Consulting Report on The Effectiveness of Intervention on Mental Distress

Naitong Chen

April 19, 2021

Abstract

In this report, change in mental distress over time is investigated in a group of participants that received mental health intervention and another control group that did not receive any treatment. The specific problems investigated are whether the participants' mental distresses decrease over time and whether the mental health intervention is effective in reducing the level of mental distress. Due to the longitudinal nature of the study, both the linear mixed effect models and the generalized estimating equation models are used to answer the two posed questions. It is found that while the participants' mental distresses in each of the two groups decrease significantly over time, mental health interventions do not seem to have a significant effect in reducing the level of mental distress.

1 Introduction

Nowadays, mental distress has become a common social issue and is affecting many people's quality of life. According to the American Psychiatric Association, up to nineteen (19) percent of adults in the United States experience some degree of mental illness [Ass21]. Amid the COVID-19 pandemic, the mental well-being of the general population has received an unprecedented amount of attention [TJ20]. It is therefore important to study the effectiveness of mental health interventions on reducing the level of mental distress.

This study investigates changes in participants' mental distress over time in an intervention group and a control group. Specifically, we try to answer whether the participants' mental distresses decrease over time and whether the mental health intervention is effective in reducing the level of mental distress.

This report begins by introducing the data recorded for this study and drawing preliminary conclusions through some explorative data analysis. The reliability of these preliminary conclusions are then checked through some model-based statistical analysis. Finally, we provide a discussion on the implication of the findings in this study. The R code used for this report can be found at `https://github.com/NaitongChen/STAT550IndividualProject`.

2 Exploratory Data Analysis

We begin by giving an overview of the data recorded in the study. A description of each variable is shown in Table 1. Note that a higher GSI score indicates a higher level of mental distress. There are five recordings of the GSI score (response) for each participant, measured respectively at the beginning of the study (0), and three (3), six (6), eighteen (18), and sixty (60) months later, as indicated by the month variable. While none of the other variables change over time, the repeated GSI measurements make this study longitudinal.

Of all 271 participants in the study, 57.6% were randomly placed in the intervention group and the remaining 42.4% were in the control group. Excluding those whose gender are unknown, 34.1% of the participants are male and 65.9% are female (Table 12). While the split between the two treatment groups is roughly balanced, there may be a slight underrepresentation of male participants in the data. However, since there are as many as 271 participants in total, there are still many participants that are male. Therefore there should not be a major impact of the unbalanced gender proportions.

SN	subject number
treatment	treatment received by each subject (1 for intervention and 2 for control)
month	measurement time (in month)
gender	gender of each subject (1 for male and 2 for female)
education	education received by each subject (in years)
GSI	Global Severity Index: an index indicating level of mental distress

Table 1: Description of all variables recorded in the study

The means and standard deviations of all continuous variables are shown in Table 2. We see that the GSI scores decreases over time while the corresponding standard deviations roughly stay consistent. Another observation is that the GSI scores are much smaller in magnitude compared to the education and month variables. This difference in scale may make the estimated effects of these explanatory variables on GSI unusually small. To aviod getting results of low interpretability, the GSI scores are scaled by a factor of ten (10) for the remainder of this report.

A common issue in longitudinal studies is missing data. As expected, with the GSI scores recorded over the span of five years, some of the response values are missing. This is common as some of the participants may have dropped out of the study for various reasons. In addition, some participants' gender and amount of education received are also missing. The count of missing values and the corresponding proportion of missing values are shown in Table 3. It is clear that more participants dropped out of the study as time went on. With the missing rates as high as 38.7%, it is important that we address the sensitivity of our subsequent analysis to these missing data in Section 3.

	education	GSI (0)	GSI (3)	GSI (6)	GSI (18)	GSI (60)
mean	13.705	1.125	1.036	0.854	0.834	0.780
sd	2.360	0.722	0.702	0.638	0.559	0.625

Table 2: Summary statistics for continuous variables

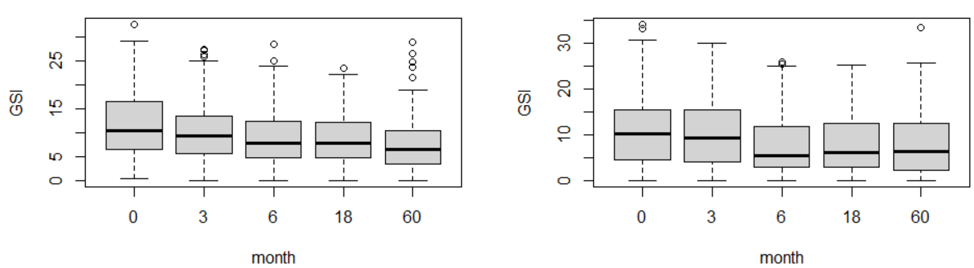
	gender	education	GSI (0)	GSI (3)	GSI (6)	GSI (18)	GSI (60)
count	4	7	10	38	52	105	98
proportion	0.015	0.026	0.037	0.140	0.192	0.387	0.362

Table 3: Missing rates of all variables in the study

2.1 Visualization of GSI

We now use side-by-side boxplots to get a rough idea of how the two main questions laid out in Section 1 can be answered. We first visualize the changes in the GSI scores over time for both the intervention group and the control group. Note that the p-values of the ANOVA and Kruskal-Wallis tests are also included in the caption for each of the two groups. In this case, ANOVA, as a generalization of the commonly known t-test, tests whether the mean GSI scores are the same across each of the five time points. Since ANOVA assumes the data to be normally distributed, which may not hold, the Kruskal-Wallis test is also conducted as a non-parametric alternative. The Kruskal-Wallis test does not make the normality assumption and is relatively robust against outliers, which can be seen to be present in Figure 1. The Kruskal-Wallis test here serves as a reference to check the reliability of the results from the ANOVA tests.

