	0.05	0.06	0
	Month	-0.02	0.001	0.001	0
	Baseline score	0.20	0.04	0.05	0
GEE with exchangeable correlation structure	Intercept	2.70	0.06	0.06	0
	Month	-0.02	0.001	0.001	0
	Baseline score	0.19	0.05	0.05	0

Overall, from the LME, the mean score is decreasing with time, which has confirmed the results from EDA showing a decreasing trend. From the LME we can further quantify this score decrement, which is on average 0.02-unit score decrement per month.

Additionally, we tried to adjust the "removal of device" in the model and repeat the above model selection procedures. We finally selected the model with both the random intercept and the random slope. The model outputs are summarized in Table S1. The model selection and diagnostic details can be found in Figure S4.

4. Conclusion

In conclusion, the "Removal of device" is negatively associated with the "good outcome" at 1 year, i.e., patients who have the "Removal of device" are less likely to have a "good outcome" at 1 year; the early score difference is negatively associated with a "good outcome", i.e., a patient with a higher early score difference will be less likely to have a "good outcome" at 5 years; on average, the score is decreasing over time. Additionally, the results of the LME model show that on average, there is a 0.02-unit score (in log-scale) decrement per month.

We do have some limitations regarding the analysis results: the data were collected by convenient sampling thus the results and conclusions may not be generalized to the whole population. The dataset contains a relatively large percentage of missing data and we only focused on the completed dataset for our analysis, which will lead to a less powered analysis, and these results could be biased if the missing is informative. If the missing is not informative, we may try multiple imputation methods to impute the missing data and do the analysis again on the imputed datasets to compare the results.

5. Reference

1. Nellans KW, Kowalski E, Chung KC. The epidemiology of distal radius fractures. Hand Clin. 2012;28(2):113-125. doi:10.1016/j.hcl.2012.02.001

2. Walenkamp MM, de Muinck Keizer RJ, Goslings JC, Vos LM, Rosenwasser MP, Schep NW. The Minimum Clinically Important Difference of the Patient-rated Wrist Evaluation Score for Patients with Distal Radius Fractures [published correction appears in Clin Orthop Relat Res. 2015 Sep;473(9):3063]. Clin Orthop Relat Res. 2015;473(10):3235-3241doi:10.1007/s11999015-4376-9

3. Neufeld ME, Broekhuyse HM, O'Brien PJ, Guy P, Lefaivre KA. The Longitudinal Short-, Medium-, and Long-Term Functional Recovery After Unstable Pelvic Ring Injuries. J Orthop Trauma. 2019;33(12):608-613. doi:10.1097

Appendix:

Figure S1.1: The boxplot of total score, pain score and function score in **log-scale** between periods

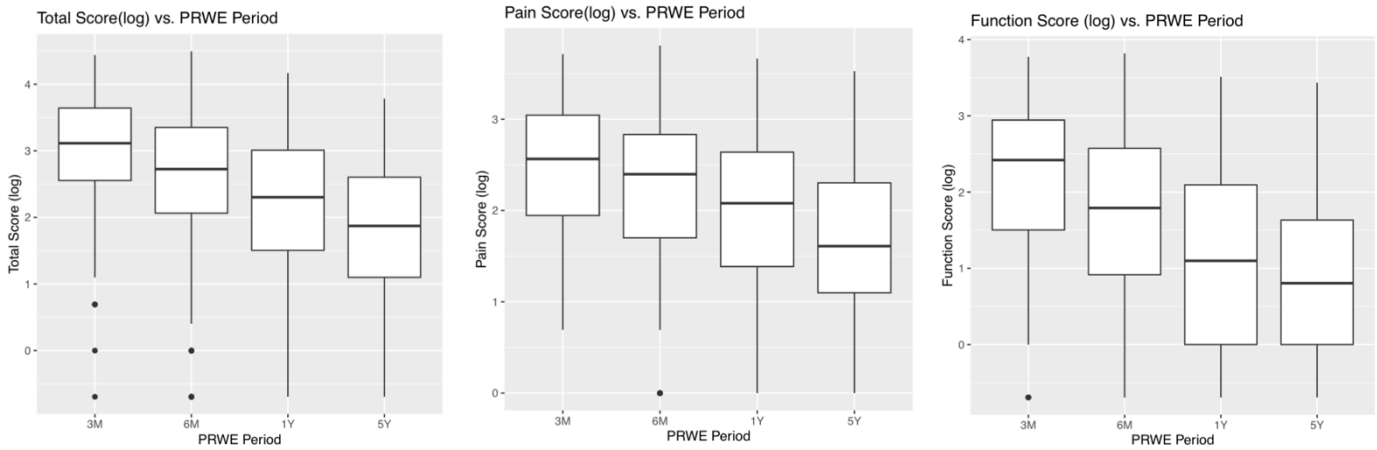


Figure S1.2: The effects of Substance Abuse, Diabetes and Mental illness on the total score in **log-scale** between periods

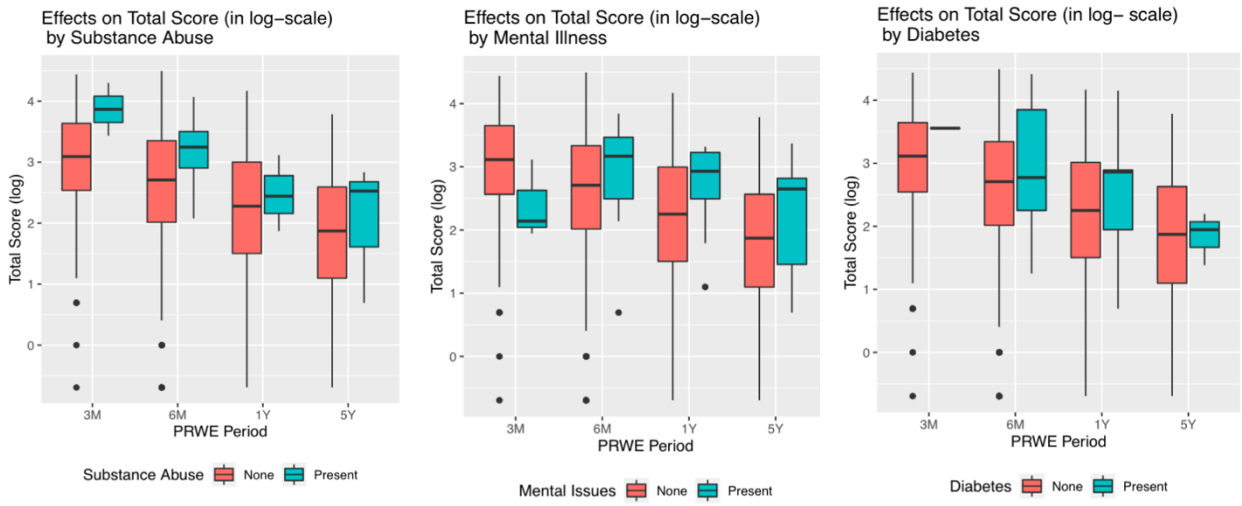


Figure S1.3: The effects of removal of device on the total score in **log-scale**

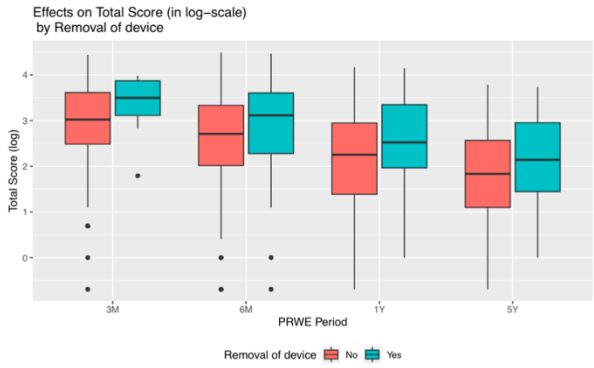


Figure S2: The Cook's distance plot for the reduced model of 1 year (left) and 5 year (right).

