An analysis of growth of the community integration psychological score in an ethnically diverse population experiencing homelessness in a permanent supportive housing program using hierarchical mixed modeling

Leah Hollis Puglisi
The University of New Mexico

Leah Puglisi
Candidate

Mathematics and Statistics
Department

This thesis is approved, and it is acceptable in quality and from for publication:

Approved by the Thesis Committee:

Erik Erhardt, Chairperson

Ronald Christensen

Fletcher Christensen

AN ANALYSIS OF GROWTH OF THE COMMUNITY INTEGRATION PSYCHOLOGICAL SCORE IN AN ETHNICALLY DIVERSE POPULATION EXPERIENCING HOMELESSNESS IN A PERMANENT SUPPORTIVE HOUSING PROGRAM USING HIERARCHICAL MIXED MODELING

by

LEAH PUGLSI

BS MATHEMATICS, THE UNIVERSITY OF NEW MEXICO

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science

Statistics

The University of New Mexico
Albuquerque, New Mexico

December, 2020

DEDICATION

To my family and friends who were there to support me through finishing my masters in the midst of a pandemic and all the wild happenings of 2020.

ACKNOWLEDGEMENTS

A very special thank you to Dr. Erik Erhardt for your advisement and for all the work and patience you had with me.

To Dr. Ronald Christensen and to Dr. Fletcher Christensen for being on my committee.

To the UNM HSC CBH team that helped guide me and allowed me to use the data and analysis as thesis material.

To Dr. Li Li for all your initial help and consideration.

To my dogs, Ollie and Poppy, for always being a perfect distraction.

An analysis of growth of the community integration psychological score in an ethnically diverse population experiencing homelessness in a permanent supportive housing program using hierarchical mixed modeling

by

Leah Puglisi

BS, Mathematics, The University of New Mexico, 2017
MS, Statistics, The University of New Mexico, 2020

ABSTRACT

Hierarchical models are becoming increasingly common in epidemiological and psychological research. When analyzing data from such studies, the nested structure of the data must be taken into account. Mixed modeling in conjunction with hierarchical mixed modeling allows researchers to ask broad questions about the population of interest. Modeling under restricted maximum likelihood estimation (REML), as opposed to full maximum likelihood estimation (ML), increases the accuracy of estimates for the random effects in the model. We use hierarchical mixed modeling under REML estimation to analyze which factors increase "community integration", a concept and a construct developed and used in the mental health service sector focusing on bettering the development of personal, social, and vocational competency of individuals experiencing homelessness enrolled in permanent supportive housing (PHS) programs. "Community integration" takes central importance because it has been shown to improve individual quality of life. From increasing their chances of finding employment to expanding their social networks in order to lessen the psychological trauma of homelessness, tracking community integration is critical to understanding the overall experience of homeless individuals. Combining organizational research theory and growth modeling theory into a 3-level model illustrates the hierarchy of variables affecting the Community Integration Scale (CIS) Psychological score. Individual growth trajectories, variation in client growth parameters at each site, and variation between sites are represented, respectively, within the first, second, and third levels of the model. For individuals experiencing homelessness, CIS Psychological scores increase most drastically between intake and 6 months, with few contextual variables affecting this change.

Key words: hierarchical model, mixed effects, REML, CIS Psychological score, individual growth

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background

During the past three years, I was part of the statistical team at the University of New Mexico School of Medicine, Department of Psychiatry, Division of Community Behavioral Health (CBH). As implied by the name, much of the analyses done and reported by CBH are from social and behavioral health science research. I am involved in the Housing Supports, Health, and Recovery for Homeless Individuals (HHRHI) project. This project investigates community integration among an ethnically diverse population with mental illness or substance use disorders who are in a permanent supportive housing program. “Community integration” is a concept and a construct developed and used in the mental health service sector focusing on bettering the development of personal, social, and vocational competency of individuals experiencing homelessness in order to live as independently as possible in their own homes and communities (Community Integration and Mental Health).

A significant part of our work is organization research. Organizational research addresses attributes of organizations holistically rather than of the individuals within organizations (Raudenbush and Bryk, 2002). The organizations we consider in this thesis are homeless shelters. We use information provided by shelter residents to ascertain information about factors that will better support community integration. We also have collected data on these individuals over multiple time points. Thus, theory of individual growth trajectory models is also used, to assess how individual growth (development of personal, social, and vocational competency) within these shelters are affected, and by which factors.

Often in such research, there is the time dependence of the data. When this is the case, one can measure individual growth over the specified time period (also known in statistics as "repeated measures"). In organizational research, both observable and latent parameters are used to study growth. In our CBH research, we are interested in the growth of feelings of community integration among a study population undergoing a specific treatment. The study population comprises individuals experiencing homelessness in New Mexico; the treatment is housing in various homeless shelters throughout the duration of the study (almost two years). We use a clinically-administered survey to assess which of several factors contribute to a feeling of community integration. The survey measures a cumulative "Community Integrations Scale" (CIS) score. The results of our research are intended to inform social reform policy and programs in New Mexico. We are contributing to a body of research focusing on the cultural welfare of homeless communities and individuals around the state.

Our data consist of nested, time-invariant variables. This structure informs our choice of statistical methods, which is known as "hierarchical modeling". Specifically, time is nested in the individuals participating in the study, and individuals, or "clients", are nested in homeless shelters, or "sites." Hierarchical modeling — also known as multilevel modeling — has become a popular statistical tool over the past three decades (Gelman and Pardoe, 2006; Raudenbush and Bryk, 2002). Hierarchical linear models (HLM) are a generalization of a generalized linear model in which data are nested due to dependencies that are defined a priori (before we observe the measurements) (Gelman 2006). HLMs are popular in many disciplines, including social science, education, political science, public health, and sociology (Bickel 2007).

Nested data are often referred to as clustered data because one could take each level of the nested data as a cluster, or each individual (observation) as a cluster within time (Huang 2018). Examples of nested data include children nested within schools, schools nested within school districts, patients in a hospital, states in a country, and, often in educational and social science research, several observations over time within each person. In our study, we have up to three time observations nested within each person and groups of people nested within one of three geographically separated sites. Many developmental research data sets collect multiple sets of observations on people over a period of time. (Raudenbush and Bryk, 2002).

There are many forms of nested data and modeling. For example, a Cluster Randomization Trial (CRT) considers each time point as a cluster, and clusters are the unit of randomization (Austin 2007). Our study does use time-of-survey as a clustering unit, but also uses the repeated measures of individuals over time. This means that we have multiple measures of a single individual, violating the assumption of observation independence which most regression models require (Rutterford et al., 2015), but HLMs take this into consideration.

We have both within-client and within-site dependencies in our study. Failing to consider this dependency structure of our nested data would result in erroneous conclusions in hypothesis testing and inference (Bickel 2007). Such errors might include under- or over-estimation of standard errors, causing inflation or deflation (respectively) in Type-1 error for regression coefficients. The resultant regression coefficients would, in turn, have larger or smaller confidence intervals or prediction intervals for new observations (McNeish and Stapleton, 2016). Statistical simulations have shown that ignoring dependencies in a data sets can impact the estimated variances and the available power to detect the treatment of covariate effects (Bell et al., 2013).

Hierarchical linear models are often conceptualized as several regression models occurring at different cluster levels. The Interclass Correlation Coefficient (ICC) is commonly reported as a measure of how subjects are correlated within a cluster and can help answer questions about variance-covariance components such as between-subject and within-subject variance (Wang, Xie, and Fisher, 2012). We use the ICC to assess the variance at all levels: unconditional, time, client, site, and the full level model with all the contextual variables from the three levels to see if variance is diminishing when more effects are added.

1.2 Objective

The goal of the research presented in this thesis is to provide evidence-based recommendations for policies affecting the homeless populations in New Mexico by understanding the factors affecting a person's sense of integration within a community. HLMs are used because we are assessing community integration scores over time in individuals experiencing homelessness who are in a permanent supportive housing program. We use this data to build an HLM by interpreting individuals' community integration self-reported score given factors such as support groups, education, income and other common factors seen in similar studies.

This thesis presents a brief discussion on latent variable models, as hierarchical modeling is innately latent variable modeling – meaning there are unobserved variables that can be detected by their effects on variables that are observable. The data analysis performed was done on data collected at sites (homeless shelters) in New Mexico from 2016-2018 in Santa Fe, Bernalillo, and Las Cruces counties. The analyses used the software SAS PROC MIXED (SAS/STAT® 14.1 User's Guide: The MIXED Procedure) to obtain frequentist estimations and prediction while using restricted maximum likelihood (REML) instead of full maximum likelihood (ML). As such, the predictions and estimations are based on model-based imputations, which is a latent measure, but are made through frequentist methodologies. The results and implications of our models are assessed and discussed. The attrition of the subjects in the sample is discussed in subsequent chapters. Lastly, a discussion on the recommendations for policy reform to foster community integration is overviewed. There are similarities to previous research considered in this thesis, as well.

1.3 Literature Review

Research results support that modest increases in community integration are rooted in disability rights, the notion of common citizenship, and was one of the driving forces behind deinstitutionalization of people experiencing homelessness (Crisanti et al., 2021). This was supported through the observation of individuals recovering from traumatic brain injuries (TBIs), serious mental illnesses (SMI) or substance abuse disorders (SUD), and their more considerable increase in quality of life when people were living together and being integrated into the community (Crisanti et al., 2021). Furthermore, Crisanti et al., (2021) report that the absence of social support can be a risk factor for those with PTSD symptoms. Poor health status can be exacerbated by people experiencing homelessness, and this population of people has increased premature mortality rates of 3 to 4 times that of the general population (Crisanti et al., 2021). People experiencing homelessness also tend to have limited social networks, and because social functioning and social relationships are tied to health, these relationships are common interest in social and behavioral studies (Umberson and Montez 2010; Crisanti et al., 2021).

Wong and Solomon (2002) found that the need for a sense of community was greater among women than men and was related to higher personal income and lower depression scores. Yanos (2007) found that higher education and length of time in one's current residence was associated with better community integration. Baumgartner and Herman (2012) found that psychiatric symptoms were negatively related to physical and social integration. Community integration has been measured with various instruments using numerous variables among different populations, but generally include physical integration, social embeddedness, and psychological embeddedness (Wong and Solomon 2002). The literature shows a high interest in community integration and in the ties it has to the homeless population and their welfare (Crisanti et al., 2021). While the data were recorded for the participants, few aspects show an impact on the self-reported CIS scores.

Multiple regression (MR) is often used to evaluate interactions between variables, test hypotheses, and estimate parameters. The use of the techniques in multiple regression has recently been extended to hierarchical linear modeling and latent curve analysis (LCA) (Preacher et al., 2006). Fitting models with a nested structure, such that each cluster or group has its' own and different regression model, was technically not possible up until about the 1990s (Raudenbush and Bryk, 2002).

Since then, computational techniques and software (such as SAS/STATA) have enabled the fitting of more complex models (Raudenbush and Bryk, 2002). Hierarchical modeling is used when the data is nested, and the parameters themselves are being modeled at each level. This means that our data could be fit under Bayesian theories and modeling if desired. In recent decades, many software has been developed to make such analyses more applicable and their estimates more reliable (Bickel 2007).

Generalized linear models are fixed effect models and are known to be the most common type model in the social and behavioral sciences (Garson, 2013; Bryk and Raudenbush, 1987) Mixed models have both random and fixed effects, and the effect is both if they contribute to the intercept and covariance structure for that model (Hand et al., 2011). Hierarchical linear modeling is a type of linear mixed model focused on the differences between groups and within groups. Random intercept models are those where only the intercept of the level-1 dependent variable is modeled as an effect of the level-2 grouping variable and other level covariates (Christensen et al., 2011).

Statistical methods make predictions, test hypotheses, and estimate parameters. Hierarchical modeling can perform these functions when modeling nested data (Bickle 2007). If instruments measure what they are designed to measure, then researchers are able to address questions about those measures. Our overall level of success depends on the quality of our data and the model, and the ability to measure the features in the model accurately.

Chapter 2: Methods

2.1. Measures and sample

In this section we introduce the study design, participants, and descriptions of the data. Data were derived from measures that evaluated the effectiveness of the Substance Abuse and Mental Health Services Administration (SAMHSA) and Center for Substance Abuse Treatment (CSAT) grant that funded a Permanent Supportive Housing (PSH) program. Three of the primary survey instruments were the Community Integration Scale (CIS citation), the Post-Traumatic Symptom Checklist - Civilian Version (PCL-C), and the Government Performance and Results Act (GPRA) (GPRA Measurement Tools). These surveys were collected through face-to-face structured interviews conducted by research assistants. The target population included chronically homeless individuals who had a severe mental illness (SMI) or a substance use disorder (SUD) diagnosis. Eligible individuals were identified through a Coordinated Assessment (Gardner et al., 2010) with a focus on selecting individuals with the greatest need due to length of housing instability and behavioral health needs. Behavioral health diagnoses were determined by a master-level independent licensed counselor through a structured face-to-face clinical interview when participants entered services.

Participants were enrolled in the evaluation between February 2016 and September 2018. They were recruited from three community-based agencies that specialize in providing behavioral health services for the homeless in the large metropolitan New Mexico areas: Albuquerque, Las Cruces, and Santa Fe. All participants were provided a housing voucher from several different state, local, and federal agencies. Of the 453 participants, 370 completed the intake interview, 286 completed a 6-month interview (77% retention), and 143 completed a discharge interview (39% retention).

Individuals that met eligibility requirements needed to provide informed consent. Those who did not consent to evaluation and data collection received integration services through the provider agencies, but were not considered in evaluation. Baseline interviews were completed within one week of enrollment in clinical services. Each participant was to complete two follow-up interviews, while the average number of interviews completed was 2.2. The first follow-up interview was completed at approximately six-months post-baseline. A discharge follow-up interview was conducted after a clinician not having any contact with the client for 90 days or more. This is, we have a convenience sample based on inclusion criteria.

2.1.1 Variables

As we are investigating the factors affecting community integration for people experiencing homelessness, we aim to gather information on variables that foster these individuals' social, personal and vocational skills while in a permanent supportive housing program. The variables collected on these individuals are those that come from the SAMSAH survey described above. The analysis conducted throughout the HHRHI project has several groups of variables. There are a series of covariates and two response variables. The response variables are the CIS Psychological and Physical scores. The cluster variables (Site, Client, Time) account for the nested structure of the observations and are either treated as random effects (Site and Person) or longitudinally (Time). The remaining covariates are treated as fixed main effects and found through statistical models.

Covariates were selected based on relevant literature on Community Integration, PSH, and clinical insight (Crisanti et al., 2021). Working with subject matter experts (SMEs) is crucial in deciding which variables to include in the models built and assessed in this study. These variables also happened to have low rates of missing data and had equal weight in their factors, so our data subset is balanced well. For example, substance use is commonly measured in studies of PSH and was measured in this evaluation but could not be included in this analysis, as it is known to be a misreported characteristic. Similarly, housing status is another variable commonly included in studies of populations in PSH programs but not included here, as individuals were given housing for participating, which is their treatment.

Cluster Variables

The repeated measures indicator variable, Time, indicates at what point a client took a survey: at intake (Time=0), 6-months (Time=1), and discharge (Time=2). The variable Client ID (Client) matches each subject to a set of surveys, and the number of observations per client over time could be 1, 2, or 3. The Site variable records the location (homeless shelter) where the client was housed: St. Luke's in Las Cruces, St. Martins in Albuquerque, and Life Link in Santa Fe. The Site and Client variables are both indicator variables modeled as random effects, while Time is treated as a factor variable modeled as fixed.

Response Variables

The Community Integration Scale (CIS; Goering et al., 2011) was used to derive measures of perceived physical and psychological integration. In a larger study, we also examined the response variable known as CIS Physical (community integration scale physical score) using the same data and variables we used to model CIS Psychological (community integration scale psychological score). In this study, we model only the CIS Psychological response variable. Both of these variables are discrete and measured on a point scale, 0 to 7 for the physical score, and 4 to 20 for the psychological score. The initial questions for the psychological score are Likert measured and then converted to a numeric variable based on how often one answered 1-5: 1 being “Strongly Disagree” and 5 being “Strongly Agree,” where there was an option to “decline” or “do not know,” treated as neutral-3. The questionnaires for both measures use the past 30 days for a reference time, and higher scores indicate higher self-perception of community integration. The questions asked intended to measure how the participant is acting in their broader community on both a physical and psychological scale. The questions for the CIS Psychological score summarized answers to questions like “Do you know who lives near you, interact with them, feel at home, like where you belong?”

Covariate Variables

The covariates used are either categorical (factor) or numeric variables. The variables treated as numeric are PTSD Score, Health Status, Age, and Income. The factor variables are Gender, Ethnicity (Hispanic/non-Hispanic), Education, Mental Health Diagnosis, self-help group Attendance, and Interaction with Family or Friends (IFF). Diagnosis is recorded by clinical staff based on diagnostic interviews, including five exclusive levels: bipolar disorder, PTSD, depressive disorders, schizophrenia spectrum disorders, and a grouped anxiety and "others" category. Education is self-reported and has three levels: less than high school (LT HS), high school (HS), and more than high school (MT HS). Ethnicity is dichotomous, recorded as 1 for Hispanic and 0 otherwise. Gender is dichotomous, recorded as 1 for male and 0 for female. These variables are from the SAMSAH and CSAT dataset and codebook, except for Diagnosis, which comes from the clinician.

The PTSD scale summarizes Likert responses to 17 questions numerically for the analysis. PTSD symptom severity is measured by the Post-Traumatic Stress Disorder Checklist (Civilian) (PCL-C) (Weathers et al. 1993). The PCL-C checklist is a self-report scale that measures core PTSD symptoms in the past month, and each item ranges from 1 ("Not at all") to 5 ("Extremely"). Responses to the 17 items were summed to yield a total severity score ranging from 17 to 85, where higher scores indicate greater symptom severity.

The GPRA interview included three questions about participation in recovery-related activities, which is the Attend variable in the study. Each question is rated as the number of groups attended in the past 30 days: voluntary self-help groups (e.g., Alcoholics Anonymous), religious/faith-affiliated recovery self-help groups, and other support/recovery groups. Attendances were tallied as a numeric measure of participation in recovery-related activities (0, 1, 2, or 3) and treated as a categorical variable in the analysis.

Overall Health Status was measured by responses to one item which asked: "How would you rate your health right now?". Responses were rated on a five-point Likert scale, ranging from 1 ("Poor") to 5 ("Excellent") and were then turned into a numeric variable on a 1-to-5 scale.

IFF measures the interaction degree with family and friends based on a yes/no question that asked: "In the past 30 days, did you have any interaction with family or friends that are supportive of your recovery", and is treated as a binary factor variable. The recovery aspect is that from a SMI or SUD diagnosis.

2.2 Statistical Methods

This section describes the statistical methods and theory used to assess the aims of the study. We do this to fully understand why and how the analysis uses the modeling scheme described in this thesis. An introduction to latent variable modeling and hierarchical modeling is discussed. An introduction of the estimator REML and a brief comparison of REML to full ML is discussed to understand why such an estimator is used. Other modeling designs, such as choice of covariance structure and choosing whether to model variables as fixed or random effects, are discussed. Lastly, we review how we handle missing data and attrition.

2.2.1 Hierarchical Mixed Models

We fit a hierarchical model with mixed effects, so we note that the properties of the fixed effects of multiple regression apply to our model, as random effect estimators are a special case of fixed effects estimators (Mundlak, 1978). A common way of finding such estimates is through what is called Ordinary Least Squares (OLS), a type of linear least-squares method for estimating the unknown parameters in linear regression models (White 1980). The fixed effects are interpreted similarly to OLS regression coefficients (Bickel 2007). Hierarchical linear models are OLS regression-based but violate the independence assumption of the errors. We can't use OLS estimators as we have effects related to each other, i.e., we may have effects related to other covariates because of the nested structure of our data. A fundamental parameter to estimate is the variance (Christensen 2016), which is one primary use of the hierarchical model so that we understand where the variance in the data is explained by the model.

Hierarchical modeling can be thought of as many regressions over different levels, so it is essential to have a baseline understanding of multiple linear regression. Since regressions are linear combinations of all the parameters in our model, we assume we have a hierarchical linear model (linear in the model parameters, not that the individual growth being studied is linear). An everyday use of regression analysis maximizes the likelihood function of our probability distribution, given the data, to find estimates of the parameters in our model. With hierarchical modeling, we could say something about the effects of say, Clients, even though we do not have information about all individuals experiencing homelessness because the variance components allow us to predict the likely effect size of non-sample homeless individuals. We use the information we obtain from the covariance parameters of these individuals experiencing homelessness, to make inferences from the model. We can say the same for the Sites.

Not only do hierarchical models use OLS regression theory, but they also use analysis of variance theory for testing the categorical factor effects. The ANOVA's primary purpose is to analyze the ratio of variance between groups and within groups, where groups are defined by each factor variable in a model (Christensen 2016). The null hypothesis for ANOVA tests is that the group means are all equal. This type of analysis tells us the differences between group means, which factors are significant, how the factors differ, and whether there are further group interactions to be analyzed. This is important in the study to determine which factors are contributing to individual growth, over time. We also note that we are treating

Time as factor and use the ANOVA to understand where the most significant individual CIS Psychological score growth is occurring.

2.2.1.1 Latent Variable and Growth Modeling Introduction

As the analysis presented in this thesis focuses on the use of hierarchical modeling for individual CIS Psychological score growth, we will describe how latent variable models theory is inherent in hierarchical modeling. In the developmental research models and latent variable models, the model aims to assess individual growth (Bickel 2007). Latent models are a class of multivariate outcome models that are useful when data are incomplete and in missing data problems when covariates are missing at random (Raudenbush and Bryk, 2002). In our study we have instances where the covariates had missing values. In a latent variable model, the complete data are those that are considered as observed and augmented by the unobserved (Raudenbush and Bryk, 2002). This unobserved imputed term is the latent variable. The observed and incomplete data are used to estimate the association among the latent variables (Raudenbush and Bryk, 2002).

Hierarchical models with data that are longitudinal and time dependent, may be regarded explicitly as latent variable models (Raudenbush and Bryk, 2002). Hierarchical models are latent variable models, with a specified number of levels. These models represent the probability of a given response as a function of characteristics of the item and latent “traits” of a person (Raudenbush and Bryk, 2002). This idea of a latent trait can be represented as a 2-level model where the item responses, traits, are nested within persons (Raudenbush and Bryk, 2002). In the 2-level model, the level-1 of the hierarchy represents the associations between the observed incomplete data and the complete latent data, and the level-2 model describes the distribution of the latent variables (Raudenbush and Bryk, 2002). The observed data is to make inferences about population distributions of latent variables (Raudenbush and Bryk, 2002).

Multiple model-based imputations are a solution to missing data problems and can produce realistic standard errors and unbiased inferences about the parameters generating the complete data (Raudenbush and Bryk, 2002). Hierarchical models have an advantage in that no matter how incomplete the data is of the dependent variable(s), the analysis and estimations are not more complicated, as long as the data are missing at random (Maas and Snijders, 2003). In this modeling schema, missing observations in the covariates can be omitted or imputed and allow the analysis of repeated measured data to be done under less strict assumptions on the covariance matrix (Maas and Snijders, 2003).

When referring to individual growth, we may think of it as a function of another latent growth parameter. The use of some observed value can be used to predict status or growth. The level-1 model usually is described as a simple linear model for growth:

$$Y_{it} = \pi_{0i} + \pi_{1i}X_{ti} + \varepsilon_{ti} \qquad (2.1)$$

Where Y_{it} is the growth of individual i at time t, X_{ti} is the observed covariates variable. The π_{0i} is the initial status of growth of student i at some starting point and π_{1i} is the unit growth rate over some time period. The ε_{ti} are the level-1 residuals that are assumed normally distributed with mean 0 and constant variance σ^2.