From Figure 1, we see that a downward trend of GSI over time is present in both groups. At the same time, the p-values from the ANOVA and Kruskal-Wallis tests (all < 0.021) indicate that there is moderate to strong evidence against that the mean GSI scores are the same across all five time points. It is then suggested that the participants' mental distresses in both groups decrease significantly over time.

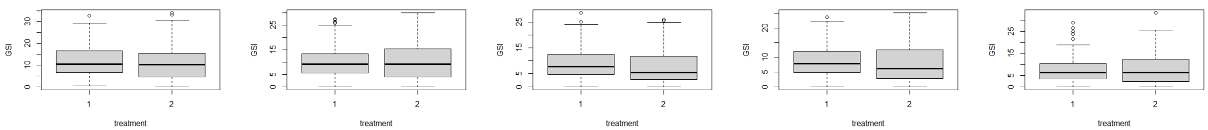


(a) intervention group (ANOVA: 0, Kruskal-Wallis: 0)

(b) control group (ANOVA: 0.021, Kruskal-Wallis: 0.006)

Figure 1: Side-by-side boxplots of the GSI scores across measurement times for each treatment group

To visualize whether the GSI scores are different between the intervention group and the control group at each of the five time points, we again use side-by-side boxplots. This is shown in Figure 2. Similarly, the p-values of two-sample t-tests and Wilcoxon tests are included in the caption to quantify the evidence we have against that the mean GSI scores are equal. Again, since the t-test assumes the data to be normally distributed and may be sensitive to outliers, the Wilcoxon test serves as a robust alternative. By the boxplots in Figure 2, the GSI scores of the two groups do not seem to differ much in any of the five times points. The p-values from the t-tests and Wilcoxon tests also indicate that, for the most part, there is no evidence against that the mean GSI scores between the two groups are the same.



(a) baseline (t-test: 0.457, Wilcoxon: 0.271) (b) 3 months (t-test: 0.995, Wilcoxon: 0.489) (c) 6 months (t-test: 0.332, Wilcoxon: 0.084) (d) 18 months (t-test: 0.405, Wilcoxon: 0.167) (e) 60 months (t-test: 0.826, Wilcoxon: 0.831)

Figure 2: Side-by-side boxplots of the GSI scores between two treatment groups across measurement times

2.2 Association between Explanatory Variables

Before summarizing our preliminary conclusions, it is worth inspecting the association between the two explanatory variables: gender and education. Since the gender variable is categorical, instead of computing the Pearson correlation coefficient, we use a side-by-side boxplot to study the association between the two variables Figure 3. Again, the p-values of t-test and Wilcoxon test are included in the caption. The plot suggests that the male participants as a group has received more education in terms of the number of years. The p-values (0.014 and 0.045) also indicates that there is moderate to strong evidence that the mean amount of education received in years among the male participants is different than that of the female participants. It is then possible that one of the two explanatory variables is enough to explain much of the variability present in the GSI scores of the corresponding participants. It is then worth inspecting whether including both of these explanatory variables leads to better fit models in Section 3.

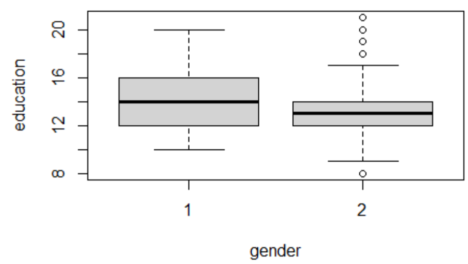


Figure 3: Side-by-side boxplot of education between the two genders (t-test: 0.014, Wilcoxon: 0.045)

2.3 Preliminary Conclusions

From the above graphical displays of the response variable, it is suggested that the participants' mental distresses in both the intervention group and the control group decrease over time. However, there does not seem to be a clear indication that the mental health intervention leads to a lower level of mental distress when compared to the control group. It is worth noting that the above preliminary conclusions are drawn without considering the gender and amount of education received of each participant. It is possible that these two explanatory variables exaggerate the effect of time or mask the effect of the mental health intervention on mental distress. Therefore, further analysis using various statistical models needs to be conducted to check the reliability of the conclusions drawn from the above exploratory data analysis.

3 Model-Based Statistical Analysis

Given the longitudinal nature of the study, in addition to considering the effect that the month and treatment variables have on reducing the level of mental distress across all participants, it is also important to address the correlation among the GSI scores obtained from each individual participant. The mixed-effects model is one of the most popular for analyzing longitudinal data for its intuitiveness and high interpretability. Particularly, it handles the individual-specific correlation by adding additional parameters at the individual level. Since the response (GSI) is continuous, and we have both continuous (month, education) and categorical (treatment, gender) explanatory variables, a natural model of choice is the linear mixed-effects model.

The linear mixed-effects model takes on the following general form:

$$y_{ij} = (\beta_0 + b_i^0) + (\beta_1 + b_i^1)t_{ij} + (\beta_2 + b_i^2)x_i^2 + \cdots + (\beta_n + b_i^n)x_i^n + e_{ij}. \quad (1)$$

In our case, y_{ij} denotes the GSI score of the ith participant at the jth time point. t_{ij} represents the month variable for the ith participant at the jth time point. $x_i^2, \cdots, x_i^n$ correspond to the other explanatory variables that do not change over time for the ith participant. Finally the e_{ij} terms are the residuals, which are assumed to be normally distributed and centred at 0.

By fixing j and ignoring the b_i terms, we recover the familiar linear regression model, where the β terms represent the effects of each explanatory variable on the response. By varying the time points j, we are able to model the effects of each explanatory variable considering the changes in the response over time. Including the b_i terms allows the individual-specific correlation to be taken into consideration. The b_i terms are the random effects. It is assumed that for each $k, b_i^k \sim N(0, \sigma_k^2)$ across all i.