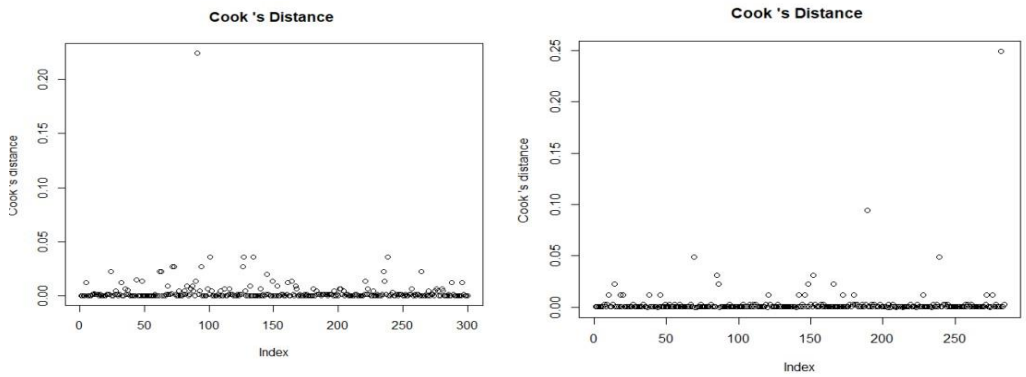


Figure S3: Confidence intervals for all individual estimates

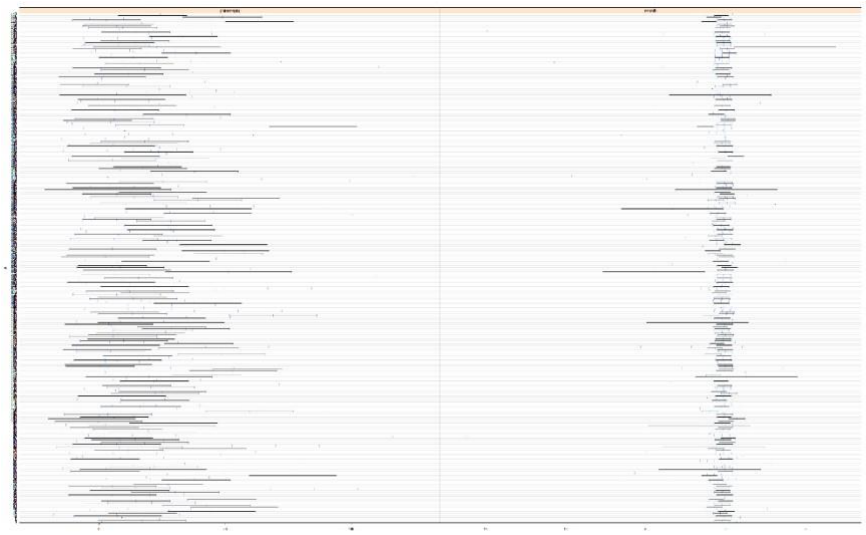


Figure S4: Q-Q plot for the random effects (left) and the residues (right):

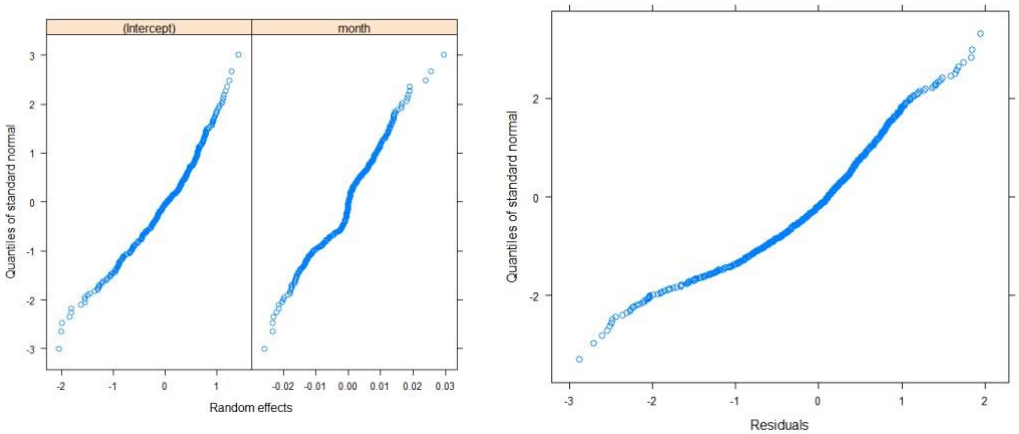
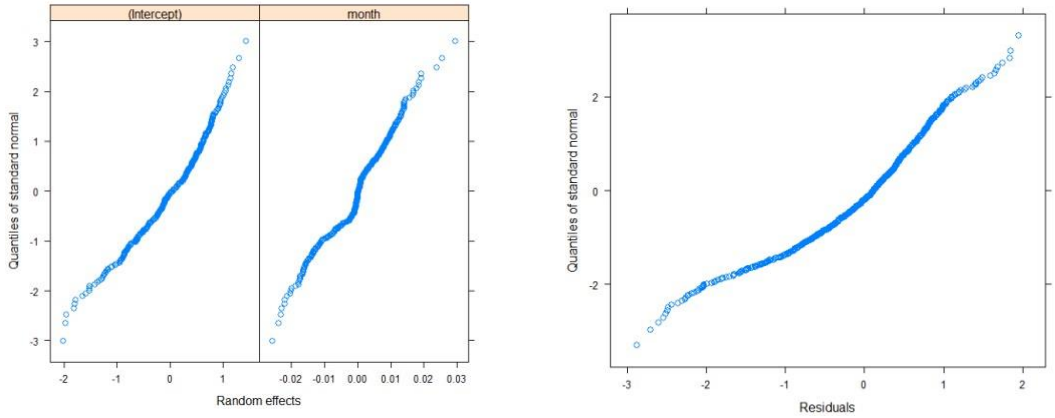


Table S1: The model outputs of LME and GEE with different working correlation matrix when adjusting "Removal of device".

Model	Variables	Estimates	SE/Naïve S.E.	Robust S.E.	p-value
LME	Intercept	2.67	0.06	-	0
	Month	-0.02	0.0020	-	0
	Baseline score	0.18	0.050	-	0.0002
	Removal of device	0.26	0.18	-	0.14
GEE with unstructured correlation structure	Intercept	2.66	0.06	0.06	0
	Month	-0.02	0.0014	0.0015	0
	Baseline score	0.19	0.046	0.049	0
	Removal of device	0.25	0.17	0.19	0.003
GEE with independent correlation structure	Intercept	2.62	0.06	0.06	0
	Month	-0.02	0.0017	0.0016	0
	Baseline score	0.21	0.036	0.051	0
	Removal of device	0.28	0.13	0.20	0.002
GEE with exchangeable correlation structure	Intercept	2.68	0.06	0.06	0
	Month	-0.02	0.0013	0.0015	0
	Baseline score	0.20	0.049	0.046	0
	Removal of device	0.21	0.19	0.21	0.005

Figure S4: Q-Q plot for the random effects (left) and the residues (right) of the model adjusting the "removal of device":



Appendix: Code

STAT 550 Project

```
library(tidyverse)
library(broom)
library(readxl)
library(ggfortify)
library(ggpubr)
library(rstatix)
library(nlme)
```

Data Import

```
# data <- read_excel("Stat_Data.xlsx", skip = 1)
# write.csv(data, "data.csv")
```

```
c1 <- read.csv("data.csv")
```

```
c1 <- c1 %>%
  mutate(Sex = Sex, across(Sex, ~if_else(.=="F", 0, 1))) %>%
  mutate(CAD = CAD, across(CAD, ~if_else(.=="None", 0, 1))) %>%
  mutate(Hypertension = Hypertension, across(Hypertension, ~if_else(.=="None", 0, 1))) %>%
  mutate(Osteoporosis = Osteoporosis, across(Osteoporosis, ~if_else(.=="None", 0, 1))) %>%
  mutate(Diabetes = Diabetes, across(Diabetes, ~if_else(.=="None", 0, 1))) %>%
  mutate(Substance.abuse = Substance.abuse, across(Substance.abuse, ~if_else(.=="None", 0, 1))) %
>%
  mutate(Alcohol.abuse = Alcohol.abuse, across(Alcohol.abuse, ~if_else(.=="None", 0, 1))) %>%
  mutate(Depression = Depression, across(Depression, ~if_else(.=="None", 0, 1))) %>%
  mutate(Anxiety.disorder = Anxiety.disorder, across(Anxiety.disorder, ~if_else(.=="None", 0, 1)))
%>%
  mutate(Psychosis = Psychosis, across(Psychosis, ~if_else(.=="None", 0, 1))) %>%
  mutate(Malignancy = Malignancy, across(Malignancy, ~if_else(.=="None", 0, 1))) %>%
  mutate(Stroke.TIA = Stroke.TIA, across(Stroke.TIA, ~if_else(.=="None", 0, 1))) %>%
  mutate(Previous.orthopedic.trauma = Previous.orthopedic.trauma, across(Previous.orthopedic.traum
a, ~if_else(.=="None", 0, 1))) %>%
  mutate(Revision.procedure = Revision.procedure, across(Revision.procedure, ~if_else(.=="Removal
 of  device", 1, 0)))
```