The level-2 model, models the level-1 parameters; the intercept and often the means, if they are random slopes. This means the initial status and rate of growth vary across the population of subjects as a function of the covariate and subject-specific random effects. The following model describes such a relationship:

$$\pi_{0i} = \beta_{00} + \beta_{01}X_i + \mu_{0i} \quad (2.2)$$

$$\pi_{1i} = \beta_{10} + \beta_{11}X_i + \mu_{1i} \quad (2.3)$$

where X_i is an indicator or dummy variable for some covariate (often gender) and we assume the random effects are bivariate normal with variance τ_{00} and τ_{11} with covariance τ_{01}.

We are using growth modeling as we measure individual growth at different time points within different organizations. When measuring growth, it is vital to notice that there will often be missing data; for example, students miss class, people miss work, or for whatever reason, one is to miss when data is collected. There may be a large amount of missing data at different time points being evaluated for different individuals. This creates an unbalanced dataset. In general, growth models do not require observations to have complete data for all time points, nor do they require the time points to be equally spaced out across observations (Bell et al., 2013). However, there must be multiple observations for each subject. Note that the models of growth trajectory models look very similar to the latent variable model presented in equations *2.1-2.3*, above. When the number of observations per individual is few, and when the time period is relatively short, linear individual growth models are most convenient in estimation and inference (Raudenbush and Bryk, 2002).

2.2.2 Developing our Hierarchical Model

The data this hierarchy is modeling is clustered; that is, the dependent variable is measured once for each subject, but the subjects themselves are somehow grouped (time nested into clients, for example). There is no ordering to the subjects within the group, so their responses should be equally correlated. In repeated measures data, the dependent variable is measured more than once for each subject. Usually, there is some covariate (often called a within-subject factor) that changes with each measurement. These data are longitudinal, as the dependent variable is measured at several time points for each subject. We say that our repeated measures data is clustered.

Repeated measures data has a hierarchical structure, as data collected from multiple individuals are nested in time groups or multiple measures across different times are nested within individuals. When analyzing data that has a hierarchical structure, standard regression techniques will not produce accurate results due to the lack of independent observations (Raudenbush and Bryk, 2002). Hierarchical linear models are being used often as a more appropriate way to analyze repeated measurements. A hierarchical linear model can be expressed as a linear mixed model (LMM). Assumptions are made on the distribution of the errors of the model and the covariance matrix of the repeated measurements. The range of assumptions can be simple or complex, restrictive, or unrestricted (Raudenbush and Bryk, 2002). Increasingly complex data calls for increasingly complex models. A linear mixed model can be expressed generally as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta_0} + \boldsymbol{Z}\boldsymbol{\beta_1} + \boldsymbol{\varepsilon} \qquad (2.4)$$

where $\boldsymbol{y}= [y_1, \dots, y_n]$ is an $n \times 1$ vector of observations, $\boldsymbol{\beta_0}$ is a $p \times 1$ vector of parameters, $\boldsymbol{\beta_1}$ is a $m \times 1$ vector of random effects and $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector of errors. $\boldsymbol{X}$ is an $n \times p$ design matrix and $\boldsymbol{Z}$ is an $n \times m$ design matrix such that:

$$\boldsymbol{Z} = \begin{bmatrix} 1 & \mathrm{z}_{11} & \dots & \mathrm{z}_{1\mathrm{m}} \\ 1 & \dots & \dots & \cdot\cdot \\ 1 & \mathrm{z}_{\mathrm{n}1} & \dots & \mathrm{z}_{\mathrm{nm}} \end{bmatrix}.$$

It is often assumed that $\boldsymbol{\beta_1} \sim \text{Norm}(0, \mathbf{G})$, $\boldsymbol{\varepsilon} \sim \text{Norm}(0, \mathbf{R})$ such that $\boldsymbol{\beta_1}$ and $\boldsymbol{\varepsilon}$ are uncorrelated.

LMM’s are best used to assess models that have a continuous outcome. HLMs, also called hierarchical mixed models (HMM), extend linear models by adding random effects to control for any heterogeneity observed between clusters while retaining fixed effects of primary interest. Note we will continue to refer to our model as a HMM. All hierarchical models that have a continuous outcome can be written in the form of LMM-and for the analysis presented our outcome is continuous. In the case of the continuous outcomes, HMM can be much more general since the matrix $\boldsymbol{Z}$ can be much more general and unique (Hox, 2002).

The algorithm for REML can fit the general LMM case, so it also fits the HMM case. When Raudenbush and Bryk discuss "HLMs," they are explicit that they mean mixed models where the random factors are nested and are not crossed. For simplicity, we will discuss the hierarchical model in terms of the full model used in assessment in this chapter, and use the reduced inferential model in Chapter 3.

One can measure the reliability of initial status and change. We first estimate the unconditional model to investigate the psychometric characteristics of the estimated individual growth parameters. One might falsely conclude that there are no relations when there is no measure of reliability. One can also measure the correlation between individual change and initial status. There are often negative correlations between initial status and growth rates due to measurement errors at the pre-test.

We begin with an unconditional model; a model with no covariates. In the unconditional model, we assess the between-subject variation and to deem that it be large enough to use hierarchical modeling. We use these covariance values to calculate the ICC, which tells us the amount of between- and within-subject variation. The unconditional model is as follows:

$$CIS\_PSY_{tis} = \beta_{000} + \gamma_{0s} + \delta_{0s} + \varepsilon_{tis}, \quad \varepsilon_{0is} \sim N(0, \sigma^2) \qquad (2.5)$$

β_{000} is the overall random intercept and value of the unconditional CIS Psychological score;

γ_{0s} is the random effect of Client;

δ_{0s} is the random effect of Site;

ε_{tis} is the random error effect;

where $t = 0,1,2$ for time the survey was taken which corresponds to intake, 6 months, and discharge, respectively;

$i = 1_s, \dots, n_s$ clients which there are 370 of them at intake, 286 at 6 months, and 143 at discharge;

$s = 1,2,3$ for the 3 different sites of homeless shelters in the larger metropolitan New Mexico cities.

If the unconditional model's intercept is significant, we can appropriately assume that subjects vary in their growth at initial assessment or intake. The intercept also tells us the average value on the outcome at time 0 (intake).

The three-level hierarchical model with a random intercept we fit is as follows:

Level-1 Time sub-model:

$$CISPSY_{tis} = \boldsymbol{\gamma_{0is}} + \beta_1 HealthStatus_{tis} + \beta_2 Attend_{tis} + \beta_3 IFF_{tis} + \beta_4 Income_{tis} + \beta_5 PTSD_{tis} + \beta_{6t} HealthStatus_{tis} * Time_{tis} + \beta_{7t} Attend_{tis} * Time_{tis} + \beta_{8t} IFF_{tis} * Time_{tis} + \beta_{9t} Income_{tis} * Time_{tis} + \beta_{At} PTSD_{tis} * Time_{tis} + \beta_{Bt} Time_{tis} + \varepsilon_{tis} \qquad (2.6)$$

where t, i, s denotes the cluster indices for Time, Client, and Site, respectively;

$\boldsymbol{\gamma_{0is}}$ is a random intercept including the random effects of both Clients and Sites;

β_1 is the fixed slope for Health Status;

β_2 is the fixed slope for Attend;

β_3 is the fixed slope for IFF;

β_4 is the fixed slope for Income;

β_5 is the fixed slope for the grand mean centered PTSD Score;

β_{6t}through β_{At} are the fixed slopes for all the covariates interacting with the Time variable where the βs are expanded below like the β_{Bt};

β_{Bt} is a fixed slope for Time;

and the error term ε_{tis}, such that $\varepsilon_{tis} \sim N(0, \sigma^2)$.

Note: $\beta_{Bt} Time_{ts} = \beta_{B1} I(Time = 1) + \beta_{B2} I(Time = 2);\ \beta_{B0} = 0$

This model is the full model at the level-1 Time sub-model of the hierarchy. We use this to assess how the variables are changing over time. This model has all the time invariant variables. We look at each individual main effect, as well as the interaction of the main effects with Time. This model is reduced based on AIC and the p-value associated with each effect. Following this model, we assess the individual characteristics of the Client level-2 model in the same way; a full model with contextual variables to the Clients, reduced to a model with significant terms and lowest AIC.

The level-2 model, models the random intercept from the level-1 1 model as it is the only parameter permitted to vary. The intercept will be modeled by Client characteristics, also determined by the exploratory data analysis and from SMEs opinions.

Level-2 Client sub-model:

$$\boldsymbol{\gamma_{0is}} = \boldsymbol{\delta_{00s}} + \pi_{01} Age_{is} + \pi_{02} Gender_{is} + \pi_{03} Education_{is} + \pi_{04} HispanicLatino_{is} + \pi_{05} Diagnosis_{is} + \pi_{06} Time_{tis} + r_{0is} \quad (2.7)$$

where $\boldsymbol{\delta_{00s}}$ is the average intercept across level-2 units;

π_{01} is the effect of Age;

$\pi_{02}$is the fixed effect of Gender;

π_{03} is the fixed effect of Education;

π_{04} is the fixed effect of Hispanic/Latino;

π_{05} is the fixed effect of Diagnosis;

π_{06} is the fixed effect of Time;

and r_{0is} is the random error.

We only now model the random intercept from the model that models the random intercept for the whole hierarchy. Details of why are again discussed in Chapter 3.

Level-3 Site sub-model:

$$\boldsymbol{\delta_{00s}} = \gamma_{000} + \beta_{00t}Time_{ts} + \mu_{00s} \quad \textit{(2.8)}$$

where $\boldsymbol{\gamma_{000}}$ is the average intercept across level-3 units, β_{001} is the effect of time in site s, and $\mu_{00s}$is the unique increment of the intercept associated with level-3 unit s.

Note that the random effects μ_{00s} are constrained to sum to zero.

We can now put the above three-level multilevel model into an HLM:

Full Model: Level-1 + Level 2 + Level 3

$$CISPSY_{tis} = \gamma_{000} + \beta_{00t}Time_{tis} + \pi_{01}Age_{is} + \pi_{02}Gender_{is} + \pi_{03}Education_{is} + \pi_{04}HispanicLatino_{is} + \pi_{05}Diagnosis_{is} + \beta_1 HealthStatus_{tis} + \beta_2 Attend_{tis} + \beta_3 IFF_{tis} + \beta_4 Income_{tis} + \beta_5 PTSD_{tis} + \beta_{6t}HealthStatus_{tis} * Time_{tis} + \beta_{7t}Attend_{tis} * Time_{tis} + \beta_{8t}IFF_{tis} * Time_{tis} + \beta_{9t}Income_{tis} * Time_{tis} + \beta_{At}PTSD_{tis} * Time_{tis} + (\mu_{00s} + r_{0is} + \varepsilon_{tis}) \quad \textit{(2.9)}$$

which can be interpreted as the outcome CIS Psychological Score is function of the average regression equation plus random error having three components; r_{0is} the random effect of unit i on the mean, μ_{00s} the random effect of unit s, and the level-1 error ε_{tis}. HLMs look at within- and between- Client and Site variation, where observations of different levels or clusters are independent of one another, but observations within clusters or levels are correlated. Higher levels of hierarchical models are also often used. This is, we present a growth model that is a three-level hierarchical model.

2.2.3 Interclass Correlation Coefficient

We begin by reinstating that the unconditional model in the analysis allows the researchers to understand how much and where there is within- and between-subject variation through the interclass correlation coefficient. In a growth model, we can say a certain amount of variance in growth exists between subjects. ICC can be useful in many statistical situations, but especially in hierarchical mixed models (Musca et al., 2011). ICC is a statistic that measures the degree of dependence among observations nested within levels (Bickel 2007). We can get the proportion of covariates variability that occurs between the clients, rather than within the clients, and between the sites, rather than within the sites and if observations within groups are dependent (Bickel 2007). We also get the proportion of the total variance in the response that is accounted for by the clustering. It can also be interpreted as the correlation among observations within the same cluster. Working with more than two levels show nesting patterns that we would otherwise not anticipate. We see that increasing distance between subjects and grouping variables does not always ensure that groups become less homogenous, as seen through a low non-significant ICC (Bickel 2007). If the ICC's variance component at the second level is large, this implies that traditional OLS could be used (Raundenbush and Bryk, 2002).

If there is no real correlation among observations within a cluster, the cluster means will not differ. It is when some clusters generally have high values, and others have relatively low values that the values within a cluster are correlated. That variance parameter estimate is the between-cluster variance. The variance of the residuals is the within-cluster variance. Their sum is the total variance in the response that is not explained by a covariate or set of covariates. When some clusters generally have high values, and others generally have low values (in other words, where there is consistency among a cluster's responses), there is variation among the clusters' means. This is the between-cluster variance. The within-cluster variance represents how far each point is to the cluster-specific mean. The ratio of the between-cluster variance to the total variance is this ICC.

When there are repeated measures, this ICC measures the dependence through a ratio of the level-2 unexplained variance (between individuals) to the level-2 plus level-1 unexplained variance (between occasions). Thus, having a high ICC reflects that individuals are similar over time. The estimated standard errors, the fixed terms (the slopes) in the hierarchical model, are used and corrected for this dependence. The "leftover" dependence is modeled through random effects. So, the random coefficients for these intercepts are permitted to vary across the cross-level interaction terms.

Another useful reason to calculate ICC is that it can help determine whether a hierarchical mixed model is necessary. If the ICC is zero, then the observations within clusters are just as similar to observations from different clusters. ICC can be theoretically meaningful to understand how much of the overall variation in the response is explained by clustering. ICC allows us to see how the between- and within-cluster variances change as variables are added to the contextual model.

Small interclass correlation can substantially deflate standard errors of the regression coefficients which tells us about the variance associated with the estimate (Bickel 2007). A smaller standard error would indicate more precise estimate of the coefficient and means we could inflate the t statistic, meaning we are more likely to find significance if there is significance. LeBreton and Senter (2008) have suggested that an ICC=.05 represents a small to medium effect. Bliese (1998) simulated conditions where the within-group ICC was 1%, where strong group-level relationships were detected that were not evident in the lower-level data. For assessing the reliability of group-level means, ICC <0.40 are poor, 0.40 to 0.75 are fair to good, and those >0.75 are excellent (Fleiss, 1986).

2.2.4 SAS Modeling

We used the MIXED procedure in SAS in conjunction with REML to fit the data/model. The REML criterion itself is optimized by a Newton method where the derivatives from the method are obtained by a mixture of implicit differentiation and direct methods (Wood 2011). The MIXED procedure profiles the likelihood with-respect-to the fixed effects and also with respect to the residual variance.

The covariance parameter estimates are the estimates of the parameters in **G** and **R** matrices. The **G** matrix specifies the subject-specific effects, and the **R** matrix specifies residual effects. Wald Z tests of the covariance parameters are computed, and they are only valid asymptotically. The asymptotic standard errors are computed from the inverse of the second derivative matrix of the likelihood with respect to each of the covariance parameters. When there are no missing values, the F tests from the ANOVA will be identical to those from a REML analysis.

The primary assumptions of the analyses performed by PROC MIXED are as follows: the data are normally distributed, the data are MAR, the means of the data are linear in terms of a particular set of parameters, the variances and covariance's of the data is in terms of a different set of parameters, and they exhibit a covariance structure the procedure can estimate. Since normally distributed data can be modeled in terms of their means and variances/covariance, the two sets of parameters in a mixed linear model specify the complete probability distribution.

Typically, in a case of repeated measures, it is appropriate to use autoregressive (AR), compound symmetry (CS), or unstructured (UN) variance structures. We use the variance covariance (VC) structure and assume that responses over time were independent conditional on the covariates.

Not only is VC the default in SAS, but it also models a different variance component for each random effect or repeated effect. In the analysis presented in this paper, we chose to model the **X** (design matrix for the fixed effects), **Z** (design matrix for the random effects), **G (**fixed**)**, and **R (**random) covariance matrices through variance components. Here we see correlated errors between time points within subjects. These correlations are presumed to be the same over time, regardless of how distant in time, repeated measures are made. So if all variables were completely independent of each other and measured on different scales, the VC structure is a reasonable pattern. When the data are repeated measures, VC is not always the best choice, but our analysis has very few repeated measurements. VC is used as it is one of the most common structures that arise from the use of random-effects parameters, which are additional unknown random variables assumed to affect the variability of the data. The variances of the random-effects parameters become the covariance parameters for this particular structure.

2.2.5 Full Maximum Likelihood and Restricted Maximum Likelihood

Estimation is often achieved by finding the value of the parameter that maximizes the likelihood. Generally, to start evaluating a likelihood function, we partition the function into two terms; the mean and variance. In full maximum likelihood estimation, we find a function so that the mean does not depend on variance, and the residual likelihood is that which only involves the variance parameter. When using REML, we have one unknown parameter and represent the log-likelihood that is free of the mean parameter.

When there are closed-form solutions to the variance components in linear regression, we can remove the bias by multiplying a correction factor, after the estimation process. When there are no closed-form solutions, one needs to obtain a bias-free estimation for variances through a more general process. This general process is restricted maximum likelihood, which is to maximize a modified likelihood function with no mean components (Patterson and Thompson, 1971; Harville, 1977) and is an increasingly common method. If one were to maximize the likelihood function with mean and variance components where we use the true mean to estimate the variance, it is unbiased. So the REML method is used to make the maximization process as unbiased as possible.

If a residual variance is a part of the mixed model, it can usually be profiled out of the likelihood. This means solving analytically for the optimal and plugging this expression back into the likelihood formula. This reduces the number of optimization parameters by one and can improve convergence properties. REML restricts negative estimates of the variance components.

In HLM's, finding the ML (REML) estimates of slope and intercept parameters requires one to integrate likelihoods over all possible values of random effects (Bolker et al., 2009).). Standard ML estimates the SDs of the random effects assuming the fixed effects are correctly specified. REML is a variant that estimates averages over some uncertainty in the fixed-effect parameters. It uses n-p (n is the sample size, and p is the number of fixed effects: the number of betas) to remove the fixed effects in the process. REML is best when trying to model random effects as ML underestimates the random effects SD (Bolker et al., 2009). REML is an estimator that estimates the random-effect parameters, such as standard deviation (SD), averaged over the values of the fixed effect parameters (Lee and Neldar, 1996). Estimated SDs for the same model using REML, instead of ML, are generally less biased (Lee and Neldar, 1996). ML avoids list-wise deletion of participants with incomplete data over time to minimize bias and maximize power (Collins et al., 2012).

REML takes account of the number of (fixed effects) parameters estimated, losing 1 degree of freedom for each. This is achieved by applying ML to the least-squares residuals, which are independent of the fixed effects. So, to compare models with different fixed effects, ML must be used. ML method underestimates the variance parameters because it assumes that the fixed parameters are known without uncertainty when estimating the variance parameters. Thus, use REML when you are interested in estimating the covariance effects. REML uses a mathematical trick to make the estimates for the variance parameters independent of the fixed effects' estimates. REML works by first getting regression residuals for the observations modeled by the fixed-effects portion of the model, ignoring any variance components. ML estimates are unbiased for the fixed effects but biased for the random effects, whereas the REML estimates are biased for the fixed effects and unbiased for the random effects (Duchateau et al., 1998).

However, the ML estimator usually has a lower mean-squared error (MSE) than the REML estimator. If one wants to be right on average, go with REML, but pay for this with greater variability in the estimates. If one wants to be closer to the actual value on average, go with ML, but pay for this with negative bias. In the simple case of a constant mean and constant variance, ML is dividing SSR with n while REML is dividing SSR by (n-1), so

REML is a generalization of this procedure. Biases get smaller for larger sample sizes under ML. For small sample sizes, REML is preferred. However, likelihood ratio tests for REML require the same fixed effects specification in both models (Lee and Neldar, 1996).

For mixed models, the likelihood function is the joint density function integrated over the random effects. These estimates are biased because they do not account for uncertainty with-respect to estimates of the means of fixed effects. REML estimates must lie within the possible parameter space. The poor behavior of REML estimators seems to be that these estimators do not have positive probability, so there is an instantaneous change in RMSE behavior between datasets for which REML does not exist and those for which they exist. The ML estimator always exists while the REML estimator may not exist with positive probability (De Oliveira and Ferreira, 2011; O'Neill 2013).

Modeling under REML can produce unbiased estimates of covariance parameters while the ML estimator is negatively biased. However, the ML estimator more often has a lower mean-squared error (MSE) than the REML estimator. If one wants to have more correct estimates use the REML estimator, but the trade-off is more considerable variability in the estimates. If one wants to be closer to the actual value, and are comfortable with a more negative bias, use the ML estimator. ML methods underestimate the variance parameters because they assume that the fixed parameters are known without uncertainty when estimating the variance parameters. Overall, ML estimates are biased for the variance components, while REML estimates are biased for the fixed components. An essential strength of using REML is that it gives higher weight to larger groups among factor variables (Bickel 2007).

2.2.6 Fixed and Random Effects

Deciding which variables are modeled as fixed or random strongly depends on the hypotheses one is testing or questions aiming to be answered. Some common questions are: on average, how does individual growth change from intake to discharge, or deciding which covariates are related to such growth or change. The mean parameters in the model are referred to as fixed-effects parameters and are associated with known covariates and can be qualitative (as in the traditional analysis of variance) or quantitative (as in standard linear regression). The covariance parameters in the model are the random effects which are these larger "subjects" one is interested in One can also think of random effects as those factors whose levels are sampled from a larger population or whose interest lies in the variation among them rather than the specific effects of each level (Bolker et al., 2009). Effects are fixed if they are interesting in themselves and random if there is interest in the underlying population (Searle, Casella, and McCulloch 1992). Mixed effect models are those that incorporate both fixed and random effects and are hierarchical in that they posit distributions for latent, unobserved parameters (Bell, Fairbrother, and Jones, 2019).

Fixed effects models control for, or partial out, the effects of time-invariant variables with time-invariant effects and this is true whether the variable is explicitly measured or not (Beck 2011). Fixed effects are more commonly used in models than random effects, as one wants to be able to talk about how each covariate is in itself is affecting the model. An easy way to view a fixed effect is that they are constant across individuals, while random effects vary. When thinking of the effects in a hierarchical model, fixed effects can be thought of as those factors whose interest lies in the specific effects of each level- effects of covariates or differences among treatments. The fixed effect assumption is that the individual-specific effects are correlated with the covariates variables. Given the above descriptions and ideas, we treat the time variable as fixed, as well as all the covariates in the models. Another reason to treat covariates as fixed effects is because they are part of experiments that cannot be "added" or "removed" as they are inherent to the study. For example, we treat age as linear, so they are fixed. Time-varying covariates are variables whose values can change across time. Although the covariates' value changes across time, the parameter value estimating each covariate's effect on the dependent variable is assumed to be constant across time.

The covariance parameters are what distinguishes the mixed linear model from the standard linear model. Random effect models assist in controlling for unobserved heterogeneity when the heterogeneity is constant over time and not correlated with covariates variables (Gormley and Matsa, 2014). The traditional use of random effects is a way to correct statistical tests when some observations are correlated. Random effects can be viewed as the source of random variation, such as experimental units and can be described as covariates variables in which the interest is in making inferences about the distribution of values (i.e., the variance among the values of the response at different levels) rather than in testing the differences of values between the particular levels.

We can also think of random effects as a way to combine information from different levels within a grouping variable. Random effects are especially useful when we have lots of levels, relatively little data on each level, and uneven sampling across levels (Fox et al., 2015). Conversely, random effects are generally ineffective when the grouping variable has too few levels. The simplest way to think of deeming what variable should be treated as random is that of a grouping variable or identification variable-such as to indicate the person, place or thing, to distinguish itself. This is why Clients and Sites are treated as random in the models. We want to see which and how all the covariates over time among the individuals change. All the clients are homeless; we want to make their lives meaningful somewhere (site) over time. In this modeling schema, we *can* talk about individual random effects. Inherently people vary from one another so they can be treated as RE.