Both of the questions that we would like to answer can be analyzed using different linear mixed-effects models. Specifically, by looking at the participants in the intervention group and the control group separately, the effect of the month variable on the GSI scores describes how mental distresses in each of the groups change over time. At the same time, by looking at both groups together, we can quantify the effect of mental health intervention by inspecting the β term associated with the treatment variable. As a result, for the remainder of this section, we present in detail the analysis procedure that addresses whether mental distress decreases over time in the intervention group as an example. The analysis results addressing both of our main questions are carefully discussed in Section 3.3.

Before proceeding to selecting the appropriate linear mixed-effects models that address our main objectives, it is worth noting that all of the entries that contain missing values, either on the explanatory variables or the response, are removed before conducting the following analysis. In addition, all of the participants with only one recorded GSI score are also excluded from the analysis below. This is because in such cases, there is no information on how the participant's mental distress has changed over time. Under the linear mixed-effects models, the removal of such observations implies that we are imposing the additional assumption that the reasons these values are missing do not depend on the missing values themselves. The missing values are more carefully treated in Section 3.4.

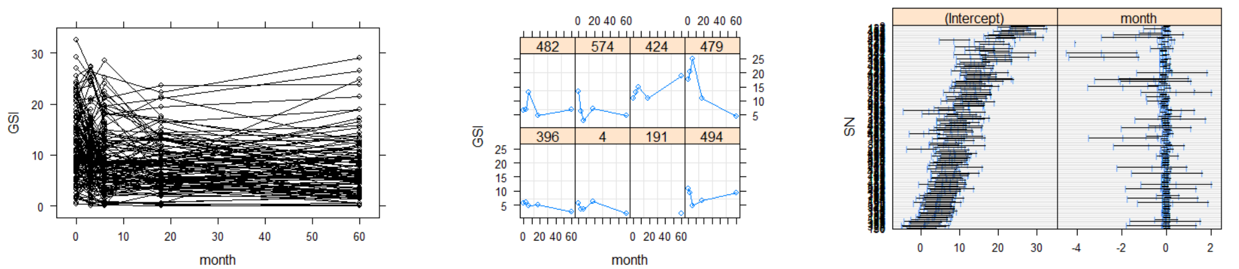
3.1 Model Selection

	month	month/gender/education
month/gender	0	0.016
month/education	0.004	0

Table 4: P values of Likelihood Ratio tests between models with different covariates under the intervention group

From Equation (1), it is clear that setting any of the β's to 0 implies the corresponding explanatory variable is not included in the linear mixed-effects model. At the same time, setting any set of b_i^k's across all i to 0 means that the random effects associated with the corresponding explanatory variable are not considered. A natural next step is then to identify the covariates and mixed effects that we would like to include in the model. By [Wu09], common model selection metrics such as the AIC, BIC, and the p-values from likelihood ratio tests (LRT) can be employed. Both the AIC and BIC try to find the balance between goodness-of-fit and model complexity. In our case, since the total number of explanatory variables considered is not large, we are not overly concerned with model complexity. Therefore, we use the p-values from LRT to select an appropriate model to answer whether the mental distresses decrease over time among participants in the intervention group. The LRT tests whether one model is a significantly better fit than the other between two nested models. By nested models, we mean that the parameters of the smaller model form a subset of those of the larger model.

We begin by selecting the explanatory variables to be included in the model. Since we're interested in the changes of mental distress over time, the month variable must be included. From Section 2, we know



(a) trellis plot of all subjects in the intervention group (b) trellis plot of randomly selected subjects (c) confidence intervals of parameters from individual linear models

Figure 4: Diagnostic plots for selection of random effects for the intervention group

that it is possible that one of the gender and education variables can explain much of the variabilities among the GSI scores. Therefore we compare the goodness-of-fit among models that include varying covariates using the LRT. The p-values are summarized in Table 4. Note that none of the mixed effects is considered here. In Table 4, the first column shows there is strong evidence (p-values < 0.004) that including either of the gender or education variables leads to a better fit model. Then the second column shows there is moderate to strong evidence (p-values < 0.016) including both covariates leads to a better fit model compared to those that only include one. As a result, we do not discard any of the covariates when studying the change in mental distress among participants in the intervention group. Similarly, as shown in Appendices B.1 and B.2, none of the covariates is discarded. Note that in Appendix B.2, the treatment variable is also included automatically because its effect is essential to detecting any effect of mental health intervention.

Now that the covariates to be included are determined for each model, the random effects can be selected following the same approach. The random effects allows us to adjust the effect that a explanatory variable or the intercept has for each participant on top of the overall average effect. Before formally deciding which random effects are to be included using the LRT, we use the plots in Figure 4 to gauge whether the inclusion of any random effects is warranted.

	no mixed effect	intercept	intercept/month
intercept	0	/	/
intercept/month	/	0.028	/
intercept/gender	/	0.784	/
intercept/education	/	0.998	/
intercept/month/gender	/	/	0.893
intercept/month/education	/	/	0.995

Table 5: P values of Likelihood Ratio tests between models with different mixed effects under the intervention group

By Figure 4a, it is clear that at the beginning of the study, the participants' levels of mental distresses are quite different in the intervention group. To better assess how the GSI scores change over time, we randomly select some of the participants and show how their GSI scores change over the course of the study in Figure 4b. While many of the participants' GSI scores exhibit an overall flat trend, there are participants whose GSI scores show clear increasing or decreasing trends acorss time. To better access the trends across time among all participants in the intervention group, we fit linear regression models without random effects on each of the participants' GSI scores individually and plot the 95% confidence intervals of the regression parameters in Figure 4c. This plot verifies the variability present in the GSI scores at the beginning of the study as many of the confidence intervals of the individual intercepts do not overlap. While most confidence intervals of the individual regression parameters for the month variable hover around 0, we do have two instances where the confidence intervals are much different than the rest. These plots suggest that a random effect term is necessary for both the intercept and the month variable.