Data Modifications

Data Clean

```
# c1 removes the rows with all NA or all "0".
c1 <- c1 %>% filter(!is.na(Total_3M) | !is.na(Total_6M) | !is.na(Total_1Y) | !is.na(Total_5Y)) %>%
  filter(Total_3M != 0 | Total_6M != 0 | Total_1Y != 0 | Total_5Y != 0)
```

Data for LME

```
# C3 is for LME use, with at least two non-NA data points
c3 <- c1 %>% filter((1*(is.na(Total_3M)) + 1*(is.na(Total_6M)) + 1*(is.na(Total_1Y)) + 1*(is.na(To
tal_5Y))) <= 2)
```

```
# regroup several attributes
easyc3 <- c3 %>%
  select(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, Substance.abu
se, Alcohol.abuse, Depression, Anxiety.disorder, Psychosis, Malignancy, Stroke.TIA, Previous.ortho
pedic.trauma, Revision.procedure, Function_Baseline, Pain_Baseline, Total_Baseline, Function_3M, P
ain_3M, Total_3M, Function_6M, Pain_6M, Total_6M, Function_1Y, Pain_1Y, Total_1Y, Function_5Y, Pai
n_5Y, Total_5Y) %>%
  mutate(SubAbuse = 1*((Substance.abuse + Alcohol.abuse)!=0), Mental.illness = 1*((Depression + An
xiety.disorder + Psychosis)!=0)) %>%
  select(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, SubAbuse, Men
tal.illness,
         Malignancy, Stroke.TIA,
         Previous.orthopedic.trauma, Revision.procedure, Function_Baseline, Pain_Baseline, Total_B
aseline,
         Function_3M, Pain_3M, Total_3M, Function_6M, Pain_6M, Total_6M, Function_1Y, Pain_1Y, Tot
al_1Y,
         Function_5Y, Pain_5Y, Total_5Y)
```

```
# longer form of easyc3
longc3 <- easyc3 %>% pivot_longer(-c(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporo
sis, Diabetes, SubAbuse, Mental.illness, Malignancy, Stroke.TIA, Previous.orthopedic.trauma, Revis
ion.procedure, Function_Baseline, Pain_Baseline, Total_Baseline), names_to = c(".value", "period"
), names_sep = "_") %>%
  mutate(month = period, across(month, ~if_else(.=="3M", 3, if_else(.=="6M", 6, if_else(.=="1Y", 1
2, 60)))))
```

```
# delete NA for lme use.
# Count missing values
lme_longc3 <- longc3 %>% drop_na(Total)
lme_longc3 %>% group_by(period) %>% count(is.na(Total_Baseline))
```

LME plot

```
fit.pre <- lmList(Total~ month|X, data = lme_longc3, na.action=na.exclude)
plot(intervals(fit.pre))
```

Data for EDA

```
# easyc1 removes some dates and some useless variables

easyc1 <- c1 %>% select(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabete
s, Substance.abuse, Alcohol.abuse, Depression, Anxiety.disorder, Psychosis, Malignancy, Stroke.TI
A, Previous.orthopedic.trauma, Revision.procedure, Function_Baseline, Pain_Baseline, Total_Baselin
e, Function_3M, Pain_3M, Total_3M, Function_6M, Pain_6M, Total_6M, Function_1Y, Pain_1Y, Total_1Y,
Function_5Y, Pain_5Y, Total_5Y)
```

```
# Combine some general categories

easyc2 <- easyc1 %>%
  mutate(SubAbuse = 1*((Substance.abuse + Alcohol.abuse)!=0), Mental.illness = 1*((Depression + An
xiety.disorder + Psychosis)!=0)) %>%
  select(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, SubAbuse, Men
tal.illness,
         Malignancy, Stroke.TIA,
         Previous.orthopedic.trauma, Revision.procedure, Function_Baseline, Pain_Baseline, Total_B
aseline,
         Function_3M, Pain_3M, Total_3M, Function_6M, Pain_6M, Total_6M, Function_1Y, Pain_1Y, Tot
al_1Y,
         Function_5Y, Pain_5Y, Total_5Y)
```

```
# longc1 is the longer form of easyc1

longc1 <- easyc1 %>% pivot_longer(-c(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporo
sis, Diabetes, Substance.abuse, Alcohol.abuse, Depression, Anxiety.disorder, Psychosis, Malignanc
y, Stroke.TIA, Previous.orthopedic.trauma, Revision.procedure), names_to = c(".value", "period"),
 names_sep = "_")
```

```
# Long form of combined data

longc2 <- easyc2 %>% pivot_longer(-c(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporo
sis, Diabetes, SubAbuse, Mental.illness, Malignancy, Stroke.TIA, Previous.orthopedic.trauma, Revis
ion.procedure), names_to = c(".value", "period"), names_sep = "_")
```

```
# Missing Value Stat
longc2 %>% group_by(period) %>% count(is.na(Total))
```

Baseline Removal

```
# rmBasec1 regulizes the Baseline scores and
# removes the Baseline records

rmBasec1 <- easyc1 %>%
  mutate(Function_3M = Function_3M - Function_Baseline,
                          Pain_3M = Pain_3M - Pain_Baseline,
                          Total_3M = Total_3M - Total_Baseline,
                          Function_6M = Function_6M - Function_Baseline,
                          Pain_6M = Pain_6M - Pain_Baseline,
                          Total_6M = Total_6M - Total_Baseline,
                          Function_1Y = Function_1Y - Function_Baseline,
                          Pain_1Y = Pain_1Y - Pain_Baseline,
                          Total_1Y = Total_1Y - Total_Baseline,
                          Function_5Y = Function_5Y - Function_Baseline,
                          Pain_5Y = Pain_5Y - Pain_Baseline,
                          Total_5Y = Total_5Y - Total_Baseline) %>%
  select(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, Substance.abu
se, Alcohol.abuse, Depression, Anxiety.disorder, Psychosis, Malignancy, Stroke.TIA, Previous.ortho
pedic.trauma, Revision.procedure, Function_Baseline, Pain_Baseline, Total_Baseline, Function_3M, P
ain_3M, Total_3M, Function_6M, Pain_6M, Total_6M, Function_1Y, Pain_1Y, Total_1Y, Function_5Y, Pai
n_5Y, Total_5Y)
```

```
# Remove baseline from combined data

rmBasec2 <- easyc2 %>%
  mutate(Function_3M = Function_3M - Function_Baseline,
                          Pain_3M = Pain_3M - Pain_Baseline,
                          Total_3M = Total_3M - Total_Baseline,
                          Function_6M = Function_6M - Function_Baseline,
                          Pain_6M = Pain_6M - Pain_Baseline,
                          Total_6M = Total_6M - Total_Baseline,
                          Function_1Y = Function_1Y - Function_Baseline,
                          Pain_1Y = Pain_1Y - Pain_Baseline,
                          Total_1Y = Total_1Y - Total_Baseline,
                          Function_5Y = Function_5Y - Function_Baseline,
                          Pain_5Y = Pain_5Y - Pain_Baseline,
                          Total_5Y = Total_5Y - Total_Baseline) %>%
  select(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, SubAbuse, Men
tal.illness, Malignancy, Stroke.TIA, Previous.orthopedic.trauma, Revision.procedure, Function_Base
line, Pain_Baseline, Total_Baseline, Function_3M, Pain_3M, Total_3M, Function_6M, Pain_6M, Total_6
M, Function_1Y, Pain_1Y, Total_1Y, Function_5Y, Pain_5Y, Total_5Y)
```

```
# Calculate the score difference between periods

ScoreDiff <- rmBasec1 %>%
  select(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, Substance.abu
se, Alcohol.abuse, Depression, Anxiety.disorder, Psychosis, Malignancy, Stroke.TIA, Previous.ortho
pedic.trauma, Revision.procedure, Function_3M, Pain_3M, Total_3M, Function_6M, Pain_6M, Total_6M,
 Function_1Y, Pain_1Y, Total_1Y, Function_5Y, Pain_5Y, Total_5Y) %>%
  pivot_longer(-c(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, Subs
tance.abuse, Alcohol.abuse, Depression, Anxiety.disorder, Psychosis, Malignancy, Stroke.TIA, Previ
ous.orthopedic.trauma, Revision.procedure), names_to = c("part", ".value"), names_sep = "_") %>%
  mutate(Diff_6M3M = `6M` - `3M`,
         Diff_1Y3M = `1Y` - `3M`,
         Diff_1Y6M = `1Y` - `6M`,
         Diff_5Y6M = `5Y` - `6M`,
         Diff_5Y1Y = `5Y` - `1Y`) %>%
  mutate(Crt_6M3M = Diff_6M3M,
         across(Crt_6M3M, ~if_else(.<=(-11.5), 1, if_else(.>=11.5, -1, 0))),
         Crt_1Y3M = Diff_1Y3M,
         across(Crt_1Y3M, ~if_else(.<=(-11.5), 1, if_else(.>=11.5, -1, 0))),
         Crt_1Y6M = Diff_1Y6M,
         across(Crt_1Y6M, ~if_else(.<=(-11.5), 1, if_else(.>=11.5, -1, 0))),
         Crt_5Y6M = Diff_5Y6M,
         across(Crt_5Y6M, ~if_else(.<=(-11.5), 1, if_else(.>=11.5, -1, 0))),
         Crt_5Y1Y = Diff_5Y1Y,
         across(Crt_5Y1Y, ~if_else(.<=(-11.5), 1, if_else(.>=11.5, -1, 0))))
```

```
# reglongc1 is the long form of rmBasec1

reglongc1 <- rmBasec1 %>%
  select(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, Substance.abu
se, Alcohol.abuse, Depression, Anxiety.disorder, Psychosis, Malignancy, Stroke.TIA, Previous.ortho
pedic.trauma, Revision.procedure, Function_3M, Pain_3M, Total_3M, Function_6M, Pain_6M, Total_6M,
 Function_1Y, Pain_1Y, Total_1Y, Function_5Y, Pain_5Y, Total_5Y) %>%
  pivot_longer(-c(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, Subs
tance.abuse, Alcohol.abuse, Depression, Anxiety.disorder, Psychosis, Malignancy, Stroke.TIA, Previ
ous.orthopedic.trauma, Revision.procedure), names_to = c(".value", "period"), names_sep = "_")
```

```
# reglongc2 is the long form of rmBasec2

reglongc2 <- rmBasec2 %>%
  select(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, SubAbuse, Men
tal.illness, Malignancy, Stroke.TIA, Previous.orthopedic.trauma, Revision.procedure, Function_3M,
 Pain_3M, Total_3M, Function_6M, Pain_6M, Total_6M, Function_1Y, Pain_1Y, Total_1Y, Function_5Y, P
ain_5Y, Total_5Y) %>%
  pivot_longer(-c(X, MRN, Age.at.injury, Sex, ISS, CAD, Hypertension, Osteoporosis, Diabetes, SubA
buse, Mental.illness, Malignancy, Stroke.TIA, Previous.orthopedic.trauma, Revision.procedure), nam
es_to = c(".value", "period"), names_sep = "_")
```