In our model, we can partition some of this error and variation into the Site and Client levels since time is not equally spaced for all the clients - otherwise, we will see unexplained variation. Modeling Clients and Sites as RE is crucial since we aim to make inferences about the general homeless population and not just that individual from their sites. This Site variable only has three, and time only has three groups. Random effects usually can't be used when the grouping variable has fewer than five levels, and random effects variance estimates are unstable with fewer than eight levels (Fox et al., 2015). Again, we use RE to remove variation for better prediction, which is conditional on the covariance, so the remaining variability is explained in the fixed effects. Random intercepts allow the outcome to be higher or lower for each site; random slopes allow fixed effects to vary for each site. We usually talk about them in terms of their variability, instead of focusing on them individually. In other words, we can now incorporate (instead of ignoring) site-to-site variability in clients' growth and improve our ability to describe how fixed effects relate to outcomes. We can also talk directly about the variability of random effects, similar to how we talk about residual variance in linear models. No general measure of weather variability is large or small exists, but subject-matter experts can consider standard deviations of random effects relative to the outcomes.

In a growth study, one would see a model with a random intercept and fixed slope, which correspond to parallel lines for different individuals. Mixed effect models include fixed effects (in these cases, estimating the population level coefficients) and random effects (to account for individual differences in response to an effect). General categories of mixed-effects models are repeated measures, longitudinal, hierarchical, or split-plot design (Gueorguieva and Krystal, 2004). Mixed-effect models, whether linear or generalized, are different in that there is more than one source of random variability. Mixed models can adjust for missingness or unknown covariates/predictions that are common to a group of observations while not fundamentally changing the statistical inferences.

SAS will produce standard errors and p-values, that are adjusted so that they account for all of the fixed effects in the model as well as the random variability among individuals. We call the variability across individuals' the residual variance (in linear models, this is the estimate of σ^2, the mean squared error (MSE)). This MSE is the variability that was unexplained by the covariates in the model (the fixed effects).

Mixed models allow us to make a "broad level" inference about the larger population of homeless people, which do not depend on a particular site. In repeated measures, where more than one measurement is taken on the same individual, the "group" effect is thought of as random because we only sample a subset of the entire population of subjects (Fox et al., 2015). Here, we present a flexible model for repeated measures longitudinal growth data within individuals that allows trends over time to incorporate individual-specific random effects. These may reflect the timing of growth events and characterize within-individual variability, which can be modeled as a function of covariates.

The fixed portion of the mixed model is doing what a linear model does; it fits an overall regression line over time. The coefficient we get for the fixed effects measures the difference in their intercepts, and the coefficient for the time measures the difference in their slopes. We measure how much each Site and Client intercept differs from the population average intercept, then find the variance of these different measures; the variance estimate for the random intercept. If that variance comes out to 0, it indicates that the intercept of time on Clients is the same for all Sites—they do not vary from each other. Time in the fixed statement measures the overall effect of time on clients across all sites. In the mixed model, we would see a different line for each person over time, where we have an overall line that people deviate from.

2.3 Data Attrition Procedure and Results

The following sections discuss the missing data problems that are in our dataset. We assume that our data is missing at random, which is an assumption of PROC Mixed. We use mean based imputation to correct for any missing data in our subset.

2.3.1 Missing Data

Determining the cause of attrition (response loss through loss of participants) and missingness (response loss through loss of information not being recorded) is essential in making a complete data subset. We need to assess the response process, the extent to which whether respondents understand survey questions to mean what we intend. Three common types of missing data are missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). Some reasons for missingness might be respondent attrition, survey structure, file-matching issues in the merge process, refusal to answer sensitive questions, and more. Deciding which missing type the data are can be a difficult task. When the missing observations are random, all the time points or "misses" are random samples of all participants. When the response process is independent of all variables, then the data are MCAR- that is, if the probability of missingness is the same for all units

(Raudenbush and Bryk, 2002; Maas and Snijders, 2003). When data are MCAR, there is no relationship between the covariates and whether or not the response is measured or not measured after controlling for some main effect-or control. To determine or assume if data are MCAR is virtually impossible (Raudenbush and Bryk, 2002).

When data are assumed to be MAR, then the probability of the missing time point is independent of the missing data given the observed data- that is, the probability a variable is missing depends on the available information (Raudenbush and Bryk, 2002). This means the response process depends on observed but not on unobserved variables. MAR is more comfortable to assume as the observed data capture confounding influences, such as variables related to both attrition and outcome (Raudenbush and Bryk, 2002). If data are MAR, the estimation of the treatment effect will be unbiased if all the data are used in the analysis. When all of this is true, the use of model-based imputation will ensure ignorable missingness (Raudenbush and Bryk, 2002). Non-ignorable missingness arises when the data are not MAR or MCAR. That is, the probability of attrition does not depend on the missing value (Raudenbush and Bryk, 2002). Our study's primary goal is to identify which covariates affect the CIS scores over time. We only have data on CIS scores for over half (53%) of the time we studied. We do not want to throw out half of our data because it does not talk about our research question, but rather impute the data.

An example of MCAR is incompleteness due to randomly failing physical system, and examples of MAR are recording failures depending on group or measurement occasion, but not otherwise on the (unrecorded) value; and termination of the observations after recording a value above or below a given threshold (Little and Rubin, 1989; Schafer, 1997). When data is MAR or MCAR valid likelihood-based statistical inference is possible without modeling the response process (Little and Rubin, 1987; Schafer, 1997). If the probability of response depends on the unobserved dependent variable (and not only as a function of the covariates), the missingness itself is informative, and it is preferable to specify a model that employs this information (Christensen et al., 2011). Under MAR, valid inferences can be made based on ML estimates of the complete data distribution given the incomplete data (Raudenbush and Bryk, 2002).

2.3.2 Imputation

Imputation is done by randomly picking a value from a distribution of the variables used. SAS PROC MIXED handles missing level combinations of classification variables by deleting the fixed-effects parameters corresponding to missing levels to preserve estimability. The procedure does not delete missing level combinations for random-effects parameters because linear combinations of the random-effects parameters are always estimable (SAS/STAT® 14.1 User's Guide: The MI Procedure). Multiple imputation (MI) is a robust and flexible option for handling missing data, so one can assure that the use of multiple imputation is reliable. SAS deals with missing data problems through appropriate multiple imputation and analytic methods that are most compatible with the analysis.

Two crucial advantages of multiple imputations are that it incorporates the variability introduced by the imputation during variance estimation, and it offers the use of appropriate statistical models for generating plausible distributions of values to replace item-missing data, not unit missing data (SAS/STAT® 14.1 User's Guide: The MI Procedure). The number of imputations depends on how much data is missing, how many records and variables included in imputation models, and other factors, where an iterative process is used to evaluate the number of imputations. There are many ways to impute data and we use mean imputation using information from related observations for our data.

To analyze a full dataset using imputation could have been achieved through SAS PROC MI. We did not use this process to deal with the attrition of losing participants over time, as we wanted to see what was happening among those who completed the surveys. This way produces biased estimates, but any imputation causes biasedness. It is important to note that we would produce more biased and unreliable estimates/results if we imputed all the missing time points for all clients and covariates where the is data missing due to attrition.

2.3.3 Application to Our Problem

Missing data problems are part of many data analyses and especially prevalent in longitudinal data sets. For the analysis presented in this paper, we assume that the data is MAR. We can assume that Gender, Ethnicity, Education, Diagnosis, Income, and Age are recorded for all the people in the survey and do not change in the relative time period (with slight but negligible change in Income and Age). We can say the data is missing at random if the probability of nonresponse to this question depends only on these other, fully recorded variables. When an outcome variable is missing at random, it is acceptable to exclude the missing cases (that is, to treat them as NA's), as long as the regression controls for all the variables that affect the probability of missingness. Meaning our model needs to have included covariates for ethnicity (and others), to avoid nonresponse bias. Both require that sufficient information be collected to "ignore" the assignment mechanism (assignment to treatment, assignment to nonresponse).

The numeric covariates Age, Income, Health Status, and PTSD score are all imputed by their grand mean, respectfully. Another necessary caution is that Age and time are linked, and if Age is used as a covariate in growth models, it should be treated as time-invariant, e.g., age at a fixed point. Rather than replacing each missing value in a dataset with one randomly imputed value, it may make sense to replace each with several imputed values that reflect our uncertainty about our imputation model. We may want our imputations to reflect not only sampling variability but also our uncertainty about the model's estimates.

Our study has only self-reported data; all the demographic variables and covariates are self-reported, except for Diagnosis. Thus, data that we are interested in using are all self-reports, and when there are many surveys to many people over time, some will not complete every survey. For covariates, we miss about 2% of data as most surveys that were taken, were completed. We miss data for all time points for each client that did not show up, where they completed 1, 2 or 3 surveys. We assume that there are no systematic differences between those for whom these data are missing and those for whom these data are present, making these data missing at random. The missingness is the same across all clients. The missingness in our data is not by survey, as a clinician administered all surveys to each client that could be located. NAs exist in the surveys as an option, but each missing point is because a client did not show up or could not be located.

The missingness in the covariate is random, as is the showing up of a client is random. We have on average (±5 max per variable) 370 individual client data at intake, 286 at 6-months, and 143 at discharge. This is the number of clients to report on each survey, not that they skipped a question as they were given the survey and asked every question through a face-to-face interview done by a clinician assistant.

As always, we should worry about the MCAR/MAR/MNAR assumptions. The surveys we are analyzing are not structurally different from surveys we do not have data on, conditional on the other information we have, assuming that missingness in response data is related to individual subjects. It is unlikely that missingness is related to individual attendance, conditional on the subject. Each survey is recorded when a clinician assistant can go to the shelter, and the client is around to get the survey within an appropriate time frame. Since we are making sure to get three survey samples from every subject based on the time with the complete response data, we should be able to make the MAR assumption.

People differ in their commitment to show up for the survey and keeping their housing/treatment. It is probably unreasonable to assume that there are no differences between clients for whom we have more consistent data and clients for whom we lack consistent data. When assuming MAR, we assume that conditional on what we know about these clients through our model. There are no systematic differences between treatment where we have consistent data and treatment where we do not. Missing data is a common complication in data analysis, and dealing with it can be tricky.

Due to attrition, our data is missing about 22% of participants at 6-months and 62% at discharge- meaning 78% of participants at 6-months and 38% of participants at discharge completed surveys. For completed surveys, each variable has very few missing data (about 2.5% max). We are missing data in the sense that we are missing the client survey at some point-poor attrition rates are the driving factors in this missing data problem. There are 54 individuals who have some missing data values. The fully observed data on all variables is 93.67% of our sample, or 799 individuals are assigned to the complete data group while the 54 individuals make up 6.33% missing data. We have a total of 74 variables in the reduced data set that will be analyzed. Out of the 74 total variables in the data set, 43 (constituting

58.1% of the full data set) are continuous variables that makeup 7 of the continuous variables needing imputation for the analysis.

The fully clean and imputed dataset (14 total variables in the model, seven being categorical) are a mix of continuous and categorical factor variables, and the missing data pattern is still considered monotone. There are 853 total observations. 799 used observations and coincidentally 54 subjects who missed at least one survey at intake, 6-months, or discharge. The maximum number of observations for each time (intake) is 370, where each client, on average, completed 2.22 surveys. Overall, the number of cases lost was about 31.5% due to attrition and missing data. That is 269 people did not complete a survey at 6-months or discharge. The valid number of cases with likewise deletion of missing data is 35 of 269 (13%) missing cases, meaning 35 clients skipped a survey but appeared again later. Two hundred thirty-four clients either missed the 6-month follow up or discharge follow-up survey (there are 5 cases where the client missed the intake survey but appeared for the others). Of the 269 missing cases, 89 clients only completed the intake interview. We now understand that the majority of missing data is due to the retention or lack thereof, of clients and not missing data in the variables themselves. The analysis used a complete data set of 799 observations and 15 variables (including the response), with grand mean imputation for those with missing values.

2.4 Why Hierarchical Mixed Modeling

One of the major features is that the data is nested of time in people the data for time is nested within that for people. We found that this clustering is essential in explaining the variance in the data and time between and within-clients. The hierarchical model also enables decomposition of the variation in these individual growth trajectories into within- and between-level components at each level. We use the entire model and all the information to provide each grouping variable with separate predictive equations for specific covariates of interest. Without accounting for the nested nature of the data and grouping variables, it is possible to lose important information about the data and the effect of variables on outcomes.

Level 1 sample size is the number of repeated measures occurring within each subject/client (Bickle 2007). Using a hierarchy allows us to investigate higher-order interactions, such as cross-level interaction terms. In addition, the first level(s) of the hierarchical model represents the association between the observed data and the “latent” true data (Raudenbush and Bryk, 2002). Using such hierarchical models, we can use contextual variables to find explanatory variables we would not have found otherwise - for example: how time or age differs and at which points (Bickle 2007). This allows analysts to define contextual effects (grouping variable effects) in a more informative way and to answer more specific questions relating to each level of a model (Bickle 2007).

We use mixed modeling so that the use of the random effects permits coefficients to vary from context to context and level to level and dependent residuals are accounted for through ICC (Bickle 2007). Cross-level interactions, which can be investigated using mixed modeling, are a specific type of contextual effect that, when modeled as an upper-level variable, moderates the association between lower-level covariates and the outcome variable in the overall model (Bickel 2007). These contextual effects also help explain variation in the previous level's dependent variable and in random components corresponding to an intercept (and slope) that vary from group to group (Bickle 2007). The random effects in the hierarchical model make it possible to draw estimates of variance and covariance components with unbalanced nested data (Raudenbush and Bryk, 2002). We use these hierarchical models to see if adding contextual variables helps in explaining variance in the model; that is as we add variables to the model, is the variance diminishing?

The question then arises as to why OLS regression cannot be used in such a situation and to what exactly these levels are contributing. In short, what are the benefits of using a hierarchical model instead of standard OLS techniques? The reasons for such methods are many and, in the following passages, we outline those which we found most important.

For our analysis, we have repeated measures data, which means dependent and similar observations as individuals were measured more than once. HMM allows for dependent observations, non-independent errors, and can be extended to deal with more complex behavior. We see that HMMs add structure to the error terms, which OLS does not. Under similar conditions, OLS and HMM will produce the same coefficient estimates because violating the assumption of independence does not bias coefficient estimates. They do, however, tend to underestimate standard errors, which is no surprise, given that multilevel modeling includes least squares regression as a special case. HMMs recognize the existence of such data hierarchies by allowing for residual components at each level in the hierarchy. In addition, modeling clustered/nested data through an HMM gives them better predictive abilities and reveals significant main effects that could not have been found in OLS (Gelman 2006).

The feature that distinguishes HMM from an ordinary regression model is the presence of two random variables; the measurement level variable and the subject level variable. Instead of one general random-effect that captures how each observation deviates from the predicted fixed-effects, there are multiple random-effects that illustrate how observations deviate within a cluster, and how each cluster deviates from the overall group. Another feature of hierarchical models is their ability to separately estimate the predictive effects of an individual covariate and its group-level mean. In a multilevel model, random variables are used to model the variation between groups, rather than using dummy variables, which would be necessary in ordinary regression models.

We use the estimated variability for each random-effect to control for variance when estimating the significance of our fixed-effects. In this way, we can model our data at the observation level (micro-level) and at the cluster level (macro-level). We then add a random-effect for our “subject”, the clients and sites, which characterizes each participant’s idiosyncratic variation from the fixed-effect estimates. The same slope can be assumed for all clients, given that each participant's individual differences would only have an effect on the intercept, not the slope. Because a different intercept for each participant within the "subject" variable must be estimated, it is most effective to use a random-intercepts model. While this accounts for the individual by-subject variation, the/our general error term captures the random deviations within the same subject, which multiple responses per subject and the dependence of each response of the subject's baseline. In order to resolve these random deviations, HMMs controls for this within-cluster variance and tests the fixed-effect estimates against the remaining between-cluster variance.

Lastly, HMMs allow for grouping of level-1 outcomes within each successive level. Cross-classified data comes from, clients in the same site who may come from different areas and clients from the same area but who may go to different sites. By assuming that the random effects come from a common distribution, a multilevel model can then share information between groups, which improves the predictions for groups that have relatively little data. Inclusion of covariates at the group level illustrates the difference in a covariate. In

HMM fewer parameters are needed in order to allow the intercepts to vary from subject to subject, which contrasts with the need for additional parameters in the dummy variable approach. This reduction in the number of parameters is particularly important with more complex models and a limited amount of data. We can allow for possible correlation by including the average of x within the group-level regression. Compared with classical regression, hierarchical modeling is almost always an improvement, proving helpful for data reduction and causal interference, and essential in prediction multilevel modeling.

Chapter 3: Analysis

3.1 Descriptive Statistics

We first present tables of the demographic statistics for the variables used in the analysis. For categorical variables, we desire that each factor is well balanced. We look at how many clients are in each homeless shelter and the attrition rate throughout the study. Then we look at each covariate and outcome, and their attrition rates.

The homeless shelter with the most retention and most clients that participated in the study is Life Link in Santa Fe, while St. Martin's, in Albuquerque, NM, had the second most participation with the most attrition, as seen in Table 3. 1. The overall trends we see in the general demographics from Table 3. 2 are at intake are that about half of the sample is Hispanic-Latino, the majority are men, the average age is about 45 years old, and the average monthly income is under $1,200. While 40% of participants have more than a high school education, those who have less than a high school education or have a high school education are each about 30%. Most of the clients are diagnosed, by a clinician, with depression; this being the more common diagnosis seen among those experiencing homelessness. The next most common diagnosis seen among the clients is post-traumatic stress disorder (PTSD); this is also a common disorder among homeless communities across the United States (Crisanti et al., 2021). The range for the PTSD score is from 17-85 points with an average of 49 points and most people report having a PTSD score of 49 to 70 points, meaning most people, overall, have high PTSD scores. The range for the outcome variable CIS Psychological is 4 to 20 with an average of 12 points.

The general trends that we see in the remaining covariates are presented in Table 3. 3. During the duration of the study, about half of the clients did not attend church, self-help groups, narcotic anonymous or alcoholic anonymous and recovery related activities. More than 1/3 of the sample reported that they saw friends or family during the study. The majority of people reported good or fair health. Make note that the variables found in the final hierarchical model and sub-models are *Time, IFF, Health Status, PTSD, Gender*, and *Education*. With attrition, the proportions of the covariates are staying about the same throughout the duration of the study. The models with these contextual variables are the models we make inferences about to propose policy recommendations. We keep only those variables that are significant as there is no sufficient evidence to conclude that there is a difference in CIS Psychological score by those variables with p-values greater than the testing point of an $\alpha = 0.05$.

Table 3. 1: Number of Clients at each Site (homeless shelter, in New Mexico cities)

Time	**Location**
Intake	N=370 St. Martins(ABQ)=126 (34%) Life Link(SF)=152 (41%) St. Luke's(LC)=91 (25%)
6-months	N=286 St. Martins(ABQ)=101(35%) Life Link(SF)=121 (42%) St. Luke's(LC)=64 (23%)

Discharge	N=143 St. Martins(ABQ)=29 (20.5%) Life Link(SF)=85 (59%) St. Luke's(LC)=29 (20.5%)

Table 3. 2: General demographics

Time	Hispanic Latino	Gender	Average Age	Average Income
Intake	N=370 Yes=173 (47%) No=197 (53%)	N=370 F=151 (41%) M=218 (59%)	N=370 44.5 years old	N=370 $816.15
6-months	N=286 Yes=135 (47%) No=151 (53%)	N=286 F=125 (44%) M=161 (56%)	N=286 44.5 years old	N=286 $974.47
Discharge	N=143 Yes=74 (53%) No=69 (47%)	N=143 F=65 (45%) M=78 (55%)	N=143 44.9 years old	N=143 $1,123.62

Table 3. 3: Covariate Demographics

Time	Education	Diagnosis	Health Status	Attend	Interact Family Friends
Intake	N=370 <High School=110 (29%) High School=113 (31%) >High School=147 (40%)	N=370 Bipolar=61 (17%) PTSD=86 (23%) Depression=122 (33%) Schizoaffective=34 (9%) Anxiety/Other=67 (18%)	N=370 Excellent= 10 (2%) Very Good=38 (10%) Good=106 (29%) Fair=143 (39%) Poor=73 (20%)	N=370 0=182 (49%) 1=114 (31%) 2=62 (17%) 3=12 (3%)	N=370 Yes=288 (78%) No=82 (22%)
6-months	N=286 <High School=79 (27.6%) High School=80 (28%) >High School=127 (44.4%)	N=286 Bipolar=50 (18%) PTSD=67 (23%) Depression=100 (35%)	N=286 Excellent= 12 (4%) Very Good=32 (11%) Good=77 (27%)	N=286 0=133 (46%) 1=99 (35%) 2=43 (15%)	N=286 Yes=209 (73%) No=77 (27%)

		Schizoaffective=32 (11%) Anxiety/Other=37 (13%)	Fair=110 (39%) Poor=55 (19%)	3=11 (4%)	
Discharge	N=143 <High School=42 (29.6%) High School=35 (24.4%) >High School=66 (46%)	N=143 Bipolar=25 (17%) PTSD=39 (27%) Depression=48 (34%) Schizoaffective=17 (12%) Anxiety/Other=14 (10%)	N=143 Excellent= 6 (4%) Very Good=21 (15%) Good=38 (27%) Fair=52 (36%) Poor=26 (18%)	N=143 0=67 (47%) 1=51 (36%) 2=20 (14%) 3=5 (3%)	N=143 Yes=102 (71%) No=41 (29%)

In Figure 3.1 we have treated the Site variable as a fixed effect and did a 2-way ANOVA of Site and Time to assess if these differences are statistically significant or not. We see that overall, the 3 Sites had about the same CIS Psychological score over time, averaging about 12.5 points. The range at intake, on average, is between 9.5-12.5 points for CIS Psychological score. We see that St. Martin has the lowest average score at intake, followed by Life Link, and St. Luke's has the overall highest average. The steepest slope is seen at St. Martin's, meaning that over time, the average CIS Psychological score increased most at this site. Both Life Link and St. Luke's have similar slopes and the general trend is that the CIS Psychological score increased over time at both Sites. St. Martins was statistically different from both Life Link and St. Luke's but St. Luke's and Life Link are not, in these difference in scores over time. Note that Site is not intended as a covariate, but rather the grouping variable for the third level of the model associated with the sampling design. We will see that the explained variance in the CIS Psychological score is not attributed to grouping at the Site level. This graph depicts the results we will see.