It is important to note, though, that each participant's gender and education variables do not change over the course of the study. Therefore when fitting individual linear regression models on each of the participants, the effect of the gender and education variables cannot be distinguished from the intercept. As a result, we only know that a random effect term is necessary for some of the intercept and the gender and education variables. To formally decide where to place the random effects among the intercept and the two covariates, we again use the p-values from the LRT.

As shown in Table 5, a stepwise procedure is used when comparing models that contain different random effect terms. We first note that adding a random effect term on the intercept leads to better fit model (p-value ≈ 0). Then among the three covariates, there is moderate evidence that adding a random effect on the month variable leads to a better fit (p-value $= 0.028$). Finally, there is no evidence against that further including random effects on either gender or education results in the same level of goodness-of-fit (p-values > 0.893). Therefore, when studying the changes in mental distress over time in the intervention group, a model that includes random effects on the intercept and the month variable is most appropriate.

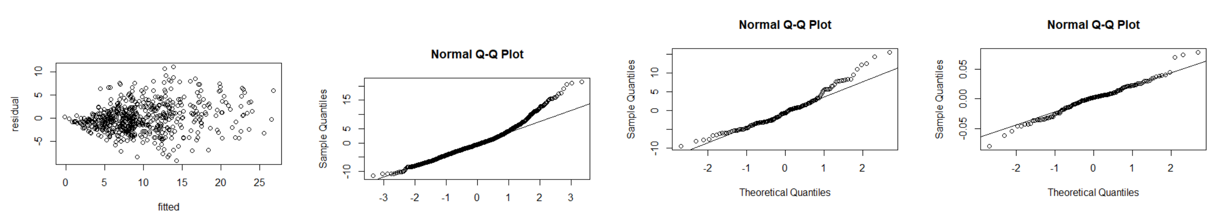
Following the same procedure, as shown in Appendix B.3, a random effect term is only added to the intercept term when studying the changes in mental distress over time in the control group. In Appendix B.4, we see that random effect terms are added to both the intercept and the month variable.

3.2 Assumption Check

Before presenting the analysis results from the linear mixed-effects models selected above, recall that there are two distributional assumptions. Particularly, both the residuals and the random effects corresponding to each explanatory variable are assumed to be normally distributed with mean 0. Again as an example, we check whether these assumptions hold for the case of studying the changes in mental distress in the intervention group using the plots in Figure 5.

Figure 5a shows the scatter plot between the fitted values and the residuals. Similar to Section 2, p-values from the t-test and Wilcoxon test quantifying the evidence against that the mean of the residuals is 0 are shown in the caption. We see that residuals seem to have a mean of 0. Figure 5b shows the residual QQ plot. If the residuals were indeed normally distributed, the dots in the plot should follow the slanted line closely. We see that this is clearly not the case for the larger residuals. This indicates that the normality assumption on the residuals may not hold. The QQ plots for both groups of the mixed effects are shown in Figures 5c and 5d. Similar to that of the residuals, while there is no evidence against that the means are 0 (p-values > 0.335), the normality assumptions seem to be violated. Overall, it seems like the normality assumptions of the linear mixed-effects model do not hold. This is also the case for the models investigating the changes in mental distress over time in the control group and the effectiveness of the treatment (Appendix C). Therefore, we run the risk of drawing misleading conclusions if we draw conclusions only based off this set of linear mixed-effects models.

It is important to note that the normality assumption violations seem to be caused by a good portion of the responses rather than a small fraction of outliers. A common remedy in this case is to apply a log transformation of the response variables. However, this does not seem to help with the current set of data.



(a) fitted value vs. residual (t-test: 1, Wilcoxon: 0.204)

(b) residual QQ plot

(c) QQ plot for the random effects on the intercept (t-test: 1, Wilcoxon: 0.335)

(d) QQ plot for the random effects on the slope (t-test: 1, Wilcoxon: 0.781)

Figure 5: Visualizing the residuals and random effects of the LME model under the intervention group

As a result, in addition to the linear mixed-effects models, a set of Generalized Estimating Equation (GEE) models are also conducted. The GEE model is commonly considered a non-parametric alternative to the mixed-effects model in longitudinal analysis. While the GEE model does not allow for individual-specific inference, there is no distributional assumptions that need to be satisfied for the model to produce reliable results. The GEE model under the linear model framework can be thought of as a generalization to the least squares method for estimating the regression parameters, which directly finds the set of parameters that minimizes the residual sum of squares without imposing the normality assumption. The removal of the normality assumption makes the GEE model more robust against the residuals' deviation from the normal distribution. Therefore, by if the analysis results are consistent between the linear mixed-effects models and the GEE models, we know that the analysis results are more reliable.

Note that for the GEE models to hold with the missing data removed, we do require that the missing values are missing completely at random. In other words, the reason for missing depends on neither the missing value itself nor any of the other variables in the model. Again, without knowing whether this assumption holds, we check how sensitive our analysis results are to these missing values in Section 3.4. Finally, as a required input for the GEE models, we need to specify a covariance structure of the responses. An unstructed covariance is selected in all of our models, which, as shown in the next section, provides very robust parameter estimates.