EDA

Histograms and Boxplots

```
easyc2 %>% ggplot(aes(x = Age.at.injury)) +
  geom_histogram(aes(y = ..density..), binwidth = 5) +
  geom_density(col = 4, size = 1) +
  labs(title = "Histogram of Age at injury") +
  geom_vline(aes(xintercept = mean(Age.at.injury)), col = "red", size = 1)
```

```
# allocate levels of factor `period`

longc1$period <- factor(longc1$period, levels= c("Baseline", "3M", "6M", "1Y", "5Y"))

reglongc1$period <- factor(reglongc1$period, levels= c("3M", "6M", "1Y", "5Y"))

longc2$period <- factor(longc2$period, levels= c("Baseline", "3M", "6M", "1Y", "5Y"))

reglongc2$period <- factor(reglongc2$period, levels= c("3M", "6M", "1Y", "5Y"))
```

```
# Find the extreme values and remove from dataset
rm_reglongc2 <- reglongc2 %>% select(X, period, Total) %>%
  group_by(period) %>%
  identify_outliers(Total) %>%
  filter(is.extreme == TRUE)

new_reglongc2 <- anti_join(reglongc2, rm_reglongc2)
```

Log Scale

```
new_reglongc2 %>% ggplot(aes(x = period, y = log(Total))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score (log)") +
  ggtitle("Total Score vs. PRWE Period")
```

```
new_reglongc2 %>% ggplot(aes(x = period, y = log(Pain))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score (log)") +
  ggtitle("Pain Score vs. PRWE Period")
```

```
new_reglongc2 %>% ggplot(aes(x = period, y = log(Function))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score (log)") +
  ggtitle("Function Score vs. PRWE Period")
```

```
new_reglongc2 %>% ggplot(aes(y = log(Total), x = period, fill = as.factor(Sex))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score (log)") +
  ggtitle("Effects on Total Score by Sex") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Sex", labels = c("Female", "Male"))
```

```
new_reglongc2 %>% ggplot(aes(y = log(Total), x = period, fill = as.factor(SubAbuse))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score (log)") +
  ggtitle("Effects on Total Score by SubstanceAbuse") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Substance Abuse", labels = c("None", "Present"))
```

```
new_reglongc2 %>% ggplot(aes(y = log(Total), x = period, fill = as.factor(CAD))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score (log)") +
  ggtitle("Effects on Total Score by Heart Issues") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Heart Issues", labels = c("None", "Present"))
```

```
new_reglongc2 %>% ggplot(aes(y = log(Total), x = period, fill = as.factor(Mental.illness))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score (log)") +
  ggtitle("Effects on Total Score by Mental Illness") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Mental Issues", labels = c("None", "Present"))
```

```
new_reglongc2 %>% ggplot(aes(y = log(Total), x = period, fill = as.factor(Diabetes))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score (log)") +
  ggtitle("Effects on Total Score by Diabetes") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Diabetes", labels = c("None", "Present"))
```

```
new_reglongc2 %>% ggplot(aes(y = log(Total), x = period, fill = as.factor(Revision.procedure))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score (log)") +
  ggtitle("Effects on Total Score by Removal of device") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Removal of device", labels = c("No", "Yes"))
```

Original Scale

```
new_reglongc2 %>% ggplot(aes(x = period, y = Total)) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Total Score vs. PRWE Period")
```

```
new_reglongc2 %>% ggplot(aes(x = period, y = Pain)) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Pain Score vs. PRWE Period")
```

```
new_reglongc2 %>% ggplot(aes(x = period, y = Function)) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Function Score vs. PRWE Period")
```

```
new_reglongc2 %>% ggplot(aes(y = Total, x = period, fill = as.factor(Sex))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Effects on Total Score by Sex") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Sex", labels = c("Female", "Male"))
```

```
new_reglongc2 %>% ggplot(aes(y = Total, x = period, fill = as.factor(SubAbuse))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Effects on Total Score by SubstanceAbuse") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Substance Abuse", labels = c("None", "Present"))
```

```
new_reglongc2 %>% ggplot(aes(y = Total, x = period, fill = as.factor(CAD))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Effects on Total Score by Heart Issues") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Heart Issues", labels = c("None", "Present"))
```

```
new_reglongc2 %>% ggplot(aes(y = Total, x = period, fill = as.factor(Mental.illness))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Effects on Total Score by Mental Illness") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Mental Issues", labels = c("None", "Present"))
```

```
new_reglongc2 %>% ggplot(aes(y = Total, x = period, fill = as.factor(Diabetes))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Effects on Total Score by Diabetes") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Diabetes", labels = c("None", "Present"))
```

```
new_reglongc2 %>% ggplot(aes(y = Total, x = period, fill = as.factor(Revision.procedure))) +
  geom_boxplot() +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Effects on Total Score by Removal of device") +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Removal of device", labels = c("No", "Yes"))
```

```
new_reglongc2 %>% group_by(period, Revision.procedure) %>% summarise(Total = mean(Total, na.rm = T
RUE), n = n()) %>%
  ggplot(aes(y = Total, x = period, col = as.factor(Revision.procedure))) +
  geom_point(size = 4) +
  xlab("PRWE Period") +
  ylab("Total Score") +
  ggtitle("Average Total Score by Removal of device")
```