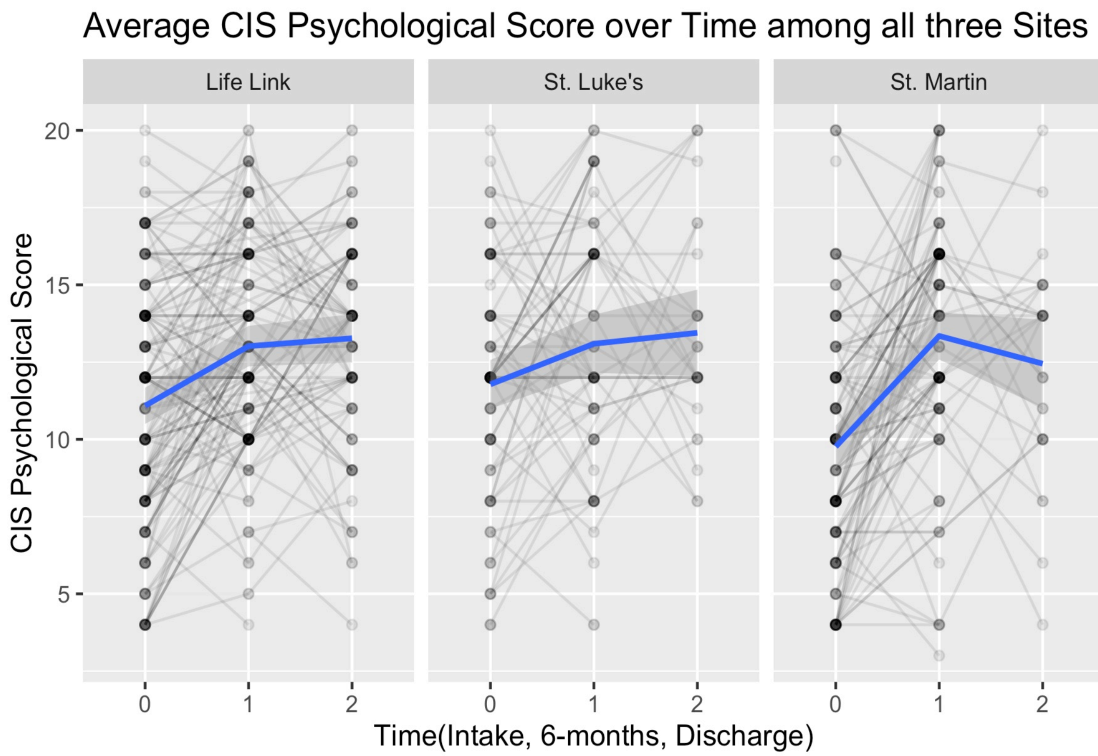

Figure 3.1 CIS Psychological score over time among all 3 sites averaged over clients

3.2 Analysis

The model in this thesis is an individual growth model and an organizational model in that we have time nested in people and people nested in a site, i.e., growth over time of an individual within an organization. We also do not have a numerically continuous time variable, dictating that there is no true linear or higher-order growth curve, though we continue to use growth modeling theory to model the data. We assess the change over time by looking at what is happening between each interview time for all the clients in each site. Ignoring such a structure can impact estimated variances, inflate Type-I error rates, and the statistical power to detect the covariate effects. When not incorporating a nested structure, obvious interpretation errors will occur in significance testing, as well (SAS/STAT® 14.1 User's Guide: The MIXED Procedure).

The differences in intercepts and slopes with few repeated measures suggest a linear growth curve is most appropriate (Raudenbush and Bryk, 2002). If there were a true time period that was continuous, a polynomial model would fit the data best as we do see curvature in the average outcome variable over time (Figure 3.1), and fitting the model with a numeric 3-point scale we have a significant quadratic term. As there are three repeated measures, we modeled the time measuring points as categorical factor. We did this to see where the most change/growth was in the clients over time and to be able to assess whether there were differences in CIS Psychological score at the three milestones of the program.

When there are few observations per individual, say three or four, using hierarchical modeling is convenient (Raudenbush and Bryk, 2002). In such a model, there is generally a growth rate parameter for each person over the data-collection period which represents the expected change during a fixed unit of time. There is an intercept parameter that tells us the estimated ability of a person at all other parameters fixed as 0. These growth rate parameters and intercept are allowed to vary at the level-2 model, which is a function of measured person characteristics. The individuals are the grouping variable and scores are nested within these individuals. The individual is at level-2, with the repeated measures at level-1. Since we have measures on all variables in the analysis for the specific times, we could use time series procedures but we only have three time points, and most often only two, so time series procedures are not appropriate to use. We then model the intercept of the level-2 model at the level-1 site level as we have modeled the intercept as random so covariates are permitted to vary with the intercept at all 1evels.

Throughout this chapter we will focus on the model building process and analysis for only the CIS Psychological score and provide output and estimates for this model, as the same process is done for the CIS Physical score (which is beyond the scope of this thesis but was also examined in the broader study). The statistics estimated from this study are not intended to be used to infer from the sample to the general population, since we do not have a simple random sample or a control group to compare against. Instead we use the results to provide policy recommendations for programs such as PSH.

The analysis consists of a 3-level hierarchical model, where all levels of the model have random intercepts and fixed slopes. There is an analysis of the unconditional model, where there are only intercept and error terms. Followed by the Level-1 model which is estimating the Time invariant variables. Then the Level-2 model, Client, where there are individual characteristics in this model. The Level-3 model, Site, where there are no covariates associated found to be Site level predictors. Lastly, the full model is assessed, which includes all the covariates from all the three sub-level models.

The model assumptions for model fit are assessed for each of these models, and then the results are analyzed. A final presentation of what each of the levels tells us individually and together is used to summarize the results. The important information we get from these results are in the examination of the variance components at each level.

3.3 Model building

In this section we illustrate how and why we chose to model the data with the current covariates seen in the inferential (reduced) model. We first discuss variable selection and then model selection. After liaising with subject matter experts (SME) on a preliminary assessment of the data and hypotheses posed for the analysis, an exploratory data analysis was done. The two primary questions of interest were:

1) Did the CIS Psychological score change over time?

2) What factors explained these changes over time?

The purpose of the exploratory analysis for this thesis is to formulate hypotheses to test and understand our data more thoroughly to assess the questions. The first assessment of the data was done using OLS procedures; by simple linear regression (SLR) models, and then by multiple regression (MR) models. These were done to assess variable and model selection before modeling the data in at hierarchical structure and algorithm under REML.

3.3.1 Variable selection

First, we examine the data set with epidemiologists and psychologists to select from the full set of hundreds of survey questions that were given to the clients during the study to fewer than 50 variables. We then created and used this set of about 50 variables to continue assessment for model selection. We look at the univariate frequency distributions of these variables to check for outliers or see if there are appropriate variable transformations. We look at the frequency distributions and tables of the categorical variables to see if such a variable had enough variability among its levels to be useful, as unbalanced data can lead to more bias in the estimates if not equally weighted (Raudenbush and Bryk, 2002). Next, bivariate relationships were assessed to identify nonlinear relationships among the covariates and response. This can also be used to identify variable transformations. Bivariate relationships between the covariates were looked at through correlation. High correlations among covariates causes variance inflation in the estimated regression coefficients of a linear model (Christensen, 2016), and we want to check if they have similar explanatory power or not. The correlation among the covariates and outcome is to check if the variable is a linear predictor of the outcome or not.

We checked the collinearity of the variables in the model through a variance inflation factor on each level and full model. Collinearity mostly affects model fit and predictability as the above estimates are used in such assessments and can be dealt with by combining those variables that have high collinearity (Christensen, 2016). After consulting the same subject matter experts, we combined variables with similar explanatory power or discarded those with “bad data”. These variables were selected as they had little to no missing values in their fields, respectively, were well balanced, did not exhibit collinearity with each other, had no high correlations among the covariates and outcome, SME knew that these were common factors in analyses similar to ours, and were likely going to be found to be meaningful in explain our data. Once the data set had near 20 variables we began model building and model fitting while cautiously specifying the model with the SMEs so that we were not overfitting a model with so many covariates needing assessment. For example, the variable PTSD is known to be a robust and reliable measure and would make for a good covariate to be checked in our model and that variables like “anxiety” and “depression” should not be treated as a covariate but rather a control.

3.3.2 Model selection

Model selection helps make an initial “complex” model more simple by removing variables that do not explain variation in the response and thus makes interpretation easier. Model selection helps to rank a set of candidate statistical models based on information-theoretic tools such as Akaike information criterion/Bayesian Information Criterion (AIC/BIC) or step-wise regression (forward, backward, or both) which uses hypothesis F-test statistics (Bolker et al., 2009). Information-theoretic approaches such as AIC and BIC measure expected predictive power and the smaller the better the model fit (Christensen, 2016). Typically, information-theoretic approaches are preferred to hypothesis testing and are most commonly used in model selection and we use the AIC value as indication (Christensen, 2016). We are aiming to find a model with the smallest AIC and this was done through manual backwards selection. We removed those variables that had large p-values (greater than 0.05). As we removed one variable from a model we examined AIC. This was done individually for all variables and interactions from the full model discussed in Chapter 2.

An important part in statistical analyses is that the covariates are centered or also standardized, as centering can decrease standard errors and should be used in hierarchical modeling (Bickel, 2007). One can either use group or grand mean centering, and sometimes using raw scores is most appropriate (Bickel, 2007). We used grand mean centering for the numeric covariates where centering was reasonable, as working with random coefficients and hierarchical regression models, grand mean centering of all the independent variables is best (Bickel 2007). This is because the intercept and slopes in the level-1 model become the outcome of the level-2 model and we have better interpretability and understanding of centered variables (Raudenbush and Bryk, 2002). The intercept of the prior level is then the expected outcome for a subject whose value on some covariate is equal to the grand mean (Raudenbush and Bryk, 2002; Bickel 2007).

If independent variables change over time, then it is best to center them, otherwise use dummy variables to indicate baseline values. When working with categorical variables use dummy variables to indicate the true value and there is little need to center these variables (Raudenbush and Bryk, 2002; Bickel 2007). The numeric variables Age, Income, and PTSD score are grand mean centered. The other covariates, Gender, Ethnicity, Education, Hispanic-

Latino, and Diagnosis are categorical variables and each has a specified baseline. The baseline for Gender is male as we see more male people experiencing homelessness than we do females. The baseline for Ethnicity is Hispanic-Latino as the study is done in New Mexico and there is a large population of Hispanic-Latino people who are experiencing homelessness. The baseline for Education is to be that the client has completed high school, as that is a common baseline in epidemiological and behavioral health studies and a general expectation for and from the population. The baseline Diagnosis is depression as that is the most common in our sample and most well-known to the public and a common diagnosis to the public.

The variable Health Status is treated as a numeric covariates in the models, and is made from Likert scale, so it makes no sense to standardize or center this variable. This is because, for example, the difference in reporting “agree” and “strongly agree” is the same measure difference of “disagree” and “strongly disagree”, and there is no true zero for this variable. The variable Interact with Family and/or Friends is a binary variable and treated as a dummy variable indicating at “yes”. The variable Attend is modeled as categorical and not centered and done to see where the true differences lie in how often one is attending NA, AA, church, and/or other self-help groups, and where there are differences.

Through the model building process, we assess the adequacy of the hierarchical model and sub-models. The first model being considered is an unconditional model. Again, we use random effects as we have variables describing individuals and these individuals are grouped into larger units, each consisting of a number of individuals. We assume that each of these groups have different regression models, or an overall regression that each individual will differ from. Since these groups are also sampled we assume the intercepts and slopes are a random sample from a population of group intercepts and slopes; this is what a random coefficient model does. Since we only have a random intercept model (and sub-models), we only have variance components models as some variables are not observed which implies they wash out into the error term causing correlation between disturbances (non independence). These disturbances have a group and individual component where individual components are independent and group components are correlated within groups, some more than others. We also only model the parameters associated with the intercept parameters as we have a random intercept model. That is the intercept at level-1 is permitted to vary among level-2 units and so on.

To present such a high number of models in the model selection process would necessitate the checking and adjustment of each effect in order to account for all changes in AIC and p-values. This would result in an excessively large table illustrating AIC values for model fit. Table 3.4 demonstrates that the full model is the best in regard to AIC while, in addition, containing the most parameters, fixed and random, to explain variance and outcome. Make note that there is at least a 4-unit difference in the reduced and full models at all levels.

Table 3.4: AIC of the Full and Reduced model used inferentially

Model	AIC Full	AIC Reduced
Unconditional	4456.4	4456.4
Level-1(Time)	4327.0	4315.5
Level-2 (Client)	4359.4	4355.7
Level-3 (Site)	4363.3	4363.3
Full	4322.1	4303.9

3.4 Assessing Model Fit

After model selection, we examined the model diagnostics: linear relationships, errors are normally distributed, homoscedasticity of errors (or, equal variance around the line), and independence of the observations. Diagnostics need to be assessed on all levels, and the full model, in the hierarchy. Thus, looking at diagnostic plots and statistics of an unconditional model, the level-1 model, level-2 model, level-3 model, and full model with contextual covariates included at each level. The models fit were done so in SAS PROC Mixed. We fit full models in this procedure that were found from the multiple regression models.

The diagnostic tools used to assess the quality of model estimation of the data described by the model are those tools one uses in standard regression models. Although, these diagnostic tools must be adjusted to reflect the dependence introduced by the nested data structure. Residual analysis now includes the assessment of distributional assumptions at each level of the model. This requires the use of level-dependent residual quantities. Similarly, the parameter estimates may be influenced at each level of the model, requiring influence diagnostics. The assumptions we the model needs to meet and that need assessment are distributional assumptions, covariate structure (handling of such), and metric of outcome variable measurement (common survey are all times). The within subject variation of the model doesn't assume uniform data collection designs across subjects but the between subject variability can accommodate different covariate structures for all intercepts and slopes. Hierarchical modeling doesn't require the same data collection design for each individual but same at each time point is required, which is true in our design, also including the same design for all individuals.

In hierarchical modeling, using a response that is logistic or continuous among the real line is preferred, as it has better interpretability properties and the data can follow a normal distribution (Douma and Weedon, 2019). The following diagnostics must all be taken cautiously as this data is nested and because we have few cluster units in the sub-models. Raudenbush and Bryk (2002) note that homogeneity may cause more bias as the standard errors will increase, if there are fewer than 10 units per group. We do not impute values over time as there are many lost cases due to attrition. Again, the main uses of this modeling is that it handles unbalanced, nested data well, even with high attrition rates, and allows for dependent observations (Raudenbush and Bryk, 2002; Bickel 2007).

We begin by looking at the correlations between the outcome and all numeric covariates to see if the covariate is a good linear predictor or not. All the correlations of the CIS Psychological score and covariates are small indicating they are weak linear predictors for CIS Psychological score, as seen in Figure 3.2. We see that the overall highest absolute correlation is between the outcome, CIS Psychological score, and PTSD, where $\rho = 0.22$ which does not cause for concern. This moderate correlation makes sense as the PTSD score is a direct measure of mental health in the homeless population (Cirsanti et al., 20121) and we are measuring an outcome that is about ones' mental wellbeing. We see that the highest correlation among the numeric predictors is 0.198 which is slightly positive but not large enough to assume that the variables are not independent or having similar explanatory power. The linear correlation among the rest of the variables is practically null. These variables do not need a transformation because they are normally distributed and they are not skewed.

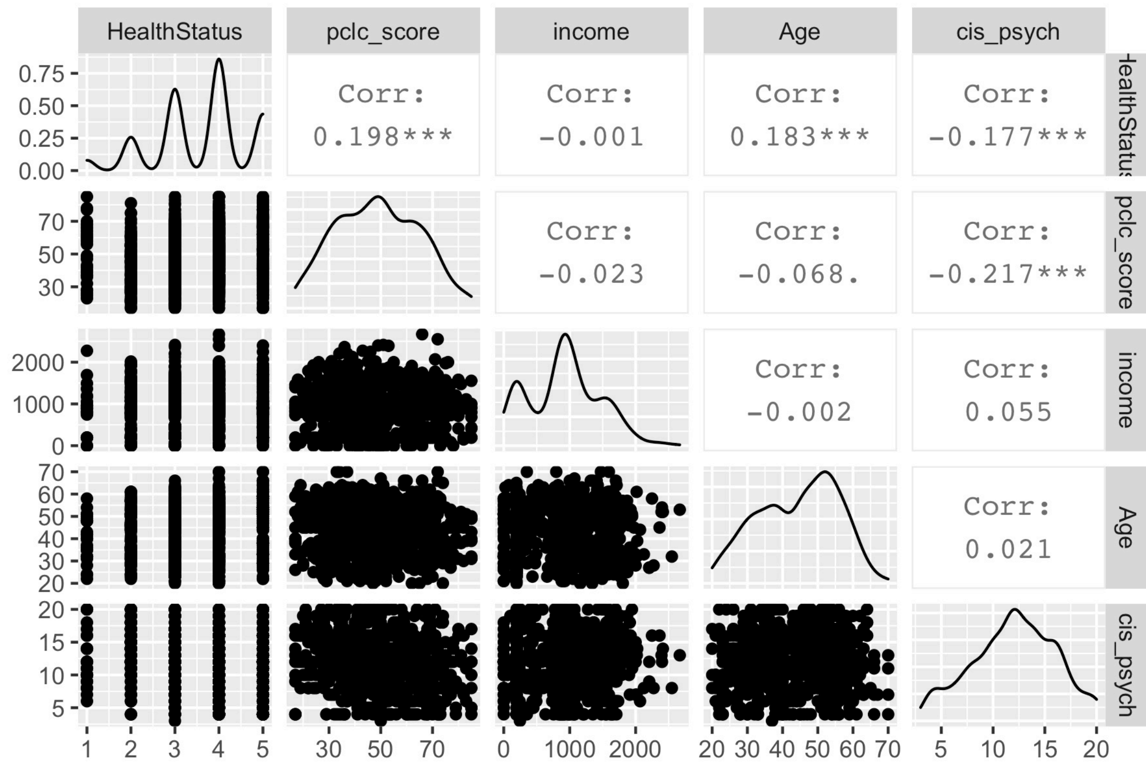

Figure 3.2: Correlation Matrix among all numeric variables for all levels in the inferential models (Health Status, PTSD Score, Income, Age, and CIS Psychological Score)

Now assessment of multicollinearity in the model is examined to see if collinearity exists or is too large to ignore. We do so by looking at the Variance Inflation Factor (VIF) of each model and the variables with the associated model. A general guideline is that a VIF larger than 5 or 10 is large, indicating that the model has problems estimating the coefficient. However, this in general does not degrade the quality of predictions. If the VIF is larger than $1/(1 - R^2)$, where R^2 is the multiple R-squared of the regression, then that predictor is more related to the other predictors than it is to the response. We compare the model VIF to each variable VIF for each model, respectively. We want to see that the variable VIF is lower that the model VIF.

In Table 3.5 we see that all of the sub-models and full model have VIFs under 2. We also see that all the variable VIFs are less than the respective model's VIF. This suggests that there is little collinearity and not high enough to be concerned with estimability problems or collinearity problems. The unconditional model cannot be assessed in this way as there are no covariates and the level-3 model only has one covariate so there are no issues with variables being too linear with each other.

Table 3.5: Model VIF to assess collinearity for all levels of the hierarchy

Model	**Model VIF**	**Variable VIF**	**DF**
Level-1	1.184133	Time: 1.026390 PTSD: 1.052575 IFF: 1.016602 Health Status: 1.044364	2 1 1 1
Level-2	1.113958	Time: 1.004980 Gender: 1.004618 Education: 1.007011	2 1 2
Level-3	N/A (only one covariate)		
Full Model	1.211974	Time: 1.031372 Gender: 1.018078 Education: 1.028969 Health Status: 1.051597 IFF: 1.031058 PTSD Score: 1.065457	2 1 2 1 1 1

Now we assess diagnostics to validate the assumptions on the residuals of the models and diagnose any influential observations. To check for normality and equal variance, there are a set of common diagnostic plots that can help assess these assumptions; qq-plot, residual vs. fitted, box plots for equal variance, added variable plots, and histograms of the residuals. It is also important to check for influential observations by looking at Cook's distance and leverage plots for the fixed and random portions of the model.

The major difference in the assumptions in OLS and in hierarchical modeling is that the errors are permitted to vary in the hierarchical models, because the error terms are more complex as they depend on each other (Bickel 2007). Observations within one cluster tend to be more alike with each other compared to observations within other clusters, violating a well-known regression assumption of observation independence (Cohen, Cohen, West, & Aiken, 2003). In our case, the errors for one individual might also be dependent. The errors are dependent within each site because errors are common to every client within the same site. This is also true as errors are dependent within each client within the repeated measures. We can loosen the constraints on the errors having common variance. This is because the errors measured at each level are then represented in a final model where there is a dependence on the covariate, which vary across sites and we exhibit the same behavior at the time and client levels.

In the mixed model, the validity and tenability of the assumptions pertain to both the structural and random components. For the structural part of the model, a properly specified model is that there the outcome is a linear function of the regression coefficients. Misspecification of the structural part of the models happens when at least one component in the error terms are associated with at least one other covariate in the model (Raudenbush and

Bryk, 2002). These specification and misspecifications, in a hierarchical model occur and must be validated at each level of the model. Misspecification at one level can affect the results at another, and because of correlated errors, we see that one equation can bias the estimates in another equation at other levels (Raudenbush and Bryk, 2002). Misspecification most often occurs when level-1 covariates are related to the outcome or other covariates, and are dropped, leading to inflated bias of the fixed effect coefficients (Raudenbush and Bryk, 2002). Thus, level-2 and higher level fixed effect estimates will be biased (Raudenbush and Bryk, 2002). The misspecification of models can be assessed through covariance testing where we want to see that the covariance among all the estimates and errors are zero (SAS/STAT® 14.1 User's Guide: The MIXED Procedure). If there are strong correlations among covariates (confounding variables), and is ignored, estimation of each level in the hierarchy will have bias in the estimates and this bias depends on the predictive power of that confounder (Raudenbush and Bryk, 2002). Formally, we assume the following:

1) Each error, for the purpose of this thesis, ε_{tis}, is independent and normally distributed with mean 0 and variance σ^2 for every level-1 unit t within each level-2 unit i and every level-2 unit i within each level-3 unit s.

   That is $\varepsilon_{tis} \sim iid\ N(0, \sigma^2)$

2) The level-1 covariates are independent of ε_{tis}

That is $Cov(X_{ptis}, \varepsilon_{tis}) = 0\ \forall\ t$, where X_{ptis} represents each p covariate

3) The random errors at level-2 and level-3 are multivariate normal, each with mean 0 and some variance τ_{eee} and some covariance among the random elements τ_{rrr}. The random error vector is independent among the second and third level units. That is the random error vectors among the client level-2 units are independent normally

distributed with mean 0 and the random error terms among the site level-3 units are independent and normally distributed with mean 0.

4) The set of level-2 covariates are independent of every error term at level-2 and the set of level-3 covariates are independent of every error term of the level-3. That is the covariance between the covariates of each level with their errors is 0.
5) The errors of level-1 and level-2 are independent, the errors of level-2 and level-3 are independent. That is the covariance of the level and subsequent levels is 0.
6) All the covariates at each level are not correlated with the random effects at all other levels.

Note that assumptions 2, 4, and 6 are about the relationships between the variables included in the structural portion of the model (the design matrices) and their error terms. Their tenability affects the bias in estimating the multilevel intercept and slopes at further levels (Raudenbush and Bryk, 2002). Assumptions 1,3 and 5 are on the random error terms and their tenability affects the consistency of the estimates of the standard errors in the multilevel intercept and slopes at further levels; the accuracy in estimating variances, slopes, hypothesis test and confident intervals (Raudenbush and Bryk, 2002). One way to deal with heteroscedastic and autocorrelation is by selecting how to model covariance structure (Bickel 2007).

When thinking of these assumptions in terms of our analysis, we assume the following:

1) Conditional on the time invariant variables, the within-client errors are normal and independent with mean 0 for each client and equal variance across clients. Conditional on the client level variables, the within-site errors are normal and independent with mean 0 in each site and equal variance across sites.

2) The time invariant covariates of CIS Psychological Score growth at level-1 are independent of the error term. All the excluded covariates are not correlated outcomes.
3) The residuals of Client effects are assumed to be bivariate normal with respective variance and covariances. The residuals of Site effects are assumed to be multivariate normal with respective variance and covariances.
4) The effect of whatever client and site covariates are excluded from the model for the intercept for time invariant covariates slopes are independent of the covariates in the client and site level models.
5) The error at level-1 is independent of the residual client effects. The error at level-2 is independent of the residual site effects.
6) The covariates of the time varying model that are excluded are relegated to the error term at level-1, are independent of the level-2 covariates in the model. The covariates of the client level model that are excluded are relegated to the error term at level-2 are independent of the level-3 covariates in the model. Also, whatever client level covariates are uncorrelated with time invariant covariates and site level covariates is uncorrelated with client level covariates.