3.3 Analysis results

3.3.1 Changes in Mental Distress over Time

	Value	Std.Error	DF	t-value	p-value
(Intercept)	11.933	2.758	465	4.326	0.000
month	-0.047	0.008	465	-5.671	0.000
gender2	2.764	0.925	141	2.990	0.003
education	-0.249	0.190	141	-1.314	0.191

Table 6: Output of Linear Mixed Model under the intervention group

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	11.162	2.484	4.494	2.538	4.397
month	-0.047	0.010	-4.477	0.008	-5.852
gender2	2.827	0.834	3.391	0.869	3.253
education	-0.194	0.170	-1.146	0.173	-1.125

Table 7: Output of GEE model under the intervention group

	Value	Std.Error	DF	t-value	p-value
(Intercept)	19.235	4.151	308	4.634	0.000
month	-0.020	0.008	308	-2.394	0.017
gender2	2.606	1.383	95	1.884	0.063
education	-0.822	0.273	95	-3.015	0.003

Table 8: Output of Linear Mixed Model under the control group

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	19.267	3.433	5.613	3.229	5.967
month	-0.027	0.013	-2.019	0.010	-2.606
gender2	2.220	1.148	1.935	1.216	1.825
education	-0.809	0.225	-3.596	0.203	-3.994

Table 9: Output of GEE model under the control group

3.3.2 Effectiveness of Mental Health Intervention

	Value	Std.Error	DF	t-value	p-value
(Intercept)	15.418	2.341	774	6.586	0.000
treatment2	-0.193	0.731	238	-0.264	0.792
month	-0.037	0.006	774	-5.893	0.000
gender2	2.737	0.776	238	3.527	0.001
education	-0.516	0.157	238	-3.295	0.001

Table 10: Output of Linear Mixed Model

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	14.685	2.046	7.176	2.053	7.154
treatment2	-0.430	0.641	-0.671	0.735	-0.586
month	-0.039	0.008	-4.666	0.006	-6.040
gender2	2.693	0.680	3.961	0.716	3.763
education	-0.455	0.136	-3.337	0.139	-3.278

Table 11: Output of GEE model

3.4 Handling Missing Data

[LR19]

4 Conclusions and Discussion

References

[Ass21] A. P. Association. *What Is Mental Illness?* https://www.psychiatry.org/patients-families/what-is-mental-illness. Online; accessed 18-April-2021. 2021.

[LR19] R. J. Little and D. B. Rubin. *Statistical analysis with missing data*. Vol. 793. John Wiley & Sons, 2019.

[TJ20] J. M. Twenge and T. E. Joiner. "Mental distress among US adults during the COVID-19 pandemic". In: *Journal of Clinical Psychology* 76.12 (2020), pp. 2170–2182.

[Wu09] L. Wu. *Mixed effects models for complex data*. CRC press, 2009.

A Summary statistics for categorical variables

	treatment		gender
intervention	0.576	male	0.341
control	0.424	female	0.659

Table 12: Summary statistics (proportion) for categorical variables

B Additional Model Selection Tables and Plots

B.1 Covariate Selection for The Control Group

	month	month + gender + education
month + gender	0	0
month + education	0	0

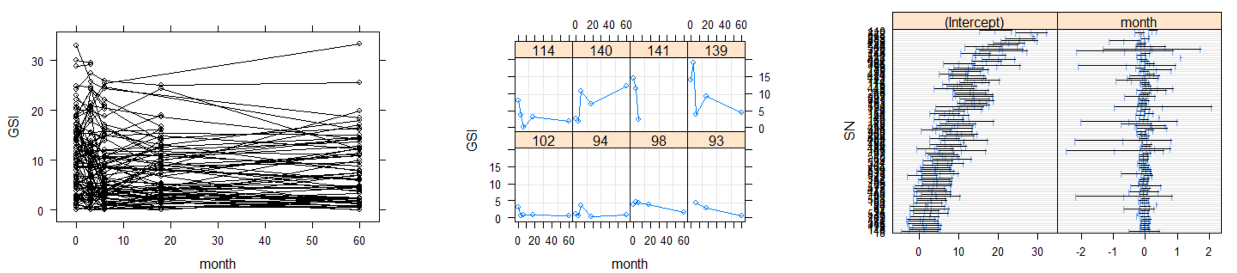
Table 13: P values of Likelihood Ratio tests between models with different covariates under the control group

B.2 Covariate Selection for Both Groups Combined

	treatment + month	treatment + month + gender + education
treatment + month + gender	0	0
treatment + month + education	0	0

Table 14: P values of Likelihood Ratio tests between models with different covariates

B.3 Random Effect Selection for The Control Group



(a) trellis plot of all subjects in the control group (b) trellis plot of randomly selected subjects (c) confidence intervals of parameters from individual linear models

Figure 6: Diagnostic plots for selection of random effects for the control group

	no mixed effect	intercept
intercept	0	/
intercept + month	/	0.166
intercept + gender	/	0.638
intercept + education	/	0.981

Table 15: P values of Likelihood Ratio tests between models with different mixed effects under the control group

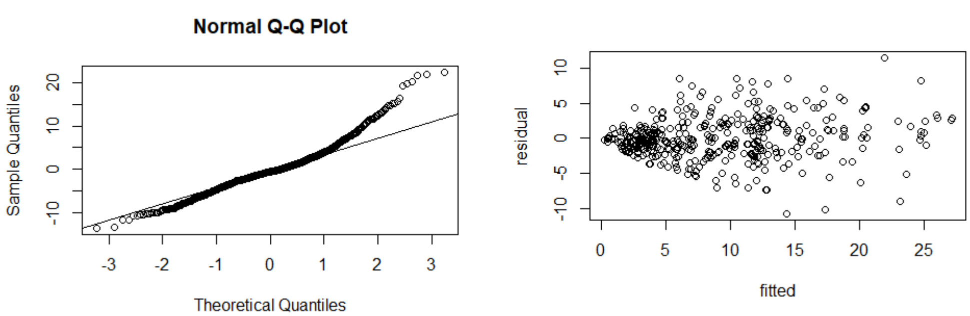
B.4 Random Effect Selection for Both Groups Combined

	no mixed effect	intercept	intercept + month
intercept	0	/	/
intercept + month	/	0.005	/
intercept + treatment	/	0.208	/
intercept + gender	/	0.487	/
intercept + education	/	0.999	/
intercept + month + treatment	/	/	0.343
intercept + month + gender	/	/	0.690
intercept + month + education	/	/	0.997