Tests on 5Y-1Y

```
rmDevice <- reglongc1 %>% select(X, period, Revision.procedure, Total)
rmDevice1 <- ScoreDiff %>% select(X, Revision.procedure, Diff_5Y1Y)
```

```
# extreme values removed
rm_df_51 <- rmDevice1 %>%
  group_by(Revision.procedure) %>%
  identify_outliers(Diff_5Y1Y) %>%
  filter(is.extreme == TRUE) %>%
  select(Revision.procedure, X, Diff_5Y1Y)

new_5Y1Y <- anti_join(rmDevice1, rm_df_51)
prop_5Y1Y <- new_5Y1Y %>% mutate(recovery = Diff_5Y1Y, across(recovery, ~if_else(.<= 11.5, 1, 0)))
```

```
# counts
prop_5Y1Y %>% filter(Revision.procedure == 0, recovery == 0) %>% count()
prop_5Y1Y %>% filter(Revision.procedure == 0, recovery == 1) %>% count()
prop_5Y1Y %>% filter(Revision.procedure == 1, recovery == 0) %>% count()
prop_5Y1Y %>% filter(Revision.procedure == 1, recovery == 1) %>% count()
```

```
# proportion test
prop_mtx <- rbind(c(569, 15), c(63, 2))
chisq.test(prop_mtx, correct = FALSE)
fisher.test(prop_mtx)
prop.test(prop_mtx)
```

Normality check

```
modelDe <- lm(Diff_5Y1Y ~ Revision.procedure,
              data = new_5Y1Y)
ggqqplot(residuals(modelDe))
```

```
shapiro_test(residuals(modelDe))
```

```
new_5Y1Y %>% group_by(Revision.procedure) %>%
  shapiro_test(Diff_5Y1Y)
```

T Wilcoxon K-W

```
t.test(Diff_5Y1Y ~ Revision.procedure, data = new_5Y1Y)
wilcox.test(Diff_5Y1Y ~ Revision.procedure, data = new_5Y1Y)
new_5Y1Y %>%
  kruskal.test(Diff_5Y1Y ~ Revision.procedure)

-1.771404+2.700000
((1.000074 + 2.857266)/2)/qnorm(0.975)
```

Tests on Periods

```
# extreme values removed
rm_df_period <- rmDevice %>% group_by(period, Revision.procedure) %>%
  identify_outliers(Total) %>%
  filter(is.extreme == TRUE) %>%
  select(X, period, Revision.procedure, Total)
new_rm_p <- anti_join(rmDevice, rm_df_period)
```

Normality Check

```
model  <- lm(Total ~ period*Revision.procedure,
             data = new_rm_p)
ggqqplot(residuals(model))
```

```
shapiro_test(residuals(model))
```

```
new_rm_p %>% group_by(period, Revision.procedure) %>%
  shapiro_test(Total)
```

```
ggqqplot(new_rm_p, "Total", ggtheme = theme_bw()) +
  facet_grid(Revision.procedure ~ period)
```

T Wilconxon K-W

```
model1 <- lm(Total ~ period * Revision.procedure, data = new_rm_p)
new_rm_p %>%
  group_by(period) %>%
  t_test(Total ~ Revision.procedure)
```

```
new_rm_p %>%
  group_by(period) %>%
  wilcox_test(Total ~ Revision.procedure)
```

```
new_rm_p %>%
  group_by(period) %>%
  kruskal_test(Total ~ Revision.procedure)
```

Summarized Analysis

```
# Score difference between periods vs. MCID
ScoreDiff %>% filter(part == "Total") %>%
  summarise(Crt_6M3M = mean(abs(Crt_6M3M), na.rm = TRUE),
            Crt_1Y3M = mean(abs(Crt_1Y3M), na.rm = TRUE),
            Crt_1Y6M = mean(abs(Crt_1Y6M), na.rm = TRUE),
            Crt_5Y6M = mean(abs(Crt_5Y6M), na.rm = TRUE),
            Crt_5Y1Y = mean(abs(Crt_5Y1Y), na.rm = TRUE))
```

```
# Counts
ScoreDiff %>% filter(Crt_6M3M != 0) %>% count()
ScoreDiff %>% filter(!is.na(Crt_6M3M)) %>% count()
ScoreDiff %>% filter(Crt_1Y3M != 0) %>% count()
ScoreDiff %>% filter(!is.na(Crt_1Y3M)) %>% count()
ScoreDiff %>% filter(Crt_1Y6M != 0) %>% count()
ScoreDiff %>% filter(!is.na(Crt_1Y6M)) %>% count()
ScoreDiff %>% filter(Crt_5Y6M != 0) %>% count()
ScoreDiff %>% filter(!is.na(Crt_5Y6M)) %>% count()
ScoreDiff %>% filter(Crt_5Y1Y != 0) %>% count()
ScoreDiff %>% filter(!is.na(Crt_5Y1Y)) %>% count()
```

```
# Recovery rate, by MCID
ScoreDiff %>% filter(part == "Total") %>%
  summarise(`3M` = mean(`3M` <= 11.5, na.rm = TRUE),
            `6M` = mean(`6M` <= 11.5, na.rm = TRUE),
            `1Y` = mean(`1Y` <= 11.5, na.rm = TRUE),
            `5Y` = mean(`5Y` <= 11.5, na.rm = TRUE))
```

```
# Counts
ScoreDiff %>% filter(part == "Total") %>% filter(`3M` <= 11.5) %>% count()
ScoreDiff %>% filter(part == "Total") %>% filter(!is.na(`3M`)) %>% count()
ScoreDiff %>% filter(part == "Total") %>% filter(`6M` <= 11.5) %>% count()
ScoreDiff %>% filter(part == "Total") %>% filter(!is.na(`6M`)) %>% count()
ScoreDiff %>% filter(part == "Total") %>% filter(`1Y` <= 11.5) %>% count()
ScoreDiff %>% filter(part == "Total") %>% filter(!is.na(`1Y`)) %>% count()
ScoreDiff %>% filter(part == "Total") %>% filter(`5Y` <= 11.5) %>% count()
ScoreDiff %>% filter(part == "Total") %>% filter(!is.na(`5Y`)) %>% count()
```

```
# proportion of recovery, by MCID and Removal of Device
ScoreDiff %>% filter(part == "Total") %>% group_by(Revision.procedure) %>%
  summarise(`3M` = mean(`3M` <= 11.5, na.rm = TRUE),
            `6M` = mean(`6M` <= 11.5, na.rm = TRUE),
            `1Y` = mean(`1Y` <= 11.5, na.rm = TRUE),
            `5Y` = mean(`5Y` <= 11.5, na.rm = TRUE))
```

```
# Counts
ScoreDiff %>% filter(part == "Total") %>% group_by(Revision.procedure) %>% filter(`3M` <= 11.5) %
>% count()
ScoreDiff %>% filter(part == "Total") %>% group_by(Revision.procedure) %>% filter(!is.na(`3M`)) %
>% count()
ScoreDiff %>% filter(part == "Total") %>% group_by(Revision.procedure) %>% filter(`6M` <= 11.5) %
>% count()
ScoreDiff %>% filter(part == "Total") %>% group_by(Revision.procedure) %>% filter(!is.na(`6M`)) %
>% count()
ScoreDiff %>% filter(part == "Total") %>% group_by(Revision.procedure) %>% filter(`1Y` <= 11.5) %
>% count()
ScoreDiff %>% filter(part == "Total") %>% group_by(Revision.procedure) %>% filter(!is.na(`1Y`)) %
>% count()
ScoreDiff %>% filter(part == "Total") %>% group_by(Revision.procedure) %>% filter(`5Y` <= 11.5) %
>% count()
ScoreDiff %>% filter(part == "Total") %>% group_by(Revision.procedure) %>% filter(!is.na(`5Y`)) %
>% count()
```

```
prop_mtx <- rbind(c(36, 85), c(2, 14))
chisq.test(prop_mtx, correct = FALSE)
fisher.test(prop_mtx)
prop_test(prop_mtx)
```

```
prop_mtx <- rbind(c(150, 169), c(12, 22))
chisq.test(prop_mtx, correct = FALSE)
fisher.test(prop_mtx)
prop_test(prop_mtx)
```

```
prop_mtx <- rbind(c(165, 97), c(15, 16))
chisq.test(prop_mtx, correct = FALSE)
fisher.test(prop_mtx)
prop_test(prop_mtx)
```

```
prop_mtx <- rbind(c(192, 58), c(20, 6))
chisq.test(prop_mtx, correct = FALSE)
fisher.test(prop_mtx)
prop_test(prop_mtx)
```