Assumptions 1, 2, 3 and 4 can be assessed by examining the data and residuals at each individual level. Assumptions 5 and 6 are concerned with cross-level associations and need closer examination through time series procedures and likelihood ratio tests. We also assume that the growth parameters vary across individuals (client covariates vary). We need to check that the fixed effect is significant and slope heterogeneity tested through likelihood ratio test, chi-square and by reliability of each parameter. When reliabilities are small the variances are likely to be close to 0 causing numerical difficulties as we are near the boundary of the parameter space (Raudenbush and Bryk, 2002).

To check for independence, a time series analysis may be appropriate, check rational subgroups of the residuals and see if there are clear different subgroups-and if so, then the residuals are not independent. If one can get a control chart and randomly group the residuals and find the groups look the same, then they are independently distributed. We also plot residuals against any time variables and any factors or regressors, where a pattern that is not random suggests lack of independence.

Presented in Table 3.6 is a summary of diagnostic statistics for the models. We use the Anderson-Darling (AD) test to determine that the sample of data is drawn from a normal distribution with the null hypothesis that the data is normally distributed. As the p-value is near 0 for levels-2 and -3, we reject the null, suggesting that the data is not normally distributed at these levels. The level-1 and full model meet the normality assumption from the AD test. The power transformation from the Box-Cox for all the models was suggested to be just above 1, and seen in Figure A.8 in the Appendix, no transformation for the response variable would be helpful in meeting the normality assumption. The standard square-root and log transformations we looked at for the response variable, but those did not help, either. We see that the p-value for the AD-test of normality is increasing with a more complex model.

Table 3.6: Model Diagnostic Statistics for all Levels of the Hierarchy

Model	**Anderson Darling (AD)**	**AD P-Value**	**Box-Cox**
Unconditional	4.326	9.49e-11	1.07
Level-1	0.41889	0.3266	1.07
Level-2	0.74808	0.05127	1.07

Level-3	1.956	5.478e-05	1.07
Full Model	0.26911	0.6799	1.07

The reliabilities, seen in Table 3.7, of the level-1 coefficients are most important to us as reliabilities will be close to 1 when the group means vary substantially across level-2 units or if the sample size is large (Raudenbush and Bryk, 2002). This measure is defined as the parameter variance divided by the parameter variance plus the error variance. The most reliably measure is PTSD and the least reliable measure is IFF, for the level-1 model. Generally, we want the reliability to be at least 0.10, and only Time and Health Status meet that criterion. For the level-2 and level-3 model Time is the only moderately reliable measure. The full model has Time, Health Status and PTSD as reliable measures.

Table 3.7: Reliability of each parameter for each level for independence assessment

Variable [Model]	**Reliability**
Time [1]	0.17
Health Status [1]	0.10
IFF [1]	0.023
PTSD [1]	0.963
Time [2]	0.17
Gender [2]	0.025
Education [2]	0.07
Time [3]	0.17
Time [Full]	0.18
Health Status [Full]	0.10

IFF [Full]	0.022
PTSD [Full]	0.963
Gender [Full]	0.025
Education [Full]	0.07

3.4.1 Model ICC for Diagnostics and Variance Beak-down

First we look at the covariance values of the unconditional model to calculate the unconditional ICC and examine how the variance is being distributed between- and within-clients and sites. There are two unconditional ICCs for this model, as there are three levels in the model. We assess these two ICCs separately and then we can add these values to get a grand unconditional model ICC. We get these variance estimates from the covariance parameter estimates, seen in Table 3.9.

The between groups variability is defined as between group (clients OR site variability)/[total variability (within clients variance(residual) + between clients within sites variance + between site variance)]. We first look at the between client variability which is

$$ICC_i = \frac{\sigma_i^2}{(\sigma_s^2+\sigma_i^2+\sigma_t^2)} = \frac{3.2655}{3.2655+0.1774+12.4541} = 0.205. \qquad (3.\ 1)$$

This means that 79.5% of the variability in CIS Psychological score occurs within clients, while 20.5% occurs between clients. We note that between clients' variance is statistically significant, so the $\mathrm{ICC_i}$ is statistically significant. This means that at least a 2-level hierarchical model is necessary. We next look at the between site variability which is

$$\mathrm{ICC_s} = \frac{\sigma_s^2}{(\sigma_s^2+\sigma_i^2+\sigma_t^2)} = \frac{0.1774}{3.2655+0.1774+12.4541} = 0.0112. \qquad (3.\ 2)$$

Meaning that 1.12% of the variance in CIS Psychological score occurs between sites, leaving 98.98% of the variance is explained within the sites. We note that the between site variance is not statistically significant, meaning ICC_s is not statistically significant. This tells us that adding a third level to the hierarchy adds no power in the distribution of residual variance and that a 2-level model is sufficient.

We can now calculate how similar clients within the same site are like each other. We divide the between-clients-within-sites variance by the sum of the between-clients-within-sites and between-sites variance:

$$ICC_{is} = \frac{\sigma_i^2}{(\sigma_s^2+\sigma_i^2)} = \frac{3.2655}{3.2655+0.1774} = 0.9485 \qquad (3.\ 3)$$

meaning that 94.85% of the between-clients variance in CIS score occurs across sites. Similarly, if we want to know the total amount of variance at each time point that is due to grouping within clients and sites, we add the between-clients-within-sites variance and between sites variance divided by total variance:

$$ICC_t = \frac{\sigma_i^2+\sigma_s^2}{(\sigma_s^2+\sigma_i^2+\sigma_t^2)} = \frac{3.2655+0.1774}{3.2655+0.1774+12.4541} = 0.2166 \qquad (3.\ 4)$$

, meaning 21.66% of the total variability in CIS Psychological score is due to time being grouped in clients and in sites. We note that most of the variability due to grouping occurs at the client level. We also note that about 5.2% of variability exists between-sites and is not significant. This means that only a 2-level model is needed. We use the site level model as we are at the cut off and just about 5% and we are wanting to consider the fact that people are nested within site, which may account for some of the homogeneity in the models. We note that sites were not randomly assigned also suggesting that a third level may be inappropriate. Very weak ICC can substantially deflate standard errors of regression coefficients and multilevel coefficients will have larger standard errors (Bickel 2007).

The ICC for the conditional models, level-1, level-2, level-3, and full model are calculated the same way. We just need to make note of the variance explained (the residual variance and covariance parameter estimates decreasing or increasing). Table 3.8 shows the ICC for these models; the ICC for clients (ICC_i), site (ICC_s) and a grand ICC (ICC_t). We see that overall the site level model offers no help in partitioning the variance into the site random effect. The most variance within-subjects is explained in the client level.

Table 3.8: Model ICCs

Model	ICC_i	ICC_s	ICC_t
Level-1	0.270	0.020	0.289
Level-2	0.294	0.011	0.304
Level-3	0.304	0.010	0.314
Full Model	0.253	0.020	0.273

<u>3.4.2 Visual Diagnostic Assessment</u>

This section describes the model diagnostic visual tests. We go through the same assumption checking and diagnostics for all models. The figures referred to in the section can be found in the Appendix, at the end of the document.

In Figure A.1 of the unconditional model normality diagnostics we see that there are strong deviations from normality in the Quantile plot and in the residual histogram. This is common for an unconditional model of nested data (Bickel 2007). We are not concerned with these diagnostics as much as we are for the level-1 to -3 and the full model. As such, we will not discuss the diagnostics of the unconditional model too thoroughly. We note that the variance between sites and their means are not too different seen in the box plots in Figure

A.2. We note that there is random noise in the time series plots indicating that there is evidence of independence in the residuals, and that there are no influential points with high leverage in either the fixed effects or random effects seen in Figure A 4 to Figure A.7.

We use the Anderson-Darling (AD) test to determine that the sample of data is drawn from a normal distribution with the null hypothesis that the data is normally distributed. As the p-value is near 0, we reject the null, suggesting that the data is not normally distributed. The test supports the diagnostic plots, that the CIS Psychological score is not normally distributed amongst the residuals in the unconditional model. These diagnosits lead us to look at transformations in the response, and re-assess normality to asuure that infrences and estimates will be as unbiased and best as can be predicted, under correct model assumptions. The standard square- root transformation and log transformations were checked, as well. There we no transformations that fixed high tails in the errors. The power transformation plot for the unconditional model, level-1 model, level-2 model, level-3 model and full model looks the same. This is because the response variable needs no power transformation as guided by a Box-Cox test, with a power of 1.0707 and the interval is just beyond 1 for all levels in the hierarchy. This means that no transformation would normalize the data for any of the models.

Next, we will take a look at the level-1 diagnostics for the model. This model includes four covariates that are understood to be time-invariant, Health Status, PTSD and Interact with family and Friends. In Figure A.9 we see that there are no large tails in the histogram or in the Quantile plot. This is because our scale for the response variable is a discrete point scale from 4 to 20 and made form Likert data. There are no extreme deviations from normality in our level-1 model as there are small amounts of residuals under the diagonal in the Quantile plot, and these tails are short. In Figure A.10 we see that the raw distribution of the CIS Psychological Score for all people in the three sites has about equal variance, this is also found in Figure A.11 of the residuals of the CIS Psychological Score for all clients in all three sites. In both Figure A.12 and Figure A.13 we see that there is random noise for the raw CIS Psychological Score and for the residuals of the level-1 model, meaning that the variables are independently normally distributed.

From Figure A.14 we do not see influential points in the fixed or random effects. Points that would cause concern are those that are well above 2 in the Cook's D plots and well above or below 1 in the covariance ratio plot also if $|COVRatio - 1| > \frac{3p}{n}$ (SAS/STAT® 14.1 User's Guide: The MIXED Procedure). Thus, we do not delete any observation as there is no concern of inflation of standard errors from leverage points. There are plots of the fitted versus the standardized residuals that show independence of the residuals.

Figure A.15, Figure A.16, and Figure A.17 also show that there are no points that are strongly influential and could be influencing the assumption of normality or not. These plots also show that there are no transformations needed amongst the covariates that would greatly change the non-violation of the normality assumption.

Next we assess the model assumptions for the level-2 model, which models the person level contextual variables. This model has two variables that pertain to the clients, Gender and Education, but are not time-invariant. The only parameter in this models that changes with time is Time itself, and of course the error. Again, we examine any extreme and or influential observations, normality of errors, constant error variance, multicollinearity, non-linearity, independence of errors, and transformations in covariates. The plots to most consider are the histogram of residuals and the Quantile plot in Figure A.18, seeing that there are some deviations from normality. All other plots are similar in their interpretation to the plots for the level-1 model in that they are within the constraints of the assumptions. See Figure A. 19 to Figure A. 26 for visual assessment of the level-2 model.

Next we examine the model fit of the level-3 model; the site level, where there is only the Time covariate. Again, we examine the same set of plots for the same set of assumptions as previously stated. We see that the p-value for the AD-test of normality is decreasing with a less complex model, suggesting the residuals are not normally distributed. This is seen in *Figure A. 27*, as there are strong deviations from normality in the histogram and from the points being far from the diagonal (normal) line in the Quantile plot. All other plots are interpreted just like for level-1 and level-2 where there are no other notable deviations in equal variance or independence of the errors. There are no transformations for the Time or response variables that would help ease the violations of normality in this level, seen in Figure A. 28 to Figure A. 35.

Finally, we assess the model fit of the full model. The full model includes all the contextual variables from each level. That is, we include the level-1 covariates, the level-2 covariates and the level-3 covariates. We do not "double" covariates if they show up in more than one model, like Time, as it is a covariate for all levels as we are asses growth over time.

Again, we examine any extreme and or influential observations, normality among errors, constant variance, multicollinearity, non-linearity, independence of errors, added-variable plots for transformations in covariates and transformation in the response.

In Figure A. 36 we see that the histogram shows very little deviations form normality. The Quantile plot for our full model looks as good as one could get. There are very few points under the diagonal, and the overwhelming majority of points fall on the diagonal. The

Fitted versus Residual plot shows no clear pattern other than a flat line at 0, meaning no transformation of the response is necessary.

In the boxplot figures for this model, there is about equal variances for all the sites of all the clients, and the means of the CIS Psychological Score in the sites are about the same seen in Figure A. 37 and Figure A. 38. The time series plots indicate that the residuals and raw score are i.i.d, seen in Figure A. 39 and Figure A. 40. In Figure A. 41 until Figure A. 44 we see that there are no extreme influential points in the model for all the variables and show that there are no need for transformations in any of the covariates or in the response variable. Overall, we have a well fit final model and have met the model assumptions.

3.5 Results

Interpreting the results from hierarchical models is more complex than when interpreting results from OLS regression analysis. In the models, we have both fixed and random effects, as well as both numerical and categorical variables. Had all the variables in the models been numeric, interpretations could have been more simple. The primary use of the level-1, level-2 and level-3 models is to assess the variance. We will note the fixed effect parameters in these models as well as the differences in the factor levels, but their values at sub-levels are not as important as the output of the covariance parameters. The importance of the fixed effects comes in when we interpret the final full level model, which includes all the covariates.

3.5.1 Unconditional Model to match output

We first assess the fully unconditional model, where no covariate variables. This model represents how much variation in the CIS Psychological score is allocated across the three different levels; time, client and site. The unconditional model is as follows

$$CIS\_PSY_{tis} = \beta_{000} + \gamma_{0s} + \delta_{0s} + \varepsilon_{tis}, \quad \varepsilon_{0is} \sim N(0, \sigma^2) \qquad (3.\ 5)$$

β_{000} is the overall random intercept and value of the unconditional score;

γ_{0s} is the random effect of Client;

δ_{0s} is the random effect of Site;

ε_{tis} is the random error effect;

$t = 1_{is}, 2_{is}, 3_{is}$ for time of survey that correspond to intake, 6 months, discharge, respectively for each client i in shelter s;

$i = 1, \ldots, n_s$ clients which there are 370 of them at intake (286 at 6 months and 143 at discharge) in shelter s;

$s = 1{,}2{,}3$ for the 3 different locations of homeless shelters in the larger metropolitan New Mexico cities.

*Note that the estimate is the Variance not Standard Deviation as we would assume as the estimate come from a Normal distribution.

Table 3.9 : Covariance Parameter Estimates for the Unconditional Model

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	Site	0.1774	0.2559	0.69	0.2441
Intercept	Client	3.2655	0.8094	4.03	<.0001
Residual	Time	12.4541	0.8627	14.44	<.0001

The null hypothesis for the covariance parameters is that the variance is equal to zero (H_0: $\sigma^2 = 0$). When the p-value for the covariance estimate is smaller than the testing point, $\alpha = 0.05$, we reject the null hypothesis concluding the variance is not equal to zero. The estimate for the covariance parameter for the RE of Site is the smallest and statistically non-significant. This means this random intercept is not explaining the variance in the model.

Thus, Sites do not vary from one another. About 1/4th (3.27) of the variance is explained at the Client level, so modeling the clients as a RE is important as clients do statistically vary from one another. The most variance in the unconditional model is being explained through the residuals. Meaning that other treatments or covariates may explain the variance in the model. The total variance for the unconditional model is 15.897.

Table 3.10 : Solution for Fixed Effects for the Unconditional Model

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.0169	0.2914	2	41.24	0.0006

The estimate for the intercept from the fixed portion of this model is 12.017 points, which is the average score for all individuals for the CIS Psychological score. Note that the intercept is significant at an $\alpha = 0.05$ level as its p-value is less than α. The intercept is meaningful as the physical score has a range of $[4,20] \in R$. On average, the CIS Psychological score would start at 12.02 with some error. The confidence intervals in hierarchical models what include random coefficients are calculated by the intercept value-t_0.05*SE to intercept value+t_0.05*SE. As these intercepts are permitted from group to group, we describe this 95% range slightly differently. The 95% confidence interval for the intercept parameter is:

95%CI($\hat{\beta}_0$): $11.446 \le \beta_0 \le 12.588$.

Where one interpretation is that we are 95% certain, the population intercept for the unconditional model is between [11.5,12.6], and note that the estimated intercept (mean CIS Psychological Score) is 12.02 which is contained in the interval. The intercept parameter in mixed models is a mixed component itself as it is a combination of the fixed intercept plus random variability from group to group. We can interpret this intercept in that the random component variance for intercept is normally distributed, we know 95% of the distribution is included in the interval. If this interval was wider, we may be concerned and adding contextual variables, cross-level interactions, could narrow this interval (Bickel 2007).

3.5.2 Level-1 Model: Repeated Measures- Time Level

This model has covariates that are time-invariant; *Time, Health Status, IFF and PTSD Score*. There are no other covariates in this model. This is a random intercept model with fixed effects on all the covariates (slopes). The variable time and IFF are being modeled as categorical as there are only 3 points of repeated measured and 2 factors in IFF. This so that we can distinctly see what is happening, on average, over time for all clients; seen through difference of least squares estimates. Health Status and PTSD score are numeric covariates. The model is as follows:

$$CISPSY_{tis} = \boldsymbol{\gamma_{0is}} + \beta_1 HealthStatus_{tis} + \beta_2 IFF_{tis} + \beta_3 PTSD_{tis} + \beta_4 Time_{tis} + \varepsilon_{tis} \quad (3.\ 6)$$

$\boldsymbol{\gamma_{0is}}$ is the random intercept- the initial status of client is, that is, the expected outcome for that client at intake (Time=0), when IFF=0, Health Status is its grand mean and PTSD is its grand mean and includes the random effect for client and site;

β_1 is the score for client is when they have the average health status;

β_2 is the score for client is when they saw friends and family;

β_3 is the score for client is when they have the average PTSD score;

β_4 is the score for client is during the study time period;

and ε_{tis} is random error.

Table 3.11 : Covariance Parameter Estimates for the Level-1 Model

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	Site	0.2664	0.3476	0.77	0.2217

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	Client	3.6581	0.7100	5.15	<.0001
Residual	Time	9.6562	0.6741	14.32	<.0001

We first note that the total variance has decreased to 13.5807 from 15.897. This means that adding the fixed effects *Time, IFF, PTSD, and Health Status* variables, partitioned some of the variance into clients over time. We see that the residual variance decreased and the clients' covariance estimate increased. Again, we see that the RE for site is statistically non-significant, while the intercept for clients is.

We can again use the covariance parameter estimates to look at the conditional ICC. We can see if the variability between groups in reduced or not. The between sites variability is now 1.96% and between client variability is now 26.9% with the inclusion of the fixed covariates. Including the contextual cross level interaction terms in multilevel regression equation has produced a slightly higher proportion of variability between groups. This may lead us to not use these contextual variables.

We see from the ANOVA Table 3.12 that *Time, IFF and Health Status and PTSD* are statistically significant fixed effect in the level-1 model. As there are variables treated as categorical, we need to see where the significant differences lie among the 3 time points (repeated measures), and between the 2 levels of IFF. First, in Table 3.13 we see the that from the fixed effects, the intercept is significant and that time 1 (6-months) is statistically different from intake and that time 2 (discharge) is statistically significant from intake. In Table 3.14, we are able to see the actual estimates for each time point.

These averages are the averages of the 3 time points over all other variables where the numeric variables are set equal to 0 and the categorical variables are averaged over all the categories. The least squares estimators handle unbalanced categories and average over any and all the categories, without weighting the categories which may cause some biased estimates as the SE will increase for the smallest of the categories. We see that all the time points are significant in the model and that the biggest difference is between intake and 6-moths followed by intake and discharge. The smallest difference is from 6-months to discharge and is statistically non-significant. These pairwise differences are seen in Table 3.15.

In Table 3.13 we see that both 6-months and discharge times are significantly different from intake. We note that from intake to 6-months has a larger increase on one's overall CIS Psychological score than from intake to discharge. We see that since "5" being the most-poor health, that this negative in front of health status means that for every point increase in Health Status, CIS Psychological score decreases by 0.0552 points. Those who did not see friends or family have a decrease in score by 1.07 points and those who have higher PTSD scores lower their CIS score by 0.039 points. This is of course while all other terms are held constant.

Table 3.12 : ANOVA Table of the Fixed Effects for the Level-1 Model: Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Time	2	403	40.76	<.0001
IFF	1	403	14.43	0.0002
Health Status	1	403	16.92	<.0001
PTSD Score	1	403	20.28	<.0001

Table 3.13 : Solution for Fixed Effects for the Level-1 Model

Effect	IFF	Time	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			13.3065	0.6072	2	21.91	0.0021
Time: 6-months		1	2.1527	0.2545	403	8.46	<.0001
Time: Discharge		2	1.9578	0.3272	403	5.98	<.0001
Time: Intake		0	0	.	.	.	.
IFF	No		-1.0697	0.2816	403	-3.80	0.0002
IFF	Yes		0	.	.	.	.
Health Status			-0.5518	0.1342	403	-4.11	<.0001
PTSD Score			-0.03892	0.008642	403	-4.50	<.0001

Table 3.14 : Least Squares Means Estimates for the Level-1 Model Categorical variables

Effect	IFF	Time	Estimate	Standard Error	DF	t Value	Pr > |t|
Time: 6-months		1	12.9477	0.3728	403	34.73	<.0001
Time: Discharge		2	12.7528	0.4256	403	29.97	<.0001
Time: Intake		0	10.7950	0.3556	403	30.36	<.0001
IFF	No		11.6303	0.3870	403	30.06	<.0001
IFF	Yes		12.7001	0.3530	403	35.98	<.0001

Table 3.15 : Difference of Least Squares Means Estimates for the Level-1 Model Categorical variables

Effect	IFF	Time	IFF	Time	Estimate	Standard Error	DF	t Value	Pr > |t|
Time		1		2	0.1949	0.3346	403	0.58	0.5607
Time		1		0	2.1527	0.2545	403	8.46	<.0001
Time		2		0	1.9578	0.3272	403	5.98	<.0001
IFF	No		Yes		-1.0697	0.2816	403	-3.80	0.0002

3.5.3 Level-2: Client/Persons level

Again, in this model we have mixed effects as we have random intercept and fixed effects in the covariates variables of the model (fixed slopes).

$$CISPSY_{tis}=\boldsymbol{\delta_{0is}} + \pi_{01}Gender_{is} + \pi_{02}Education_{is} + \pi_{03}Time_{tis} + r_{0is} \qquad (3.\ 7)$$

$\boldsymbol{\delta_{00s}}$ is the random intercept- the initial status of client is, that is, the expected outcome for that client at intake (Time=0), when education is level is high school and gender is male and includes the random intercepts for client and site;

π_{01} is the fixed estimate for Gender;

π_{02} is the fixed estimate for Education;

π_{03} is the fixed estimate for Time;

and r_{0is} is random error.

Table 3.16 : Covariance Parameter Estimates for the Level-2 Model

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	Site	0.1514	0.2304	0.66	0.2555
Intercept	Client	4.2434	0.7925	5.35	<.0001

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Residual	Time	10.0504	0.7123	14.11	<.0001

Seen in Table 3.16, the total variance in the level-2 persons' level is 14.4452, which is a 1-unit increase from the time level model. Most of the explained variance is from the random effect of clients and is a significant effect. We note that the variance in this model is more than the variance in the more complex model. This model is explaining more variance than the unconditional model. We again look at the conditional ICC, where 1.0% of the variance in CIS Psychological score is between sites and 29.4% is between clients. Clients are still a significant RE.

We see in , that of the Fixed Effects for the Level-2 Model: Type 3 Tests of Fixed Effects we see that Time, Gender, and Education are significant fixed effects in the model. Make note that Education is marginally significant, but when removing this effect, the AIC value increased more than 4 units. Thus, keeping Education in the model as the primary indication for model fit is smallest AIC.

Table 3. 17: ANOVA table of the Fixed Effects for the Level-2 Model: Type Tests of the Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Time	2	404	52.54	<.0001
Gender	1	404	4.18	0.0417
Education	2	404	2.65	0.0718

Table 3.18 shows that females and males do statistically differ in their CIS Psychological score. We see that having a higher education (MT HS) is statistically different from only having a HS education. Having a higher education increases ones CIS

Psychological score by 0.9581 points. Table 3.19 tells us the estimate, again, on average over all other variables, for each factor in each variables in the model. In Table 3.20 we see that, again, HS and MT HS differ and Genders differ and all other differences are not significant.