Table 16: P values of Likelihood Ratio tests between models with different mixed effects

C Additional Assumption Check Plots

C.1 Control Group



(a) residual QQ plot

(b) fitted value vs. residual (t-test: 1, Wilcoxon: 0.501)

Figure 7: Visualizing the residuals of the LME model under the control group

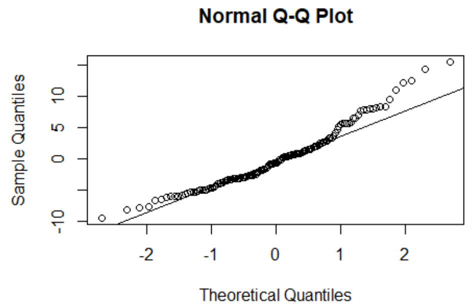
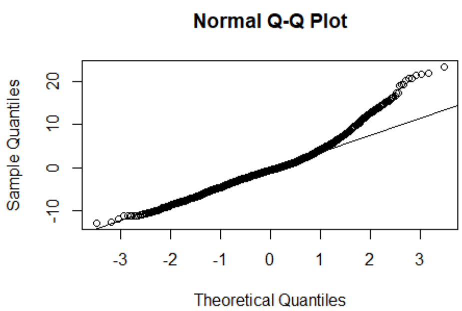
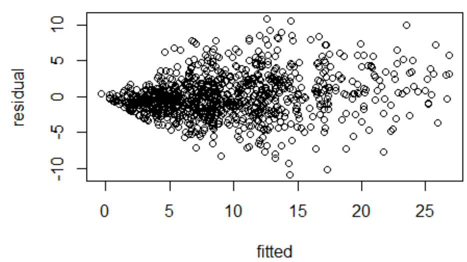


Figure 8: QQ plot for the random effects on the intercept (t-test: 1, Wilcoxon: 0.405)

C.2 Both Groups Combined

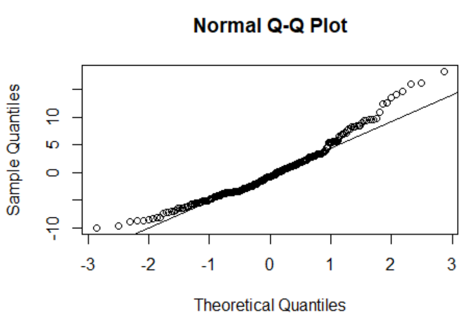


(a) residual QQ plot

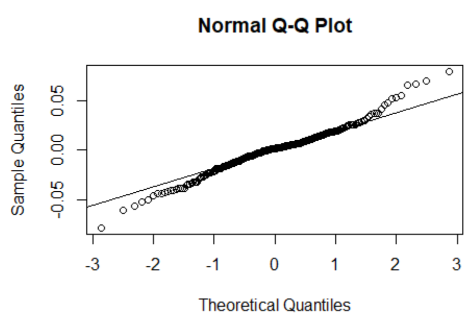


(b) fitted value vs. residual (t-test: 1, Wilcoxon: 0.143)

Figure 9: Visualizing the residuals of the LME model



(a) QQ plot for the random effects on the intercept (t-test: 1, Wilcoxon: 0.207)



(b) QQ plot for the random effects on the slope (t-test: 1, Wilcoxon: 0.786)

Figure 10: Visualizing the random effects

D Pooled Analysis Results from Multiple Imputation

D.1 Intervention Group

	Estimate	Std.Error	t.value	df	P value
(Intercept)	11.526	2.809	4.104	231.840	0.000
month	-0.045	0.011	-3.999	23.557	0.001
gender2	2.734	0.968	2.824	239.207	0.005
education	-0.222	0.199	-1.115	107.937	0.268

Table 17: Output of pooled Linear Mixed Model under the intervention group

	Estimate	Std.Error	t.value	df	P value
(Intercept)	11.102	2.789	3.981	158.677	0.000
month	-0.048	0.012	-4.090	12.698	0.001
gender2	2.733	0.942	2.902	355.509	0.004
education	-0.188	0.198	-0.947	78.666	0.347

Table 18: Output of pooled GEE model (naive) under the intervention group

	Estimate	Std.Error	t.value	df	P value
(Intercept)	11.102	2.692	4.124	137.732	0.000
month	-0.048	0.012	-3.939	14.758	0.001
gender2	2.733	0.908	3.009	307.277	0.003
education	-0.188	0.191	-0.985	67.256	0.328

Table 19: Output of pooled GEE model (robust) under the intervention group

D.2 Control Group

	Estimate	Std.Error	t.value	df	P value
(Intercept)	19.695	4.011	4.910	1786.900	0.000
month	-0.032	0.012	-2.571	10.407	0.027
gender2	2.277	1.315	1.732	5112.727	0.083
education	-0.828	0.268	-3.096	802.808	0.002

Table 20: Output of pooled Linear Mixed Model under the control group

	Estimate	Std.Error	t.value	df	P value
(Intercept)	20.598	4.119	5.000	208.709	0.000
month	-0.027	0.014	-1.999	14.750	0.064
gender2	1.821	1.328	1.371	440.506	0.171
education	-0.870	0.280	-3.103	105.337	0.002

Table 21: Output of pooled GEE model (naive) under the control group

	Estimate	Std.Error	t.value	df	P value
(Intercept)	20.598	3.695	5.575	135.062	0.000
month	-0.027	0.014	-1.949	16.329	0.069
gender2	1.821	1.302	1.399	406.382	0.163
education	-0.870	0.242	-3.590	58.813	0.001