- logistic regression
- LME modeling
- GEE and log-transformed LME

logistic regression

```
library(stargazer)
library(Rmisc)
library(reshape2)
library(tidyr)
library(stringr)
library(readxl)
library(nlme)
library(gee)
library(ggplot2)
library(boot)
da<- data.frame(read_excel("new2.xlsx"))
c1 <- read.csv("cleandata.csv")
ct3m_b<-c1[,27]-c1[,22]
ct6m_b<-c1[,32]-c1[,22]
ct1_b<-c1[,37]-c1[,22]
ct5_b<-c1[,42]-c1[,22]


da$Aditional.procedures<-as.factor(da$Aditional.procedures)
da$Revision.procedure<- as.factor (ifelse (da$Revision.procedure == "Removal of  device",1,0))
da$MRN<-as.factor(da$MRN)
da[,5:22]<- ifelse (da[,5:22] == "None", 0,1)
HD<- ifelse (da$CAD + da$Arrhythmia == 0, 0,1)
SA<- ifelse (da$Substance.abuse + da$Alcohol.abuse == 0, 0,1)
MI<- ifelse (da$Depression + da$Anxiety.disorder +da$Psychosis == 0, 0,1)
da<- cbind (da, HD, SA, MI)
da$Sex<-as.factor(da$Sex)
da$ISS<-as.numeric(da$ISS)
da[,5:22]<-lapply(da[,5:22], factor)
da[,55:57]<-lapply(da[,55:57], factor)
da[,23:47]<-lapply(da[,23:47], as.numeric)

t3m_b<-da[,31]-da[,26]
t6m_b<-da[,36]-da[,26]
t1_b<-da[,41]-da[,26]
t5_b<-da[,46]-da[,26]
da3<-cbind(da,t1_b, t5_b,t6m_b )

###### 1 year
da3_1y<-subset(da3, !is.na(t1_b))
ind_1y<- ifelse (da3_1y$t1_b< 11.5,1,0)
da3_1y<-cbind(da3_1y,ind_1y)

fit1_y<-glm(ind_1y~ Revision.procedure +
              #da3_1y$Aditional.procedures+
              Age.at.injury + Sex +ISS + HD+
              Hypertension+Osteoporosis + Diabetes + SA + MI
              + Malignancy + Stroke.TIA + Previous.orthopedic.trauma ,
            data= da3_1y, family= "binomial")

summary(fit1_y)
step(fit1_y)
fit1_y_select<-glm(formula = ind_1y ~ Revision.procedure + Osteoporosis + Diabetes + MI +
```

```
                     Malignancy, family = "quasibinomial", data = da3_1y)
summary(fit1_y_select)
anova(fit1_y, fit1_y_select)
pchisq(7.9057, df=8)
#0.56
cooks.distance(fit1_y_select)
plot(cooks.distance (fit1_y_select), ylab ="Cook 's distance ", main ="Cook 's Distance ")
#which.max(cooks.distance (fit1_y_select))
#stargazer(fit1_y_select, title = "Predictors in reduced model", dep.var.lables = "Good outcom
e")

####### 5 year
da3_5y<-subset(da3, !is.na(t5_b))
ind_5y<- ifelse (da3_5y$t5_b< 11.5,1,0)
da3_5y<-cbind(da3_5y,ind_5y)
str(da3_5y)

fit5_y<-glm(ind_5y~ Revision.procedure +
              #da3_1y$Aditional.procedures+
              Age.at.injury + Sex +ISS + HD+
              Hypertension+Osteoporosis + Diabetes + SA + MI
            + Malignancy + Stroke.TIA + Previous.orthopedic.trauma ,data= da3_5y, family= "binom
ial")

lapply(da3_5y[c("CAD", "Hypertension", "Osteoporosis" , "Diabetes" , "Substance.abuse" ,
                "Alcohol.abuse" , "Depression" ,"Anxiety.disorder",
                "Psychosis" , "Malignancy" , "Stroke.TIA" , "Previous.orthopedic.trauma")], uniq
ue)

summary(fit5_y)
step(fit5_y)
fit5_y_select<-glm(formula = ind_5y ~ Sex + ISS + HD + Diabetes + SA,
                   family = "quasibinomial", data = da3_5y)
summary(fit5_y_select)
anova(fit5_y, fit5_y_select)
plot(cooks.distance (fit5_y_select), ylab ="Cook 's distance ", main ="Cook 's Distance ")
# which.max(cooks.distance (fit5_y_select))
# # 445
# stargazer(fit5_y_select, title = "Predictors in reduced model for 5 year", dep.var.lables = "G
ood outcome")
#### Analysis 4 5 year #####

da4<-cbind(da,t1_b, t5_b,t6m_b )
da4_5y<-subset(da4, !is.na(t5_b))
ind_5y<- ifelse (da4_5y$t5_b< 11.5,1,0)
da4_5y<-cbind(da4_5y,ind_5y)

fit6_y<-glm(ind_5y~ Revision.procedure +
              #da3_1y$Aditional.procedures+
              Age.at.injury + Sex +ISS + HD+
              Hypertension+Osteoporosis + Diabetes + SA + MI
            + Malignancy + Stroke.TIA + Previous.orthopedic.trauma +t6m_b
            ,data= da4_5y, family= "binomial") # Psychosi s only has one level in this dataset
summary(fit6_y)
```

```
step(fit6_y)

fit6_sel<-glm(formula = ind_5y ~ Sex + ISS + HD + Osteoporosis + Diabetes +
                SA + Previous.orthopedic.trauma + t6m_b, family = "binomial",
              data = da4_5y)
summary(fit6_sel)
# stargazer(fit6_sel, title = "Predictors in reduced model (with early PWRE score)",
#           dep.var.lables = "Good outcome")

## Bootstrap ##
attach(da4_5y)
test<-function(formula,data,indices){             #function to  get the target statistics
  d<-data[indices,]
  fit<-glm(formula,data = d, family = binomial)
  return(fit$coeff)
}
## check single level covariates##

# lapply(da4_5y[c("SA", "Hypertension", "Osteoporosis" , "Diabetes" , "SA" , "Alcohol.abuse" ,
 "Revision.procedure" ,"Anxiety.disorder",
#                  "Age.at.injury" , "Malignancy" , "Stroke.TIA" ,"Sex", "Previous.orthopedic.tra
uma","HD" , "MI")], unique)
# bootresult<-boot(data<-da4_5y,statistic = test, R=1000000,formula = ind_5y~ Revision.procedure
+
#                     Age.at.injury + Sex +ISS + HD+
#                     Hypertension+Osteoporosis + Diabetes + SA + MI
#                   + Malignancy + Stroke.TIA + Previous.orthopedic.trauma +t6m_b) # calling boot
function
# boot.ci(bootresult,conf=0.95,index = 1,type = "bca")




### LME Without log-transformation ##
### 407 observations
clme<-cbind(c1,ct3m_b, ct6m_b, ct1_b, ct5_b )
names(clme)[51:54]<-c("3", "6", "12", "60")
clme_long<-melt(clme,id=c("X", "Revision.procedure"), measure.vars = c("3", "6", "12", "60"))
clme_long$variable <-as.numeric(as.character(clme_long$variable))
clme_long<- na.omit(clme_long)
names(clme_long)<-c("ID","Rd", "month", "score")
ggplot(clme_long,aes (x=month, y=score, group = ID)) + geom_line()+ geom_point()

write.csv(clme_long, file = "cl.csv")
# cfit.pre<-lmList(score~month|ID, data = clme_long)
# intervals(na.omit(fit.pre))
# plot(intervals(cfit.pre))

#### Base one original

dlme<-cbind(da,t3m_b, t6m_b, t1_b, t5_b )
names(dlme)[58:61]<-c("3", "6", "12", "60")

dlme_long<-melt(dlme,id=c("MRN", "Revision.procedure"), measure.vars = c("3", "6", "12", "60"))
```

```
dlme_long$variable <-as.numeric(as.character(dlme_long$variable))
dlme_long<- na.omit(dlme_long)
str(dlme_long)
names(dlme_long)<-c("ID","Rd", "day", "score")
dlme_long<-dlme_long%>% arrange(ID)

ggplot(dlme_long,aes (x=day, y=score, group = ID)) + geom_line()+ geom_point()
#
# fit.pre<-lmList(score~day|ID, data = dlme_long)
# intervals(na.omit(fit.pre))
# plot(intervals(fit.pre))

# fit.lme1<-lme(score~day, random = ~1|ID, data = dlme_long)
# fit.lme2<-lme(score~day+I(day^2), random = ~1|ID, data = dlme_long)
# fit.lme3<-lme(score~day, random = ~1+day|ID, data = dlme_long)
# fit.lme4<-lme(score~day+I(day^2), random = ~1+day|ID, data = dlme_long)
# summary(fit.lme1)

# GEE
# fit.uns <- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "unstructured")
# fit.ind<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "independence")
# fit.ex<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "exchangeable")

#fit.AR<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "AR-M")

# anova(fit.lme1, fit.lme3)
#plot (fit.lme1, resid (., type = "p") ~ fitted (.))
# within residue plot
# qqnorm(fit.lme4, ~resid(.))
# between individual errors
# qqnorm(fit.lme4, ~ ranef(.))
# summary(fit.uns)
# Estimate Naive S.E.   Naive z Robust S.E.   Robust z
# (Intercept) 18.4576377 0.87677791 21.051668  0.95643224  19.298427
# day         -0.1974098 0.01993359 -9.903376  0.01570122 -12.572901
# Rd1          7.3925277 2.30047056  3.213485  2.89142113   2.556711
# summary(fit.ind)
# Estimate Naive S.E.   Naive z Robust S.E.   Robust z
# (Intercept) 18.2265503  0.7347453 24.806624  0.96917859  18.806183
# day         -0.2001099  0.0225596 -8.870279  0.01688478 -11.851500
# Rd1          7.5355069  1.7524635  4.299951  2.88446347   2.612447
# summary(fit.ex)
# Estimate Naive S.E.    Naive z Robust S.E.   Robust z
# (Intercept) 18.7237308 0.84867716  22.062254  0.95372671  19.632176
# day         -0.1934191 0.01607502 -12.032277  0.01559903 -12.399430
# Rd1          7.7511678 2.54701469   3.043236  3.25985462   2.377765


# ##### log-score ####
#
# lgs<-log(dlme_long$score+0.49)
# dlme_long2<-cbind(dlme_long,lgs)
# dlme_long2<-na.omit(dlme_long2)
# lfit.lme1<-lme(lgs~day, random = ~1|ID, data = dlme_long2)
```

```
# lfit.lme2<-lme(lgs~day+I(day^2), random = ~1|ID, data = dlme_long2)
# lfit.lme3<-lme(lgs~day, random = ~1+day|ID, data = dlme_long2)
# lfit.lme4<-lme(lgs~day+I(day^2), random = ~1+day|ID, data = dlme_long2)
#
# qqnorm(lfit.lme1, ~resid(.))
# # between individual errors
# qqnorm(lfit.lme1, ~ ranef(.))




######## Add baseline as covariates in the LME #####

c1b<-c1
names(c1b)[c(27,32,37,42)]<-c("3", "6", "12", "60")
c1b_long<-melt(c1b,id=c("X", "Revision.procedure", "Total_Baseline"), measure.vars = c("3", "6",
"12", "60"))
c1b_long$variable <-as.numeric(as.character(c1b_long$variable))
c1b_long$Total_Baseline <-as.numeric(as.character(c1b_long$Total_Baseline))
c1b_long<- na.omit(c1b_long)
names(c1b_long)<-c("ID","Rd","Baseline", "month", "score")
lgs2<-log(c1b_long$score+0.49)
lgb<-log(c1b_long$Baseline+0.49)
c1b_long<-cbind(c1b_long, lgs2,lgb)
c1b_long<-c1b_long%>% arrange(ID)
ggplot(c1b_long[1:100,],aes (x=month, y=lgs2, group = ID)) + geom_line()+ geom_point()

#c2b with baseline as 0
c2b<-c1
names(c2b)[c(22,27,32,37,42)]<-c("0", "3", "6", "12", "60")
c2b_long<-melt(c2b,id=c("X", "Revision.procedure"), measure.vars = c("0","3", "6", "12", "60"))
c2b_long$variable <-as.numeric(as.character(c2b_long$variable))
#c2b_long$Total_Baseline <-as.numeric(as.character(c2b_long$Total_Baseline))
c2b_long<- na.omit(c2b_long)
names(c2b_long)<-c("ID","Rd", "month", "score")
clgs2<-log(c2b_long$score+0.01)

c2b_long<-cbind(c2b_long, clgs2)
c2b_long<-c2b_long%>% arrange(ID)
ggplot(c2b_long[1:60,],aes (x=month, y=clgs2, group = ID)) + geom_line()+ geom_point()+  labs(x
 = "Month", y = "Total score (in lgo-scale) ")+
  theme(axis.title.x = element_text(size = 15),axis.title.y = element_text(size= 15),
        plot.title = element_text(hjust = 0.5, size =20),
        axis.text = element_text(size = 15))


###### fitting by lme without transform ####
# bfit.lme1<-lme(score~month+Rd+Baseline, random = ~1|ID, data = c1b_long)
# bfit.lme2<-lme(score~day+I(day^2), random = ~1|ID, data = dlme_long)
# bfit.lme3<-lme(score~day, random = ~1+day|ID, data = dlme_long)
# bfit.lme4<-lme(score~day+I(day^2), random = ~1+day|ID, data = dlme_long)
# summary(bfit.lme1)

# anova(fit.lme1, fit.lme3)
```

```
#plot (fit.lme1, resid (., type = "p") ~ fitted (.))
# within residue plot
# qqnorm(bfit.lme1, ~resid(.))
# between individual errors
# qqnorm(bfit.lme1, ~ ranef(.))
# GEE
# bfit.uns <- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "unstructured")
# bfit.ind<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "independence")
# bfit.ex<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "exchangeable")

#fit.AR<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "AR-M")



### with log trans
lbfit.lme1<-lme(lgs2~month+lgb, random = ~1|ID, data = c1b_long)
lbfit.lme2<-lme(lgs2~month+lgb, random = ~1+month|ID, data = c1b_long)
lbfit.lme3<-lme(lgs2~month+Rd+lgb, random = ~1|ID, data = c1b_long)
lbfit.lme4<-lme(lgs2~month+Rd+lgb, random = ~1+month|ID, data = c1b_long)
summary(lbfit.lme1)
summary(lbfit.lme2)
summary(lbfit.lme3)
summary(lbfit.lme4)

anova(lbfit.lme1, lbfit.lme2)
anova(lbfit.lme3, lbfit.lme4)


# within residue plot
qqnorm(lbfit.lme4, ~resid(.))
# between individual errors
qqnorm(lbfit.lme4, ~ ranef(.))
# GEE
bfit.uns <- gee(lgs2~ month+lgb+Rd, id= ID, data= c1b_long, corstr = "unstructured")
bfit.ind<- gee(lgs2~ month+lgb+Rd, id= ID, data= c1b_long, corstr = "independence")
bfit.ex<- gee(lgs2~ month+lgb+Rd, id= ID, data= c1b_long, corstr = "exchangeable")
summary(bfit.uns)
summary(bfit.ind)
summary(bfit.ex)
#bfit.AR<- gee(lgs2~ month+Rd+lgb, id= ID, data= c1b_long, corstr = "AR-M")
```