Table 3.18 : Solution for Fixed Effects: Level-2 Model

Effect	Gender	Time	Education	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept				9.9877	0.4224	2	23.65	0.0018
Time: 6-months		1		2.4528	0.2567	404	9.55	<.0001
Time: Discharge		2		2.2664	0.3325	404	6.82	<.0001
Time: Intake		0		0	.	.	.	.
Gender	Female			0.6478	0.3170	404	2.04	0.0417
Gender	Male			0	.	.	.	.
Education: MT HS			1	0.8988	0.3911	404	2.30	0.0220
Education: LT HS			2	0.3820	0.3533	404	1.08	0.2802
Education: HS			3	0	.	.	.	.

Table 3.19 : Least Squares Means: Estimates for the Level-2 Model Categorical variables

Effect	Gender	Time	Education 1	Estimate	Standard Error	DF	t Value	Pr > |t|
Gender	Female			12.6355	0.3114	404	40.58	<.0001
Gender	Male			11.9877	0.3382	404	35.45	<.0001

Effect	Gender	Time	Education 1	Estimate	Standard Error	DF	t Value	Pr > |t|
Education : MT HS			1	12.7835	0.3679	404	34.75	<.0001
Education : LT HS			2	12.2667	0.3321	404	36.94	<.0001
Education : HS			3	11.8847	0.3650	404	32.56	<.0001
Time: 6-months		1		13.1914	0.3190	404	41.36	<.0001
Time: Discharge		2		13.0049	0.3831	404	33.94	<.0001
Time: Intake		0		10.7386	0.3014	404	35.63	<.0001

Table 3.20 : Difference of Least Squares Means: Estimates for the Level-2 Model categorical variables

Effect	Gender	Time	Education	Gender	Time	Education	Estimate	Standard Error	DF	t Value	Pr > |t|
Gender	Female			Male			0.6478	0.3170	404	2.04	0.0417
Education			MT HS			LT HS	0.5168	0.3658	404	1.41	0.1586
Education			MT HS			HS	0.8988	0.3911	404	2.30	0.0220
Education			LT HS			HS	0.3820	0.3533	404	1.08	0.2802
Time		1			2		0.1865	0.3419	404	0.55	0.5857
Time		1			0		2.4528	0.2567	404	9.55	<.0001

Effect	Gender	Time	Education	Gender	Time	Education	Estimate	Standard Error	DF	t Value	Pr > |t|
Time		2			0		2.2664	0.3325	404	6.82	<.0001

3.5.4 Level-3 Model: Site

We now model the final level of the hierarchy as this is the corset grouping variable assumed in the data. There are 2 random effects in the intercepts and 1 fixed effect for the slope of the time variable.

$$CISPSY_{tis} = \gamma_{000} + \beta_{00t}Time_{ts} + \mu_{00s} \quad (3.\ 8)$$

where $\boldsymbol{\gamma_{000}}$ is the average intercept across level-3 units;

β_{00t} is the effect of time in site s;

μ_{00s} is the unique increment of the intercept associated with level-3 unit s.

Table 3.21 : Covariance Parameter Estimates for the Level-3 Model

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	Site	0.1458	0.2286	0.64	0.2618
Intercept	Client	4.4366	0.7900	5.62	<.0001
Residual	Time	10.0238	0.7044	14.23	<.0001

In the level-3 model there is a total variance of 14.6064 (Table 3.21), which is slightly more than the level-2 model and more than the level-1 but less than the unconditional model. This means, so far, that the level-1 model is explain the most variability. We see again that time is a significant covariate and that time 0 and 1 are significantly different and that 0 and 2 are significantly different. We see that in Table 3.24 there is the highest, on average, score at

6-months. The difference in score from 6-months to discharge is slightly different (0.1743) but that difference is not significantly different.

We again look at the conditional ICC, where 1.0% of the variance in CIS Psychological score is between sites 30.4% is between clients. Note that Clients are still a significant RE, while Sites are not.

Table 3.22 : ANOVA of the Fixed Effects for the Level-3 Model: Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Time	2	406	52.46	<.0001

Table 3.23 : Solution for Fixed Effects for the Level-3 Model

Effect	Time	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		10.7935	0.2969	2	36.36	0.0008
Time: 6-months	1	2.4418	0.2562	406	9.53	<.0001
Time: Discharge	2	2.2675	0.3320	406	6.83	<.0001
Time: Intake	0	0	.	.	.	.

Table 3.24 : Least Squares Means: Estimates for the Level-3 Model categorical variables

Effect	Time	Estimate	Standard Error	DF	t Value	Pr > |t|
Time: 6-months	1	13.2353	0.3142	406	42.12	<.0001
Time: Discharge	2	13.0610	0.3792	406	34.44	<.0001
Time: Intake	0	10.7935	0.2969	406	36.36	<.0001

Table 3.25 : Difference of Least Squares Means: Estimates for the Level-3 Model categorical variables

Effect	Time	Time	Estimate	Standard Error	DF	t Value	Pr > |t|
Time	1	2	0.1743	0.3417	406	0.51	0.6102
Time	1	0	2.4418	0.2562	406	9.53	<.0001
Time	2	0	2.2675	0.3320	406	6.83	<.0001

3.5.5 Full Model: Time, Client, and Site level covariates included

Finally, we present the full model. This includes all the covariates from all 3 levels of the hierarchy.

$$CISPSY_{tis} = \gamma_{0is} + \beta_{00t}Time_{tis} + \pi_{01}Gender_{0is} + \pi_{02}Education_{0is} +$$
$$\beta_1 HealthStatus_{tis} + \beta_2 IFF_{tis} + \beta_3 PTSD_{tis} + (\mu_{00s} + r_{0is} + \varepsilon_{tis}) \qquad (3.\ 9)$$

where γ_{0is} is the score when all other covariates are set to their averages or baselines (the intercept) and includes the random effects of both Client and Site;

β_{00t} is the FE estimate for Time;

π_{01} is the FE estimate for Gender;

π_{02} is the FE estimate for Education;

β_1 is the FE estimate for Health Status;

β_2 is the FE estimate for IFF;

β_3 is the FE estimate for grand mean centered PTSD Score;

and r_{0is} the random effect of unit i on the mean, μ_{00s} the random effect of unit s, and the level-1 error ε_{tis}.

In Table 3.26 the total variance for the full model is 13.341, where the most explained variance comes from the random effect of clients within sites. We see that the residual variance is the second smallest in this model, where level-1 has the smallest residual variance by 1/100th of a unit point. We note that the site level variance estimate is largest in this model but is not a significant random effect. Below we see that all the covariates (fixed effects) are significant, noting that Education is marginally significant. When we removed this effect, the AIC was larger, thus we keep it in the full and final model. We again look at the conditional ICC, where 2.0% of the variance in cis score is between sites 25.3% is between clients. Clients are still a significant RE. We use a combination of the fixed effects estimates from Table 3.27 and the difference in least squares means to make our inferences about for the full model.

Table 3.26 : Covariance Parameter Estimates for the Full Model

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	Site	0.2680	0.3449	0.78	0.2186
Intercept	Client	3.3694	0.7051	4.78	<.0001
Residual	Time	9.7036	0.6827	14.21	<.0001

In Table 3.27 we see that all the variables in the model are significant, except for education, which is marginally significant. We further investigate how these variables are contributing to the CIS Psychological score. From the fixed effect solutions in Table 3.28 we want to take note of the estimates and p-values of the numeric covariates, Health Status and PTSD Score. Each 1-unit increase in Health Status corresponds, on average to a 0.5625-point decrease in the CIS Psychological score. This makes sense as better Health Status is rated lower, and worse Health Status is rated higher. For each 1-point increase in PTSD Score, on average the CIS Psychological score decrease by 0.03889-points. This again makes sense that there is a decrease as a higher PTSD score means worse and we expect the change to be this small as the PTSD score scale is 17 to 85, while the response is a 0 to 20 scale.

We now interpret the categorical variables in how they differ from one another. As seen in Table 3.29, those who interact with their family and friends have a 1-point higher score on average, than those who do not, and the difference is statistically significant. Females have on average, a 0.5 point higher CIS Psychological score than males, and the difference is statistically significant.

The following interpretations are on those statistics and estimates found in Table 3.29 and Table 3.30. Those with more than a high school experience have the highest on average score, and the difference in CIS Psychological score for those who have more than a high school education and those who have less than a high school education is 0.829 points. That is, those who have a more than less than high school education has a 0.829 point higher CIS Psychological score, and is statistically significant. The difference in CIS Psychological score for those who have more than a high school education and those who have a high school education is 1.154 points. That is, those who have a more than a high school education has a 1.154 point higher CIS Psychological score, and is statistically significant. The difference in CIS Psychological score for those who have less than a high school education and those who have a high school education is 0.325 points. That is, those who have a more than a high school education has a 0.325 point higher CIS score, and is not statistically significant.

We now examine what is happening over time, when including all the covariates variables, on average. The score at intake is the smallest averaged over all the variables, whole the CIS score at 6-months is the largest on average of over all the variables. The difference if CIS score from intake to 6-months is 2.158 and is statistically significant. That is, that from intake to 6-months the CIS score increase by 2.158 points. The difference if CIS score from intake to discharge is 1.949 and is statistically significant. That is, that from intake to discharge the CIS score increase by 1.949 points. The difference if CIS score from 6-month to discharge is 0.209 and is NOT statistically significant. That is, that from intake to discharge the CIS score decreases by 0.209 points. These results are congruent with Figure 3.1.

Table 3.27 : ANOVA table of the Fixed Effects for the Full Model: Type 3 Tests of Fixed Effects

Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Time	2	401	40.62	<.0001
IFF	1	401	17.23	<.0001
Health Status	1	401	17.82	<.0001
PTSD Score	1	401	20.39	<.0001
Gender	1	401	3.26	0.0717
Education	2	401	5.09	0.0066

Table 3.28 : Solution for Fixed Effect of the Full Model

Effect	Gender	IFF	Education	Time	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept					12.5870	0.6700	2	18.79	0.0028
Time: 6-months				1	2.1579	0.2550	401	8.46	<.0001
Time: Discharge				2	1.9485	0.3276	401	5.95	<.0001
Time: Intake				0	0	.	.	.	.
IFF		No			-1.1676	0.2813	401	-4.15	<.0001
IFF		Yes			0	.	.	.	.
Health Status					-0.5625	0.1332	401	-4.22	<.0001
PTSD Score					-0.03889	0.008612	401	-4.52	<.0001
Gender	Female				0.5415	0.2999	401	1.81	0.0717
Gender	Male				0	.	.	.	.
Education: MT HS			1		1.1540	0.3724	401	3.10	0.0021
Education: LT HS			2		0.3249	0.3367	401	0.96	0.3352
Education: HS			3		0	.	.	.	.

Table 3.29 : Least Squares Means: Estimates for the Full Model categorical variables

Effect	Gender	IFF	Education1	Time	Estimate	Standard Error	DF	t Value	Pr > |t|
IFF		No			11.5372	0.3880	401	29.74	<.0001
IFF		Yes			12.7048	0.3545	401	35.84	<.0001
Gender	Female				12.3917	0.3641	401	34.04	<.0001
Gender	Male				11.8503	0.3860	401	30.70	<.0001
Education: MT HS			1		12.7820	0.4075	401	31.37	<.0001
Education: LT HS			2		11.9529	0.3831	401	31.20	<.0001
Education: HS			3		11.6280	0.4082	401	28.48	<.0001
Time: 6-months				1	12.9101	0.3743	401	34.50	<.0001
Time: Discharge				2	12.7007	0.4267	401	29.76	<.0001
Time: Intake				0	10.7522	0.3572	401	30.10	<.0001

Table 3.30 : Difference of Least Squares Means: Estimates for the Full Model categorical variables

Effect	Gender	IFF	Education	Time	Gender	IFF	Education	Time	Estimate	Standard Error	DF	t Value	Pr >|t|
IFF		No				Yes			-1.1676	0.2813	401	-4.15	<.0001
Gender	Female				Male				0.5415	0.2999	401	1.81	0.0717
Education			MT HS				LT HS		0.8291	0.3487	401	2.38	0.0179
Education			MT HS				HS		1.1540	0.3724	401	3.10	0.0021
Education			LT HS				HS		0.3249	0.3367	401	0.96	0.3352
Time				1				2	0.2094	0.3348	401	0.63	0.5322
Time				1				0	2.1579	0.2550	401	8.46	<.0001
Time				2				0	1.9485	0.3276	401	5.95	<.0001

We now put all the information together in Table 3.31, Table 3.32 and Table 3.33 that break down inferential statistics for model selection and for variance explanation of the hierarchical model. As deviance is a measure of error, we are looking for smaller deviance difference. The smaller the better, meaning one model (the conditional model is preferred to the unconditional model). We, as always want a smaller AIC value and larger R^2 value. Looking at Table 3.30 there are multiple measures to assess model fit. We can see that the full model [5] is best besides the unconditional model based on R^2, which is not the primary statistic we use to assess model fit. The R^2 value is a summary statistic calculated by dividing the sum of residual and intercept for the conditional model by the sum of residual and intercept for the null model, subtracting the result from 1 and multiplying the result by 100 (Bickel). We also note that the full model is explaining more variance than any other model. If we base our decision on deviance difference, we see that the level-3 model or model 4 is the best model. The deviance statistic is calculated as the AIC values of the null model (unconditional model) minus the AIC of the conditional model. We use the AIC values of the null model and AIC values of all the conditional models –levels 1,2,3 and full- to calculate the deviance statistics. Overall, the full model is the best model for this data. This conclusion is supported by first off, the model diagnostics in previous sections of the chapter, from having the smallest AIC and largest R^2, and importantly explain (capturing) the most variance in the data.

We see overall that the models are explaining the variance differently. Certain models are accounting for significant portions of the variance and more so than others but that doesn't mean that including contextual variables were a waste of degrees of freedom. We have significant fixed effects in our models, so the benefit we get is equal to the cost of these degrees of freedom. In Table 3.32 and Table 3.33 we want to see that adding contextual variables is aiding in explaining the variance. We see that the model with the least residual variance is the level-1 model but close to that of the full level model. This means that adding in fixed effects helped to partition the variance in the fixed and random effects.

Table 3.31 : Model Selection Criterion Table for all 5 Models in this Thesis

Model [#]	**AIC**	**R^2**	**Deviance**	**Significant Parameters**	**Total Variance**

Unconditional [1]	4456.4	84.103%	NA	3	15.8970
Level-1(Time) [2]	4315.5	14.57%	141.1	6	13.5807
Level-2 (Client) [3]	4355.2	9.13%	101.2	5	14.4452
Level-3 (Site) [4]	4363.3	8.12%	93.1	3	14.6064
Full [5]	4303.9	16.08%	152.9	8	13.3410

Table 3.32: CIS Psychological Score: Random Intercept with Three Levels; Estimates of Covariance Parameters- A Level-by-Level breakdown of variance explained

Random Effect	***Variance Component***	***DF***	***Wald Z***	***P-value***
Level-1				
Variance				
Residual (ε_{tis})	9.6562	403	14.32	<.0001
Level-2				
Variance				
Client (r_{0is})	4.2434	404	5.35	<.0001
Site (r_{1is})	0.1514	404	0.66	0.2555
Level-3				
Variance				
Client (u_{00s})	4.4366	406	5.62	<.0001
Site (u_{t0s})	0.1458	406	0.64	0.2618

Table 3.33: CIS Psychological Score: Random Coefficient with Three Levels; Random Intercept for all levels; Estimates of Covariance Parameters- A complete breakdown

Parameter	***Estimate***	***Std. Error***	***Wald Z***	***Sig. Level***
Unconditional Model				
Site intercept variance	0.1774	0.2559	0.69	0.2441
Client intercept variance	3.2655	0.8094	4.03	<.0001
Residual Variance	12.4541	0.8627	14.44	<.0001
Level-1 Model				
Site intercept variance	0.2664	0.3476	0.77	0.2217
Client intercept variance	3.6581	0.7100	5.15	<.0001
Residual Variance	9.6562	0.6741	14.32	<.0001
Level-2 Model				
Site intercept variance	0.1514	0.2304	0.66	0.2555
Client intercept variance	4.2434	0.7925	5.35	<.0001
Residual Variance	10.0504	0.7123	14.11	<.0001
Level-3 Model				
Site intercept variance	0.1458	0.2286	0.64	0.2618
Client intercept variance	4.4366	0.7900	5.62	<.0001
Residual Variance	10.0238	0.7044	14.23	<.0001
Full Model				
Site intercept variance	0.2680	0.3449	0.78	0.2186
Client intercept variance	3.3694	0.7051	4.78	<.0001
Residual Variance	9.7036	0.6827	14.21	<.0001

It is important to note that these effects cannot necessarily be interpreted causally for observational data from a non-random sample from the population of interest. We could have found misleading conclusion and find contextual effects where none necessarily exist. We might have found that group-level correlations can be mistakenly attributed to individual-level causes, but our setting is slightly different in that both individual- and group-level data are available.

The Full Hierarchy of the Reduced Inferential Model

Level – 1: Time level with time varying covariates, random intercept and fixed slopes

$$\mathrm{C}ISPSY_{tis} = \boldsymbol{\gamma_{0is}} + \beta_1 HealthStatus_{tis} + \beta_2 IFF_{tis} + \beta_3 PTSD_{tis} + \beta_4 Time_{tis} + \varepsilon_{tis} \quad (3.\ 10)$$

where t, i, s denotes the cluster indices for Time, Client, and Site, respectively;

$\boldsymbol{\gamma_{0is}}$ is a random intercept, where there are 2 components making up this estimate-that of the random intercept for Client and random intercept for Site;

β_1 to β_4 are fixed slopes, respectively;

ε_{tis} the error term, such that $\varepsilon_{tis} \sim N(0, \sigma^2)$.

Level- 2: Client level with person characteristic covariates, random intercept and fixed slopes Here we only model the level-1 intercept at level-2 with covariates as we have a random intercept only model; this will apply to the level-3 models as well:

$$\boldsymbol{\gamma_{0is}} = \boldsymbol{\delta_{00s}} + \pi_{01} Gender_{is} + \pi_{02} Education_{is} + \pi_{03} Time_{tis} + r_{0is} \quad (3.\ 11)$$

$\boldsymbol{\delta_{00s}}$ is the average intercept across level-2 units;

π_{01} to π_{03} are the fixed slopes;

r_{0is} is the random error.

Level- 3: Site level with no covariates other than time, random intercept and fixed slopes

$$\boldsymbol{\delta_{00s}} = \gamma_{000} + \beta_{00t} Time_{ts} + \mu_{00s} \qquad (3.\ 12)$$

γ_{000} is the average intercept across level-3 units;

β_{00t} is the effect of time in site s;

μ_{00s} is the unique increment of the intercept associated with level-3 unit s.

Full Model: All covariates from all levels, random intercept and fixed slopes

$$CISPSY_{tis} = \gamma_{000} + \beta_{00t} Time_{tis} + \pi_{01} Gender_{0is} + \pi_{02} Education_{0is} + \beta_1 HealthStatus_{tis} + \beta_2 IFF_{tis} + \beta_3 PTSD_{tis} + (\mu_{00s} + r_{0is} + \varepsilon_{tis}) \qquad (3.\ 13)$$

which can be interpreted as the outcome CIS Psychological Score is function of the average regression equation plus random error having three components; r_{0is} is the random effect of unit i on slope, μ_{00s} is the random effect of unit s, and the level-1 error ε_{tis}. The estimates for this model are those found from the solution for fixed effect parameters in Table 3.28, Table 3.29, and Table 3.30.

Chapter 4: Discussion

4.1 What we learned from data and model

The findings from this study support that individuals experiencing homelessness with better-reported health status have higher community integration scores. Further, this study supports Yanos (2007) that those experiencing homelessness who have a higher education have higher community integration scores than those with lower educational attainment. We also found that those with higher PTSD scores had lower community integration scores, which is similar to those findings from Baumgartner and Herman (2012). We relate that those who interact with their family and friends increased their feeling of connectedness in their communities.

Implications are that those with better social networks and lower rates of PTSD felt better in their communities. We also found that those who attend more self-help groups and churches didn't predict CIS Psychological growth. We did not find supporting evidence that there were any discrepancies between women and men with depression who have higher income tied to higher community welfare. We found that women had higher CIS Psychological while averaged over other significant factors in this study. Overall, we found that better all-around welfare for the homeless community creates higher rates of community integration.

We see that those who have higher self-reported physical and psychological health have higher CIS Psychological scores at baseline. When there are self-help groups (i.e. alcoholic anonymous) and churches for individuals experiencing homelessness to attend at least once a month, they have a higher CIS *Physical* integration score, but there was no significant increase in the CIS Psychological score for any amount of attendances through the duration of the study. Allowing individuals to congregate with family and friends also increases these CIS scores. These increases in the CIS scores are seen to increase over time from intake to 6-months, while homeless individuals are housed. We note that those with higher educational attainment after high school increases CIS Psychological score more than those with less than a high school education. Thus, for those who are experiencing homelessness and who have permanent housing, we know which factors are influencing their self-reported CIS Psychological scores.

4.2 Limitations and Future work

Inference is only as good as how well your model fits the data, and how well the data is measured. Random sampling is important for inference from a sample to a population. In this observational study, we had no random design or control group, and our sample was not a random sample since the data is self-reported and survey data. A nonrandom sample may produce bias. Another source of bias is that self-reported data can produce inaccuracies from social desirability (Larson 2019). The data are also unbalanced, and there are attrition issues where the sample size decreases over time. Unbalanced data and time-varying covariates variables allure one to use HLMs to make accurate estimates and predictions, as OLS would produce inefficient estimates of fixed effects. As there is no cross-validation to assess if the analysis results could be generalizable to an independent dataset, we cannot use our model as a predictive model, which is one of the main goals in any statistical analysis.

A significant limitation to the analysis in this paper is that the *time* variable is not continuous, limiting the assessment of growth over time. We treat the variable as a categorical to assess what is happening between the repeated measures.

Another major limitation of this study is that the level-1 and level-3 models clusters were few. There are only three repeated measures and only three sites where we are grouping the data. Bickel suggests at least a 20/20 rule, meaning 20 observations in 20 cluster. Raudenbush and Bryk (2002) suggests that even 12 observations in a clustering group are small, and we have some number of observations in 3 clusters (91, 126 and 52, for example at intake). Small clustering units can impact biases and estimation in any direction in all levels of the hierarchical model. When modeling the variance components under REML when the number of subjects per group is few, there are higher biases (Stalter 2018). Slater also found that, in general, higher biases were found when the ICC was large. However, we do not see extremely high ICC in the unconditional or conditional models.

The work we'd consider to further this study would be to implement the suggested changes to the program, so there is more accessibility. This could give us an indication if the PSH program is improving the feeling of higher community integration and social embeddedness for those experiencing homelessness. Other work could include sensitivity analyses of a few types. One might be to run the analysis with a different covariance structure, like AR(1) or unrestricted to see the actual differences in the inferential statistics. One might be tempted to run the models under ML, but again, we have random effects and under ML these effects would be negatively biased. So you may trade precision and bias in that analysis. If there was an indicator for who the interviewers were for the surveys, we could possibly use that as a grouping variable and random effects or a level of itself. Adding levels to a multilevel model does become statistically cumbersome and needs to be well thought out, especially when higher levels in the model could give little to no information to the researcher and you could over scope the task at hand (Bickel 2007).

Chapter 5: Conclusion

5.1 Analytical Conclusion

The overall takeaway from the analysis of growth in community integration score(s), psychological (and physical), in those experiencing homelessness, is that the CIS Psychological score increases most between intake and 6-months, with no measurable change thereafter. As we do not have a control group of those with treatment (housing) we cannot directly say that supporting permanent housing allows these individuals to feel better in their psychological and physical welfare. We now understand that there are essential factors that contribute to the well beings of homeless communities in New Mexico. From the results of the analysis we suggest social and policy reform to better support permanent housing communities.