Table 22: Output of pooled GEE model (robust) under the control group

D.3 Both Groups Combined

	Estimate	Std.Error	t.value	df	P value
(Intercept)	14.858	2.439	6.091	135.244	0.000
treatment2	-0.171	0.733	-0.234	802.931	0.815
month	-0.039	0.009	-4.167	15.060	0.001
gender2	2.604	0.829	3.139	108.962	0.002
education	-0.467	0.172	-2.715	58.348	0.009

Table 23: Output of pooled Linear Mixed Model

	Estimate	Std.Error	t.value	df	P value
(Intercept)	14.644	2.449	5.980	98.809	0.000
treatment2	-0.223	0.724	-0.307	832.892	0.759
month	-0.039	0.011	-3.678	8.707	0.005
gender2	2.591	0.810	3.197	140.824	0.002
education	-0.446	0.173	-2.582	48.524	0.013

Table 24: Output of pooled GEE model (naive)

	Estimate	Std.Error	t.value	df	P value
(Intercept)	14.644	2.356	6.215	84.692	0.000
treatment2	-0.223	0.742	-0.300	917.614	0.764
month	-0.039	0.011	-3.594	9.554	0.005
gender2	2.591	0.786	3.298	124.401	0.001
education	-0.446	0.168	-2.655	43.400	0.011

Table 25: Output of pooled GEE model (robust)