LME modeling

```
### LME Without log-transformation ##
### 407 observations
clme<-cbind(c1,ct3m_b, ct6m_b, ct1_b, ct5_b )
names(clme)[51:54]<-c("3", "6", "12", "60")
clme_long<-melt(clme,id=c("X", "Revision.procedure"), measure.vars = c("3", "6", "12", "60"))
clme_long$variable <-as.numeric(as.character(clme_long$variable))
clme_long<- na.omit(clme_long)
names(clme_long)<-c("ID","Rd", "month", "score")
ggplot(clme_long,aes (x=month, y=score, group = ID)) + geom_line()+ geom_point()

write.csv(clme_long, file = "cl.csv")
# cfit.pre<-lmList(score~month|ID, data = clme_long)
# intervals(na.omit(fit.pre))
# plot(intervals(cfit.pre))

#### Base one original

dlme<-cbind(da,t3m_b, t6m_b, t1_b, t5_b )
names(dlme)[58:61]<-c("3", "6", "12", "60")

dlme_long<-melt(dlme,id=c("MRN", "Revision.procedure"), measure.vars = c("3", "6", "12", "60"))
dlme_long$variable <-as.numeric(as.character(dlme_long$variable))
dlme_long<- na.omit(dlme_long)
str(dlme_long)
names(dlme_long)<-c("ID","Rd", "day", "score")
dlme_long<-dlme_long%>% arrange(ID)

ggplot(dlme_long,aes (x=day, y=score, group = ID)) + geom_line()+ geom_point()

# fit.pre<-lmList(score~day|ID, data = dlme_long)
# intervals(na.omit(fit.pre))
# plot(intervals(fit.pre))

# fit.lme1<-lme(score~day, random = ~1|ID, data = dlme_long)
# fit.lme2<-lme(score~day+I(day^2), random = ~1|ID, data = dlme_long)
# fit.lme3<-lme(score~day, random = ~1+day|ID, data = dlme_long)
# fit.lme4<-lme(score~day+I(day^2), random = ~1+day|ID, data = dlme_long)
# summary(fit.lme1)
```