That is the implication from the results of this analysis to the policy recommender and practitioners is that among people experiencing homelessness enrolled our program for permanent housing, more attention is needed for those with lower education, little to no access to physical and mental health professionals, no common place for external support from family and friends, and for men more so than for women. To support better facilities and housing, we recommend that there is better access to those experiencing homelessness.

Having better access to health facilities, mental and physical, could influence a more excellent feeling of belonging and integration. Having access to at least one self-help group, like churches, libraries, and areas of the congregation could further influence a feeling of belonging and integration. Finding ways to educate or foster education could result in a higher feeling of community integration. Most importantly, these increases of CIS scores are attributed to the amount of time those experiencing homelessness who were in permanent supportive housing. This is only until 6-months into the program, so after their CIS score growth is level, these individuals may be improving in other vocational and social ways that are more stable.

5.2 Statistical Modeling Conclusion

Statistical analyses are driven by questions, hypotheses, and the extent of the quantity and quality of the data. The analysis presented in this paper used hierarchical mixed models, which allows for technically accurate estimation and modeling of nested data. Using this technique, we could contextually break down the model into sub-models, to explain how and where the variance is explained. For instance, we can ask how variables measured at one level affect relationships occurring at other levels.

A researcher has to make many model specification decisions to draw the best inference from the data. In a hierarchical framework, the following are specifications that one must consider to model the data accurately: selecting appropriate contextual covariates, determining if the ICC is high enough to warrant several appropriate levels, deciding which components are to be modeled as fixed or random (vary across higher levels), and determining the structure to model the variance-covariance components (Bickel 2007). The model validity is more complex in hierarchical modeling, and we relax the independent and constant-variance assumptions, allowing correlated errors with non-constant variance. Allowing relaxed assumptions from the GLM accounts for the nested structure of the data.

Using mixed models allows us to use random effects to account for correlated residuals. The random portion of mixed models allows for individual effects (such as, random variation across individuals). Using mixed-effects allows us to assume that there is more than one source of random variability in the data. To more accurately estimate the covariance parameters, we use REML instead of the more common ML estimator. There are specific ways to model the variance components in different statistical software. As we modeled the data in SAS PROC Mixed, we use the "variance components" method, over say, AR(1), or UN structures. The use of VC is because our data was survey data from selected individuals at non-random sites.

Appendix: Model Diagnostic Figures

Unconditional Model

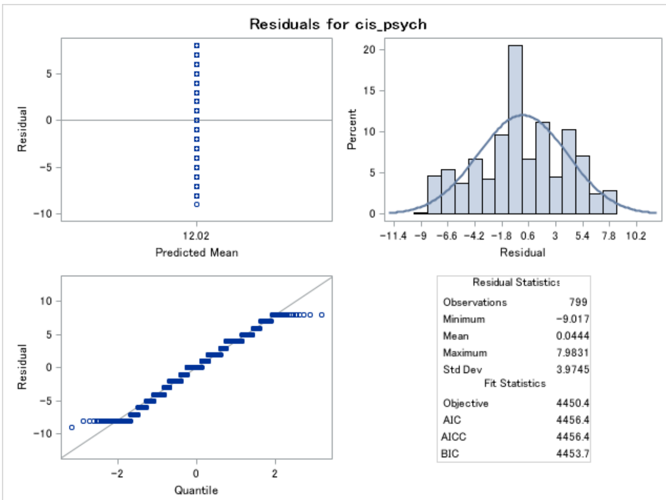

Figure A.1: Residual Distributional Diagnostics for the Unconditional Model-These diagnostics show that the unconditional model does not follow a normal curve; that is, the residuals are not normally distributed with mean zero and equal variance. We see this in the large deviations from the normal blue curve in the histogram and outliers in the quantile plot.

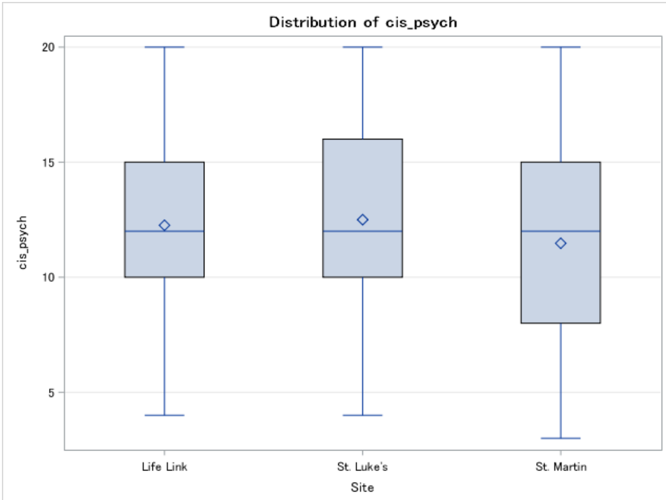

Figure A.2: Box plots of the response, CIS Psychological score, among the three different sites-This diagnostic plot shows that the 3 sites are distributed differently in how the boxes are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same. Although, we would say they exhibit equal variance.

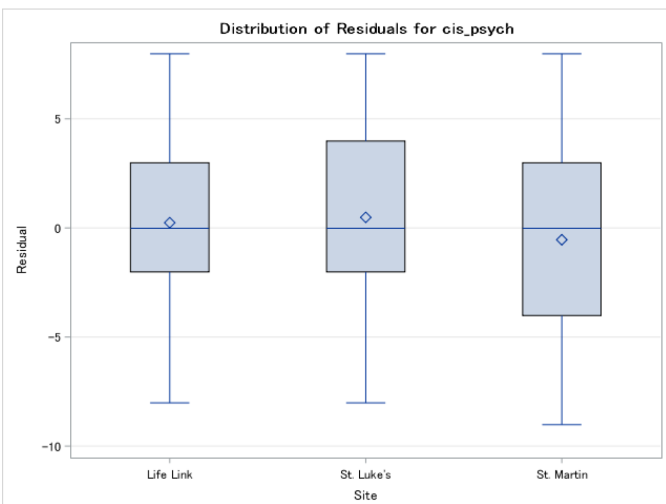

Figure A.3: Box plots of the residuals for the model among the three different sites-This diagnostic plot shows that the 3 sites are distributed differently in the residuals in how the boxes are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same, with equal variance.

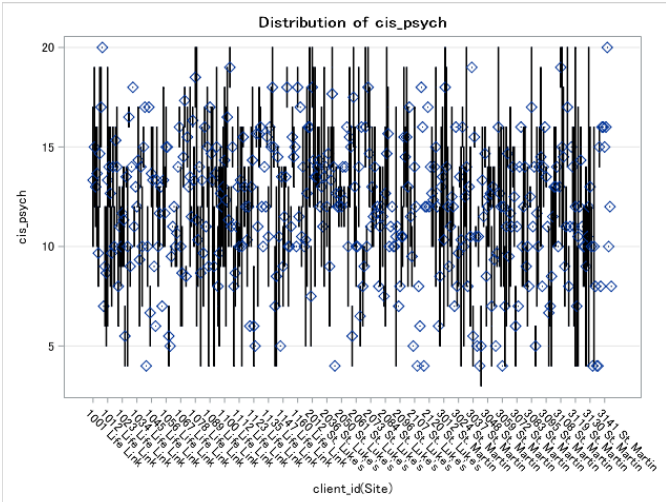

Figure A 4: Time Series plot of the response of clients within sites- This diagnostic plot shows independence among the CIS Psychological score. We see that there is random noise, an indication of independence in the model.

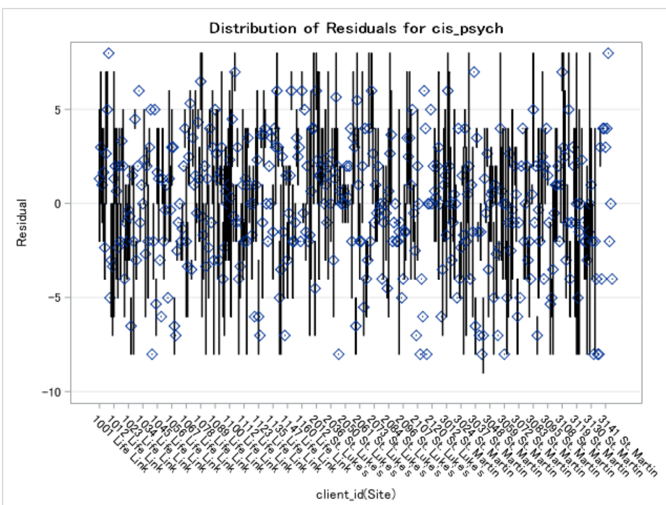

Figure A.5: Time Series plot of the residuals of clients within sites- This diagnostic plot shows independence among the CIS Psychological residuals. We see that there is random noise, an indication of independence in the residuals.

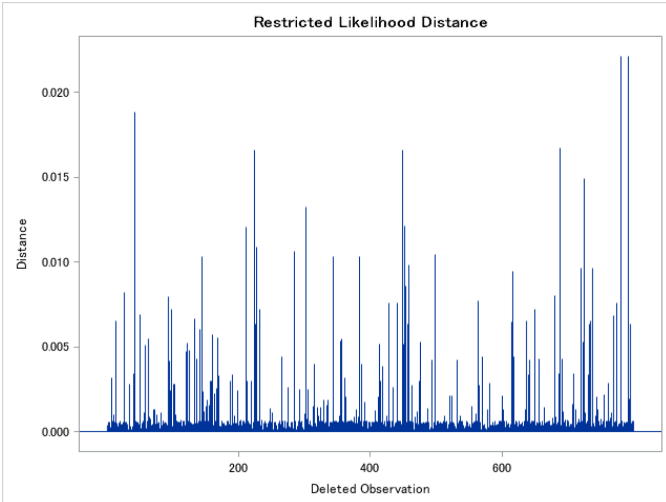

Figure A.6: REML Distance plot of the full data set- This plot is used to diagnose overall influence points. There are no high influence points from the fixed effects.

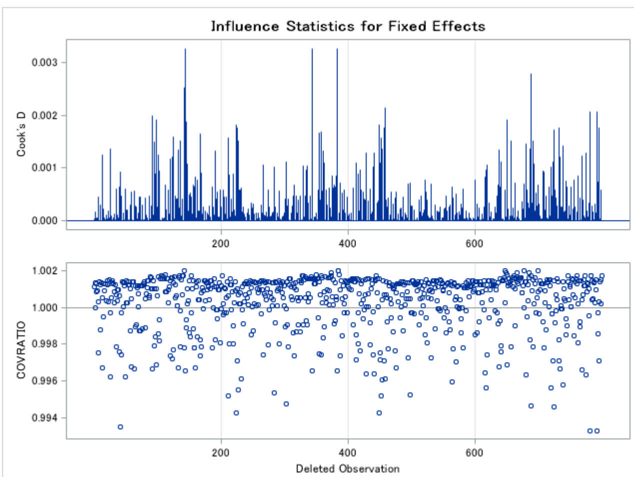

Figure A.7 Cook's Distance For Fixed Effects and Influential points for Random Effects. There are no high influential observations in the fixed or random effects.

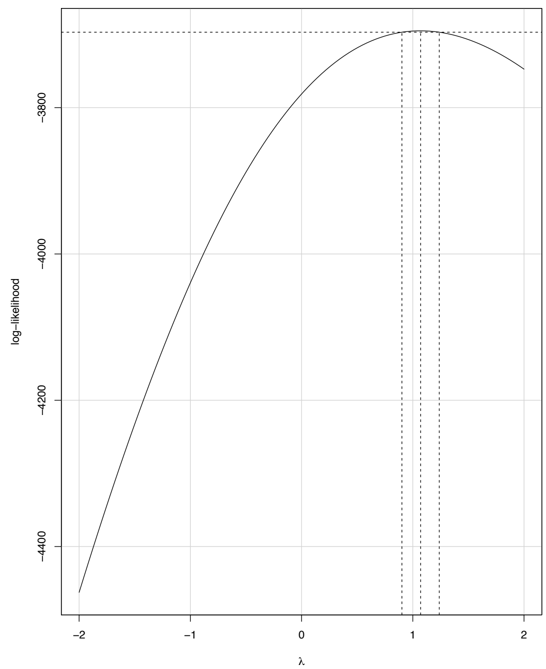

Figure A.8: Box-Cox plot for the the Level-1 Model; all levels of the model including the unconditional through to the full model all have a lambda (power) value of about 1 with narrow bands; this is not the exact plot for each level of the model as they all have different residual values but they all exhibit a similar look with similar axis values. This means no power transformation in the outcome would result in bettering the normality assumption.

Level-1 Model

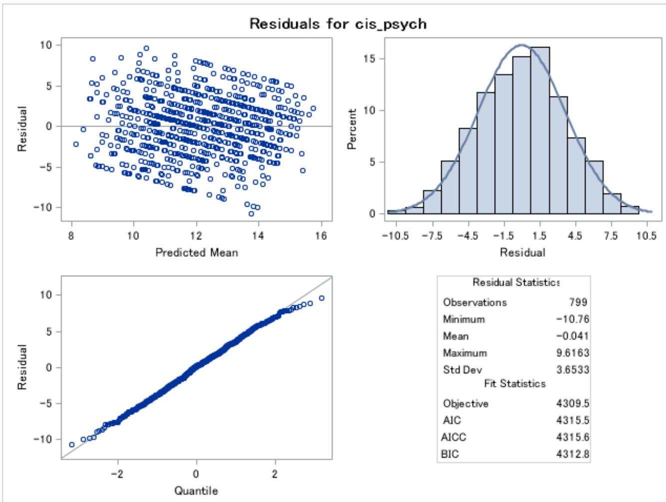

Figure A.9: Residual Distributional Diagnostics for the Unconditional Model- The scatter plot shows independence in the residuals; the histogram shows normality in the distribution of the residuals; the qq-plot looks like there is normality in the residuals, with few off diagonal points.

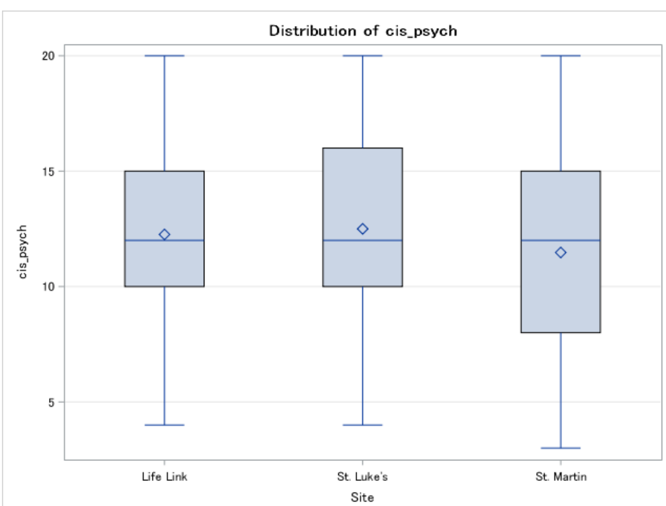

Figure A.10: Box plots of the response, CIS Psychological score, among the three different sites- This diagnostic plot shows that the 3 sites are distributed differently in how the boxes *are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same. These show equal variance.*

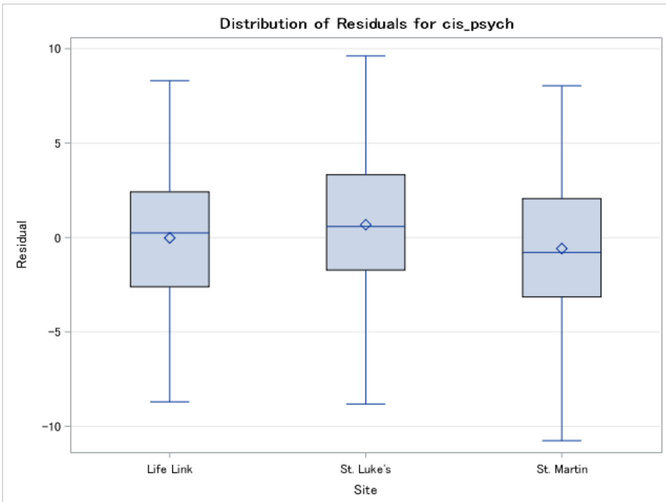

Figure A.11 : Box plots of the residuals for the model among the three different sites- This diagnostic plot shows that the 3 sites are distributed differently in the residuals in how the boxes are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same, with equal variance.

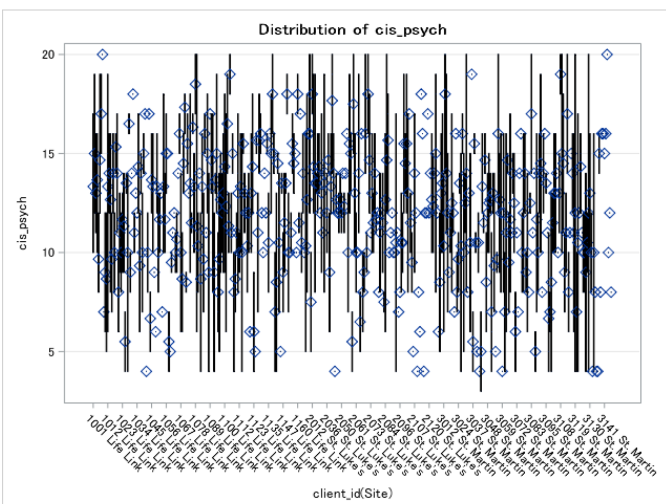

Figure A.12: Time Series plot of the response of clients within sites- there is random noise in this plot of the outcome, meaning that we can assume there is independence in the variable.

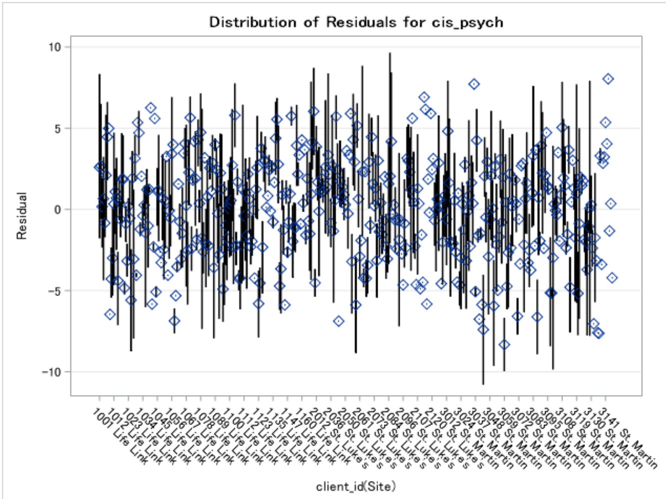

Figure A.13 : Time Series plot of the residuals of clients within sites- there is random noise in this plot of the outcome residuals, meaning that we can assume there is independence in the residuals.

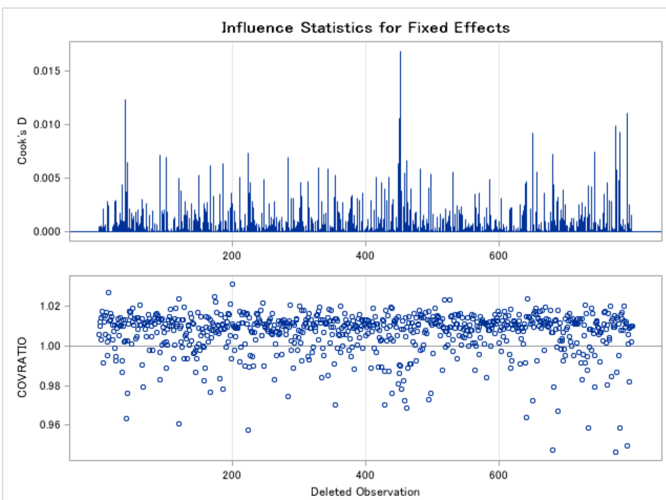

Figure A.14: Cook's Distance for Fixed Effects and Influential points for Random Effects- There are no high influential observations in the fixed or random effects.

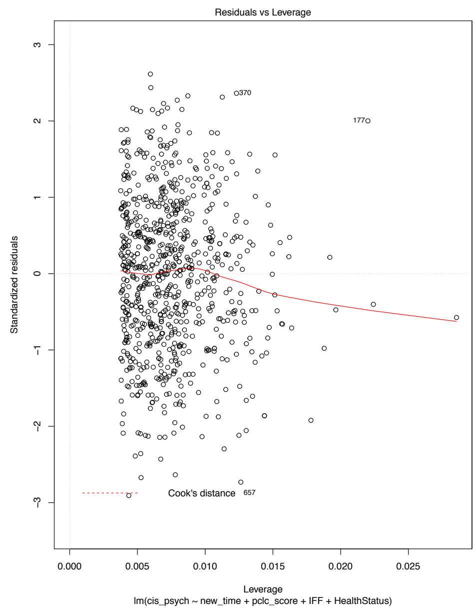

Figure A.15 : Level-1 Residuals vs Leverage points- There are no high leverage points that cause concern for removing of an observation.

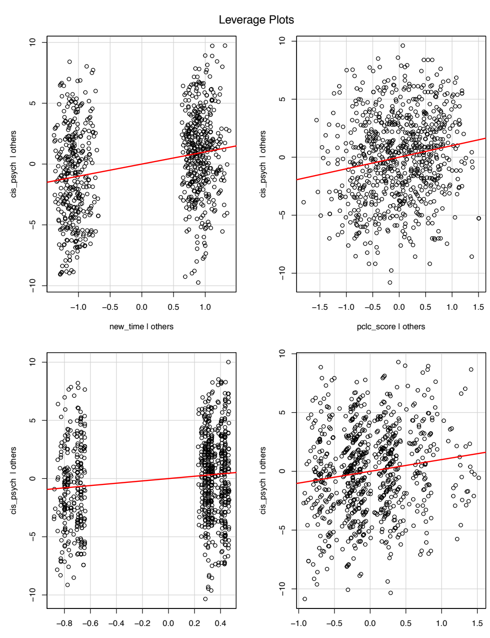

Figure A.16: Leverage plots of the numeric predictors in the Level-1 model- these variables are not needing an transformation or seem to be linearly dependent among the covariates..

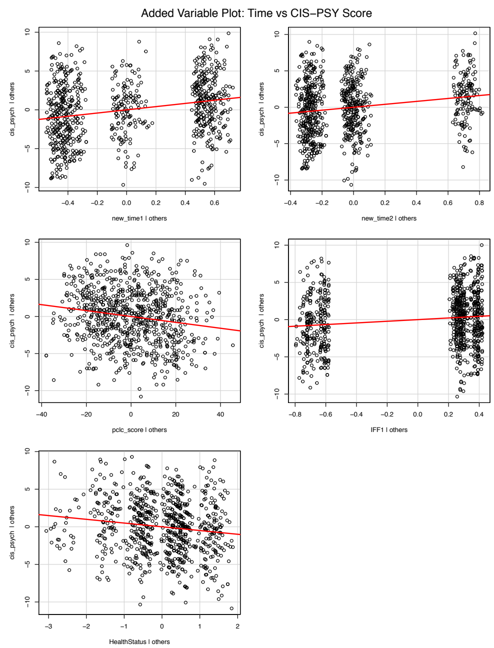

Figure A.17: Added Variable plots for the level-1 predictors- these variables are not needing an transformation or seem to be linearly dependent with the predictor.