E R Code Excerpt (Intervention Group)

```
# load packages
library(nlme)
library(gee)
library(corrplot)
library(lattice)
library(lmtest)
library(mitml)

# load data
dat = read.csv("data_dep.txt", TRUE, " ")

# removing variables not in the description
# the remaining variables are:
```

```r
# subject #, treatment, gender, education, month, GSI
dat = dat[,c(1,2,3,5,10,20)]

# make SN, treatment and gender factors
dat[,2] = as.factor(dat[,2])
dat[,3] = as.factor(dat[,3])

# scale GSI by 10
dat[,6] = 10 * dat[,6]

dat = as.data.frame(dat)
colnames(dat)[2] = "treatment"
colnames(dat)[3] = "gender"
colnames(dat)[4] = "education"

# EDA

# summary statistics
# treatment
print(length(which(dat$treatment == 1)) / length(dat$SN))
print(length(which(dat$treatment == 2)) / length(dat$SN))
# gender (ignore missing)
print(length(which(dat$gender == 1)) / (length(which(dat$gender == 1)) +
                                        length(which(dat$gender == 2))))
print(length(which(dat$gender == 2)) / (length(which(dat$gender == 1)) +
                                        length(which(dat$gender == 2))))
# education
print(mean(dat$education, na.rm = T))
print(sd(dat$education, na.rm = T))
# GSI
# month = 0
print(mean(dat$GSI[which(dat$month == 0)], na.rm = T))
print(sd(dat$GSI[which(dat$month == 0)], na.rm = T))
# month = 3
print(mean(dat$GSI[which(dat$month == 3)], na.rm = T))
print(sd(dat$GSI[which(dat$month == 3)], na.rm = T))
# month = 6
print(mean(dat$GSI[which(dat$month == 6)], na.rm = T))
print(sd(dat$GSI[which(dat$month == 6)], na.rm = T))
# month = 18
print(mean(dat$GSI[which(dat$month == 18)], na.rm = T))
print(sd(dat$GSI[which(dat$month == 18)], na.rm = T))
# month = 60
print(mean(dat$GSI[which(dat$month == 60)], na.rm = T))
print(sd(dat$GSI[which(dat$month == 60)], na.rm = T))

# missing rates
# gender
print(sum(is.na(dat[,3]))/5)
print((sum(is.na(dat[,3]))/5)/(length(dat[,1])/5))
# education
print(sum(is.na(dat[,4]))/5)
print((sum(is.na(dat[,4]))/5)/(length(dat[,1])/5))
# GSI
# month = 0
print(sum(is.na(dat$GSI[which(dat$month == 0)])))
print(sum(is.na(dat$GSI[which(dat$month == 0)]))/(length(dat[,1])/5))
# month = 3
print(sum(is.na(dat$GSI[which(dat$month == 3)])))
print(sum(is.na(dat$GSI[which(dat$month == 3)]))/(length(dat[,1])/5))
# month = 6
print(sum(is.na(dat$GSI[which(dat$month == 6)])))
print(sum(is.na(dat$GSI[which(dat$month == 6)]))/(length(dat[,1])/5))
# month = 18
print(sum(is.na(dat$GSI[which(dat$month == 18)])))
print(sum(is.na(dat$GSI[which(dat$month == 18)]))/(length(dat[,1])/5))
# month = 60
```

```r
print(sum(is.na(dat$GSI[which(dat$month == 60)])))
print(sum(is.na(dat$GSI[which(dat$month == 60)]))/(length(dat[,1])/5))

# boxplots over time
dat_control = dat[which(dat$treatment == 2),]
boxplot(GSI ~ month, dat_control)
summary(aov(GSI ~ month, dat_control))
kruskal.test(GSI ~ month, dat_control)$p.value

dat_trt = dat[which(dat$treatment == 1),]
boxplot(GSI ~ month, dat_trt)
summary(aov(GSI ~ month, dat_trt))
kruskal.test(GSI ~ month, dat_trt)$p.value

# boxplots between groups
dat_t1 = dat[which(dat$month == 0),]
boxplot(GSI ~ treatment, dat_t1)
t.test(GSI ~ treatment, dat_t1)
wilcox.test(GSI ~ treatment, dat_t1)

dat_t2 = dat[which(dat$month == 3),]
boxplot(GSI ~ treatment, dat_t2)
t.test(GSI ~ treatment, dat_t2)
wilcox.test(GSI ~ treatment, dat_t2)

dat_t3 = dat[which(dat$month == 6),]
boxplot(GSI ~ treatment, dat_t3)
t.test(GSI ~ treatment, dat_t3)
wilcox.test(GSI ~ treatment, dat_t3)

dat_t4 = dat[which(dat$month == 18),]
boxplot(GSI ~ treatment, dat_t4)
t.test(GSI ~ treatment, dat_t4)
wilcox.test(GSI ~ treatment, dat_t4)

dat_t5 = dat[which(dat$month == 60),]
boxplot(GSI ~ treatment, dat_t5)
t.test(GSI ~ treatment, dat_t5)
wilcox.test(GSI ~ treatment, dat_t5)

# association between covariates
# keep one row from each SN
inds = seq(1, length(dat[,1]), 5)
dat_single = dat[inds,]

boxplot(education ~ gender, dat_single)
t.test(education ~ gender, dat_single)
wilcox.test(education ~ gender, dat_single)

# confirmatory data analysis
# group data
dat_cc = na.omit(dat)
dat_cc = dat_cc[-c(1:21),]
dat_cc_g = groupedData(GSI ~ month | SN, data = dat_cc)
dat_cc_trt = dat_cc[which(dat_cc$treatment == 1),]
dat_cc_trt_g = groupedData(GSI ~ month | SN, data = dat_cc_trt)

# covariate selection
model_base = lm(GSI ~ month, data = dat_cc_trt)
model_d1 = lm(GSI ~ month + gender, data = dat_cc_trt)
model_d4 = lm(GSI ~ month + education, data = dat_cc_trt)
model_both = lm(GSI ~ month + gender + education, data = dat_cc_trt)

lrtest(model_base, model_d1)
lrtest(model_base, model_d4)

lrtest(model_d1, model_both)
```

```
lrtest(model_d4, model_both)

# diagnostic plots
xyplot(GSI ~ month, group = SN, data = dat_cc_trt, col="black", type="b")
plot(dat_cc_trt_g[335:370,])

fit_lm = lmList(GSI ~ month | SN, dat_cc_trt_g)
plot(intervals(fit_lm))

# selecting random effects
model_base = lm(GSI ~ month + gender + education, data = dat_cc_trt)
model_intercept = lme(GSI ~ month + gender + education,
                      random= ~ 1 | SN, data = dat_cc_trt_g,
                      control = lmeControl(opt='optim'))
model_month = lme(GSI ~ month + gender + education,
                  random= ~ month | SN, data = dat_cc_trt_g,
                  control = lmeControl(opt='optim'))

model_d1 = lme(GSI ~ month + gender + education,
               random= ~ gender | SN, data = dat_cc_trt_g,
               control = lmeControl(opt='optim'))

model_d4 = lme(GSI ~ month + gender + education,
               random= ~ education | SN, data = dat_cc_trt_g,
               control = lmeControl(opt='optim'))

model_d1d4 = lme(GSI ~ month + gender + education,
                 random= ~ gender + education | SN, data = dat_cc_trt_g,
                 control = lmeControl(opt='optim'))

model_monthd1 = lme(GSI ~ month + gender + education,
                    random= ~ month + gender | SN, data = dat_cc_trt_g,
                    control = lmeControl(opt='optim'))

model_monthd4 = lme(GSI ~ month + gender + education,
                    random= ~ month + education | SN, data = dat_cc_trt_g,
                    control = lmeControl(opt='optim'))

anova(model_intercept, model_base)

anova(model_month, model_intercept) # winner
anova(model_d1, model_intercept)
anova(model_d4, model_intercept)

anova(model_monthd1, model_month)
anova(model_monthd4, model_month)

# fit lme
model_month = lme(GSI ~ month + gender + education,
                  random= ~ month | SN, data = dat_cc_trt_g,
                  control = lmeControl(opt='optim'))
summary(model_month)$tTable

# lme assumption check
qqnorm(model_month$residuals)
qqline(model_month$residuals)
fitted = fitted(model_month)
residual = resid(model_month)
plot(fitted, residual)
t.test(resid(model_month))
wilcox.test(resid(model_month))

qqnorm(ranef(model_month)[,1])
qqline(ranef(model_month)[,1])
t.test(ranef(model_month)[,1])
wilcox.test(ranef(model_month)[,1])
```

```
qqnorm(ranef(model_month)[,2])
qqline(ranef(model_month)[,2])
t.test(ranef(model_month)[,2])
wilcox.test(ranef(model_month)[,2])

# fit gee
gee_trt <- gee(GSI ~ month + gender + education, SN, corstr = "unstructured",
               data = dat_cc_trt)
summary(gee_trt)$coef

# multiple imputation
dat_to_impute = dat[-c(1:130),]
type_vec = c(-2, 2, 1, 1, 3, 1)
imp = jomoImpute(dat_to_impute, type = type_vec, seed = 1, m = 5)
implist <- mitmlComplete(imp)

implist_trt = implist
for (i in 1:5) {
  implist_trt[[i]] = implist[[i]][which(implist[[i]]$treatment == 1),]
}

# pool lme
lme.p1.control <- with(implist_con,
                       lme(GSI ~ month + gender + education,
                           random= ~ 1 | SN,
                           control = lmeControl(opt='optim')))
testEstimates(lme.p1.control)

# pool gee
gee.p1.treatment <- with(implist_trt,
                         gee(GSI ~ month + gender + education, SN,
                             corstr = "unstructured"))
qhat <- sapply(gee.p1.treatment, coef)

# robust
uhat <- qhat
for (i in 1:5) {
  uhat[,i] = diag(gee.p1.treatment[[i]]$robust.variance)
}

testEstimates(qhat = qhat, uhat = uhat)

# naive
uhat <- qhat
for (i in 1:5) {
  uhat[,i] = diag(gee.p1.treatment[[i]]$naive.variance)
}

testEstimates(qhat = qhat, uhat = uhat)
```