GEE and log-transformed LME

```
# GEE
# fit.uns <- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "unstructured")
# fit.ind<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "independence")
# fit.ex<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "exchangeable")

#fit.AR<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "AR-M")

# anova(fit.lme1, fit.lme3)
#plot (fit.lme1, resid (., type = "p") ~ fitted (.))
# within residue plot
# qqnorm(fit.lme4, ~resid(.))
# between individual errors
# qqnorm(fit.lme4, ~ ranef(.))
# summary(fit.uns)
# Estimate Naive S.E.   Naive z Robust S.E.   Robust z
# (Intercept) 18.4576377 0.87677791 21.051668  0.95643224  19.298427
# day         -0.1974098 0.01993359 -9.903376  0.01570122 -12.572901
# Rd1          7.3925277 2.30047056  3.213485  2.89142113   2.556711
# summary(fit.ind)
# Estimate Naive S.E.   Naive z Robust S.E.   Robust z
# (Intercept) 18.2265503  0.7347453 24.806624  0.96917859  18.806183
# day         -0.2001099  0.0225596 -8.870279  0.01688478 -11.851500
# Rd1          7.5355069  1.7524635  4.299951  2.88446347   2.612447
# summary(fit.ex)
# Estimate Naive S.E.    Naive z Robust S.E.   Robust z
# (Intercept) 18.7237308 0.84867716  22.062254  0.95372671  19.632176
# day         -0.1934191 0.01607502 -12.032277  0.01559903 -12.399430
# Rd1          7.7511678 2.54701469   3.043236  3.25985462   2.377765


# ##### log-score ####
#
# lgs<-log(dlme_long$score+0.49)
# dlme_long2<-cbind(dlme_long,lgs)
# dlme_long2<-na.omit(dlme_long2)
# lfit.lme1<-lme(lgs~day, random = ~1|ID, data = dlme_long2)
# lfit.lme2<-lme(lgs~day+I(day^2), random = ~1|ID, data = dlme_long2)
# lfit.lme3<-lme(lgs~day, random = ~1+day|ID, data = dlme_long2)
# lfit.lme4<-lme(lgs~day+I(day^2), random = ~1+day|ID, data = dlme_long2)
#
# qqnorm(lfit.lme1, ~resid(.))
# # between individual errors
# qqnorm(lfit.lme1, ~ ranef(.))




######## Add baseline as covariates in the LME #####

c1b<-c1
names(c1b)[c(27,32,37,42)]<-c("3", "6", "12", "60")
c1b_long<-melt(c1b,id=c("X", "Revision.procedure", "Total_Baseline"), measure.vars = c("3", "6",
"12", "60"))
```

```
c1b_long$variable <-as.numeric(as.character(c1b_long$variable))
c1b_long$Total_Baseline <-as.numeric(as.character(c1b_long$Total_Baseline))
c1b_long<- na.omit(c1b_long)
names(c1b_long)<-c("ID","Rd","Baseline", "month", "score")
lgs2<-log(c1b_long$score+0.49)
lgb<-log(c1b_long$Baseline+0.49)
c1b_long<-cbind(c1b_long, lgs2,lgb)
c1b_long<-c1b_long%>% arrange(ID)
ggplot(c1b_long[1:100,],aes (x=month, y=lgs2, group = ID)) + geom_line()+ geom_point()

#c2b with baseline as 0
c2b<-c1
names(c2b)[c(22,27,32,37,42)]<-c("0", "3", "6", "12", "60")
c2b_long<-melt(c2b,id=c("X", "Revision.procedure"), measure.vars = c("0","3", "6", "12", "60"))
c2b_long$variable <-as.numeric(as.character(c2b_long$variable))
#c2b_long$Total_Baseline <-as.numeric(as.character(c2b_long$Total_Baseline))
c2b_long<- na.omit(c2b_long)
names(c2b_long)<-c("ID","Rd", "month", "score")
clgs2<-log(c2b_long$score+0.01)

c2b_long<-cbind(c2b_long, clgs2)
c2b_long<-c2b_long%>% arrange(ID)
ggplot(c2b_long[1:60,],aes (x=month, y=clgs2, group = ID)) + geom_line()+ geom_point()+  labs(x
 = "Month", y = "Total score (in lgo-scale) ")+
  theme(axis.title.x = element_text(size = 15),axis.title.y = element_text(size= 15),
        plot.title = element_text(hjust = 0.5, size =20),
        axis.text = element_text(size = 15))


###### fitting by lme without transform ####
# bfit.lme1<-lme(score~month+Rd+Baseline, random = ~1|ID, data = c1b_long)
# bfit.lme2<-lme(score~day+I(day^2), random = ~1|ID, data = dlme_long)
# bfit.lme3<-lme(score~day, random = ~1+day|ID, data = dlme_long)
# bfit.lme4<-lme(score~day+I(day^2), random = ~1+day|ID, data = dlme_long)
# summary(bfit.lme1)
#
# anova(fit.lme1, fit.lme3)
#plot (fit.lme1, resid (., type = "p") ~ fitted (.))
# within residue plot
# qqnorm(bfit.lme1, ~resid(.))
# between individual errors
# qqnorm(bfit.lme1, ~ ranef(.))
# GEE
# bfit.uns <- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "unstructured")
# bfit.ind<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "independence")
# bfit.ex<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "exchangeable")

#fit.AR<- gee(score~ day + Rd, id= ID, data= dlme_long, corstr = "AR-M")



### with log trans
lbfit.lme1<-lme(lgs2~month+lgb, random = ~1|ID, data = c1b_long)
lbfit.lme2<-lme(lgs2~month+lgb, random = ~1+month|ID, data = c1b_long)
```

```
lbfit.lme3<-lme(lgs2~month+Rd+lgb, random = ~1|ID, data = c1b_long)
lbfit.lme4<-lme(lgs2~month+Rd+lgb, random = ~1+month|ID, data = c1b_long)
summary(lbfit.lme1)
summary(lbfit.lme2)
summary(lbfit.lme3)
summary(lbfit.lme4)

anova(lbfit.lme1, lbfit.lme2)
anova(lbfit.lme3, lbfit.lme4)


# within residue plot
qqnorm(lbfit.lme4, ~resid(.))
# between individual errors
qqnorm(lbfit.lme4, ~ ranef(.))
# GEE
bfit.uns <- gee(lgs2~ month+lgb+Rd, id= ID, data= c1b_long, corstr = "unstructured")
bfit.ind<- gee(lgs2~ month+lgb+Rd, id= ID, data= c1b_long, corstr = "independence")
bfit.ex<- gee(lgs2~ month+lgb+Rd, id= ID, data= c1b_long, corstr = "exchangeable")
summary(bfit.uns)
summary(bfit.ind)
summary(bfit.ex)
#bfit.AR<- gee(lgs2~ month+Rd+lgb, id= ID, data= c1b_long, corstr = "AR-M")
```

```
# c1 <- c1 %>%
#   mutate(Sex = Sex, across(Sex, ~if_else(.=="F", 0, 1))) %>%
#   mutate(CAD = CAD, across(CAD, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Hypertension = Hypertension, across(Hypertension, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Osteoporosis = Osteoporosis, across(Osteoporosis, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Diabetes = Diabetes, across(Diabetes, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Substance.abuse = Substance.abuse, across(Substance.abuse, ~if_else(.=="None", 0,
 1))) %>%
#   mutate(Alcohol.abuse = Alcohol.abuse, across(Alcohol.abuse, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Depression = Depression, across(Depression, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Anxiety.disorder = Anxiety.disorder, across(Anxiety.disorder, ~if_else(.=="None", 0,
 1))) %>%
#   mutate(Psychosis = Psychosis, across(Psychosis, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Malignancy = Malignancy, across(Malignancy, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Stroke.TIA = Stroke.TIA, across(Stroke.TIA, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Previous.orthopedic.trauma = Previous.orthopedic.trauma, across(Previous.orthopedic.t
rauma, ~if_else(.=="None", 0, 1))) %>%
#   mutate(Revision.procedure = Revision.procedure, across(Revision.procedure, ~if_else(.=="Remo
val of  device", 1, 0)))
```