Level-2 Model

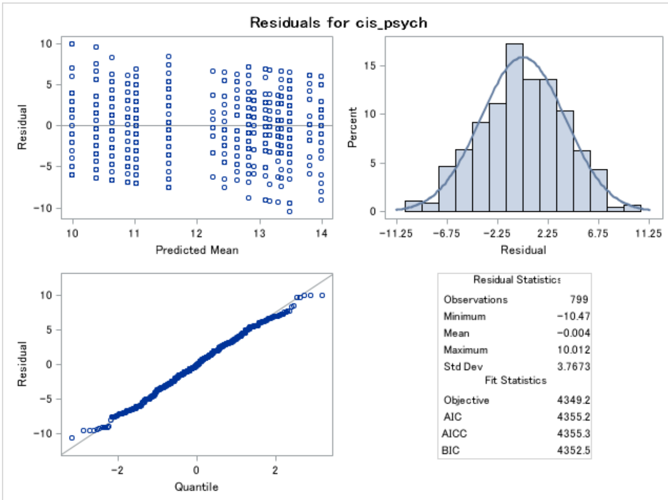

Figure A.18: Residual Distributional Diagnostics for the Level-2 Model- The scatter plot shows independence in the residuals; the histogram shows normality in the distribution of the residuals; the qq-plot looks like there is normality in the residuals, with few off diagonal points. This model exhibits more deviance from normality than the level-1 model.

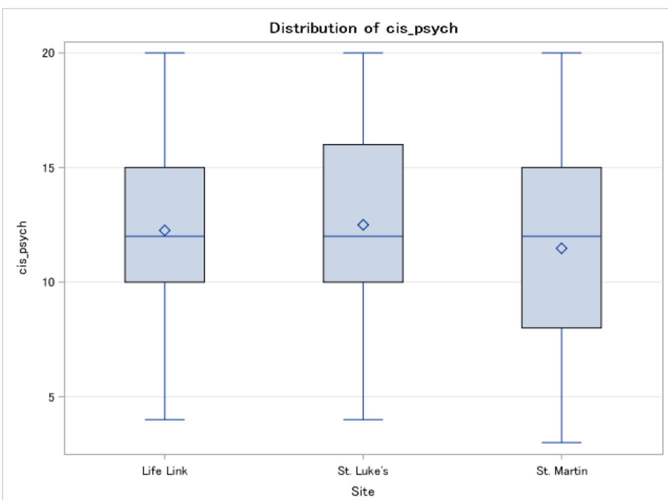

Figure A. 19: Box plots of the response, CIS Psychological score, among the three different sites- This diagnostic plot shows that the 3 sites are distributed differently in how the boxes are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same. These show equal variance.

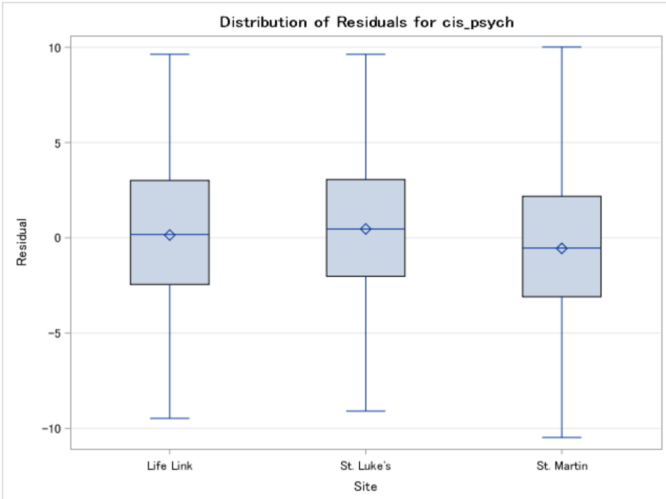

Figure A. 20: Box plots of the residuals for the model among the three different sites- This diagnostic plot shows that the 3 sites are distributed differently in the residuals in how the boxes are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same, with equal variance.

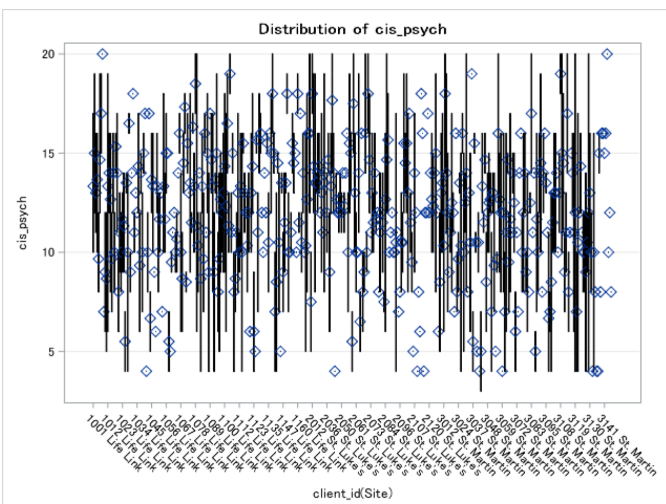

Figure A. 21: Time Series plot of the response of clients within sites- There is random noise in this plot of the outcome, meaning that we can assume there is independence in the variable.

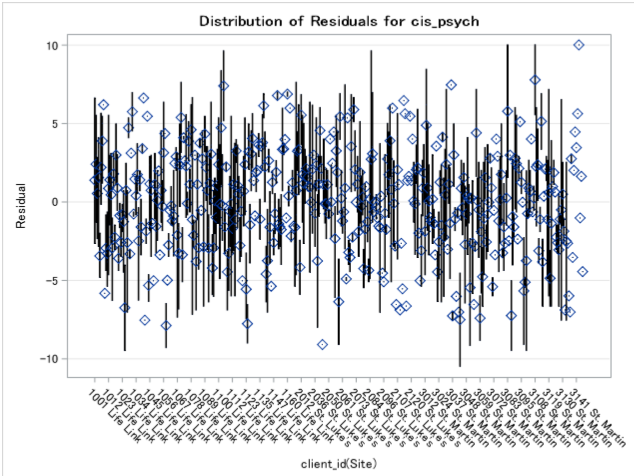

Figure A. 22: Time Series plot of the residuals of clients within sites there is random noise in this plot of the outcome residuals, meaning that we can assume there is independence in the residuals.

Figure A. 23: Cooks Distance For Fixed Effects and Influential points for Random Effects- There are no high influential observations in the fixed or random effects.

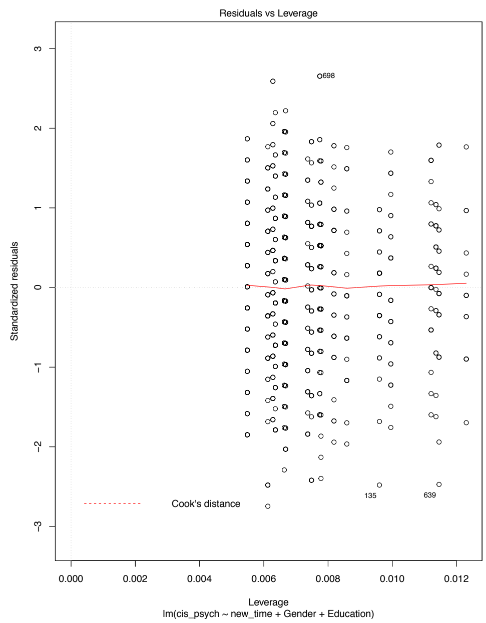

Figure A. 24: Level-2 Residuals vs Leverage points- There are no high leverage points that cause concern for removing of an observation.

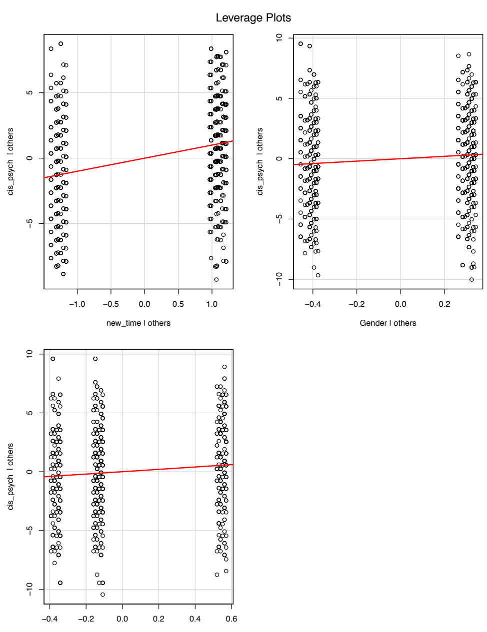

Figure A.25: Level-2 Leverage plots for all the predictors- This plot shows there are no linear dependencies in the variables for this model.

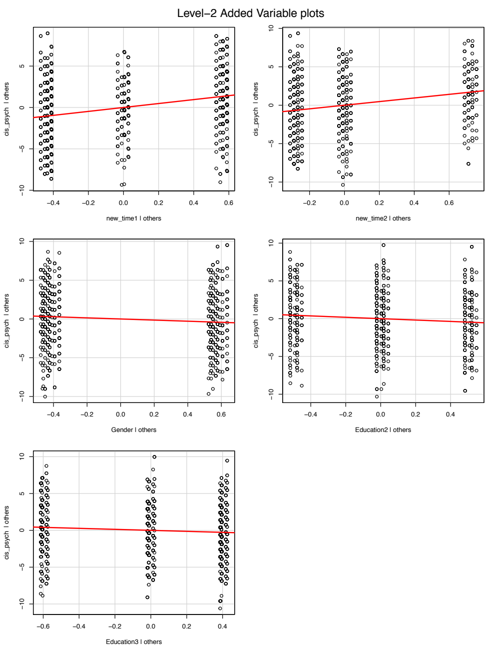

Figure A. 26: Level-2 predictors Added variable plots- This plot shows that there are no linear dependencies among the outcome and covariates.

Level-3

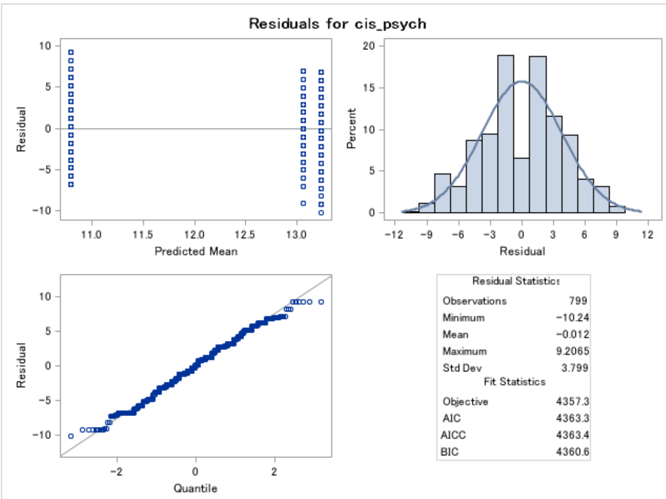

Figure A. 27: Distributional Diagnostics for the Level-3 Model- The scatter plot shows independence in the residuals; the histogram shows deviations from normality in the distribution of the residuals; the qq-plot looks like there is again, deviations from normality in the residuals, with many off diagonal points. This model exhibits more deviance from normality than the level-1 model, and level-2 model.

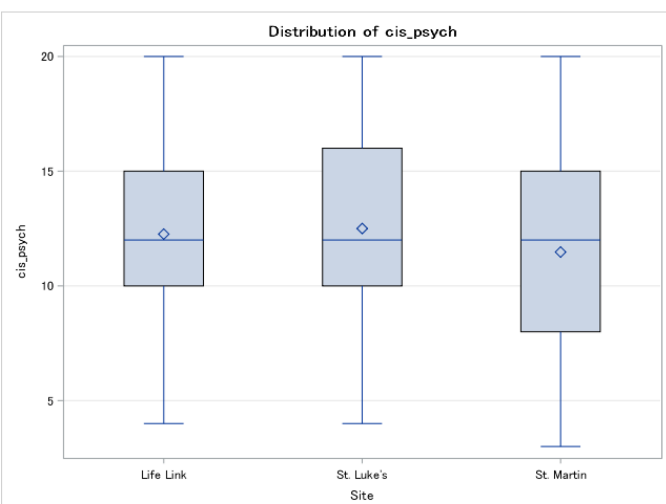

Figure A. 28: Box plots of the response, CIS Psychological score, among the three different sites-This diagnostic plot shows that the 3 sites are distributed differently in how the boxes *are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same. These show equal variance.*

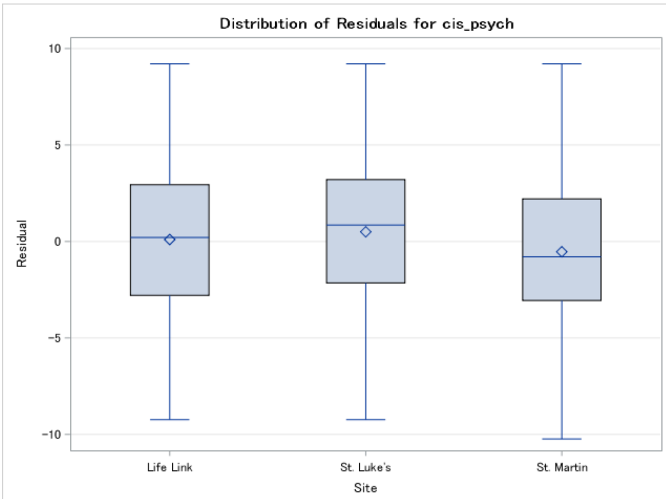

Figure A. 29: Box plots of the residuals for the model among the three different sites- This diagnostic plot shows that the 3 sites are distributed differently in how the boxes are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same. These show equal variance in the residuals.

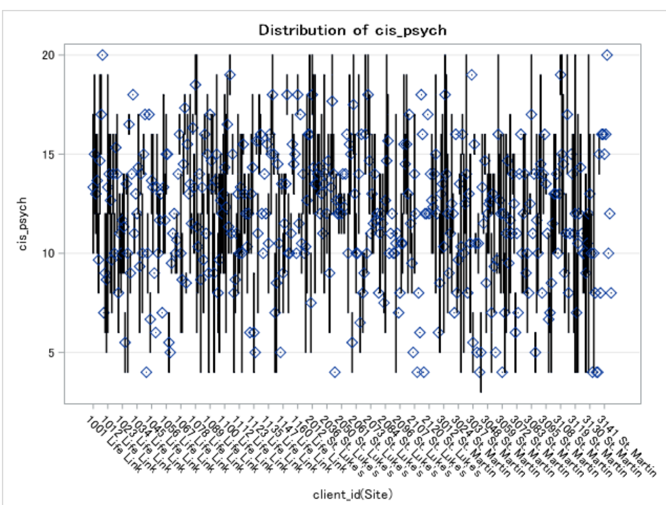

Figure A. 30 : Time Series plot of the response of clients within sites- There is random noise in this plot of the outcome, meaning that we can assume there is independence in the variable.

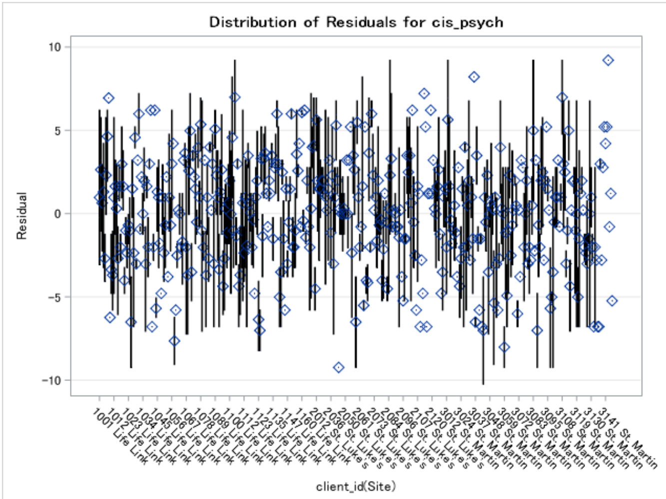

Figure A. 31: Time Series plot of the residuals of clients within sites- There is random noise in this plot of the outcome residuals, meaning that we can assume there is independence in the residuals.

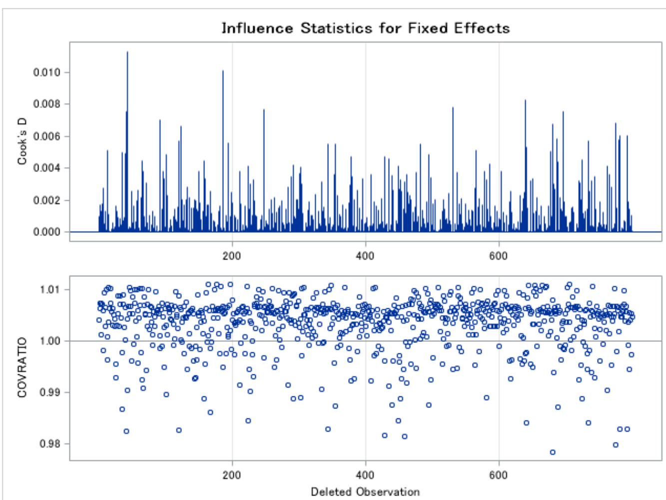

Figure A. 32: Cooks Distance For Fixed Effects and Influential points for Random Effects- There are no high influential observations in the fixed or random effects.

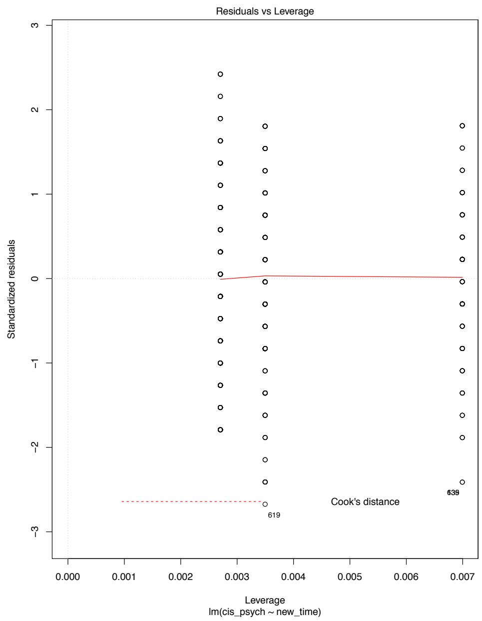

Figure A. 33: Level-3 Residuals vs Leverage points- There are no high leverage points that cause concern for removing of an observation.

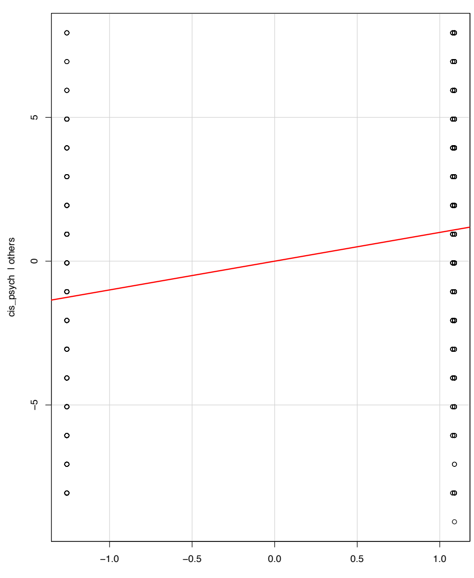

Figure A. 34: Level-3 Leverage plots for all the predictors- This plot shows there are no linear dependencies in the variables for this model.

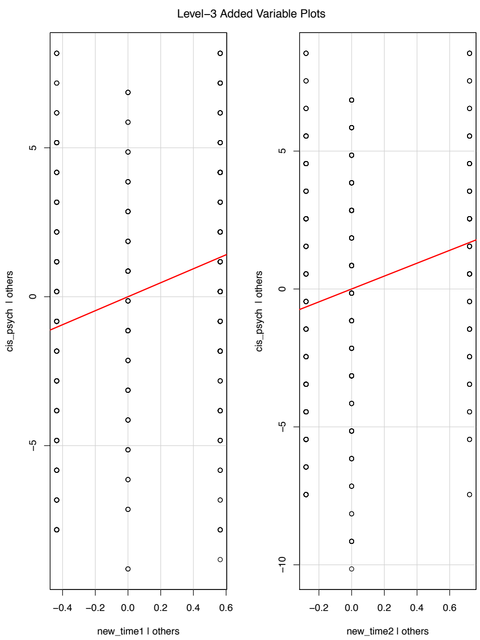

Figure A. 35: Added Variable plots for the level-3 Model- This plot shows that there are no linear dependencies among the outcome and covariates.

Full model

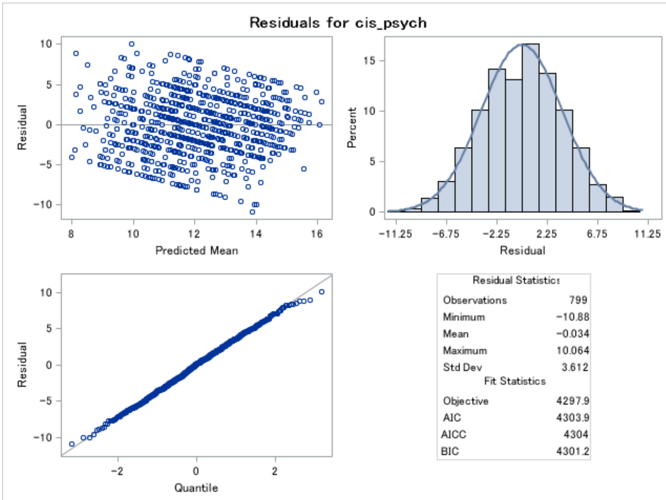

Figure A. 36: Distributional Diagnostics for the Full Model Residuals- This is the model of primary interest where we really care about model fit. We see that this model fits the data well as there is no clear pattern in the scatter plot, normality in the histogram, and the qq-plot has very few off diagonal points.

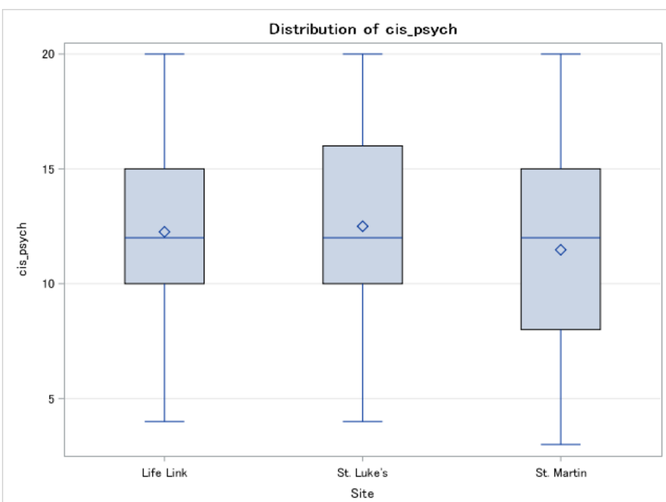

Figure A. 37: Box plots of the response, CIS Psychological score, among the three different sites- This diagnostic plot shows that the 3 sites are distributed differently in how the boxes are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same. These show equal variance.

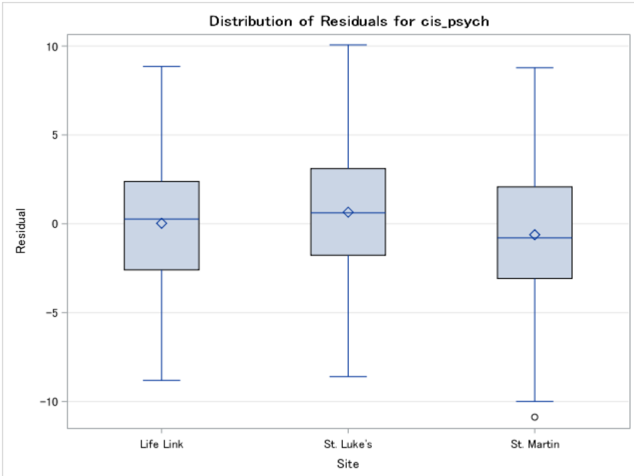

Figure A. 38: Box plots of the residuals for the model among the three different sites- This diagnostic plot shows that the 3 sites are distributed differently in how the boxes are of different sizes, but that the mean CIS Psychological score for all three sites are roughly the same. These show equal variance among the residuals.

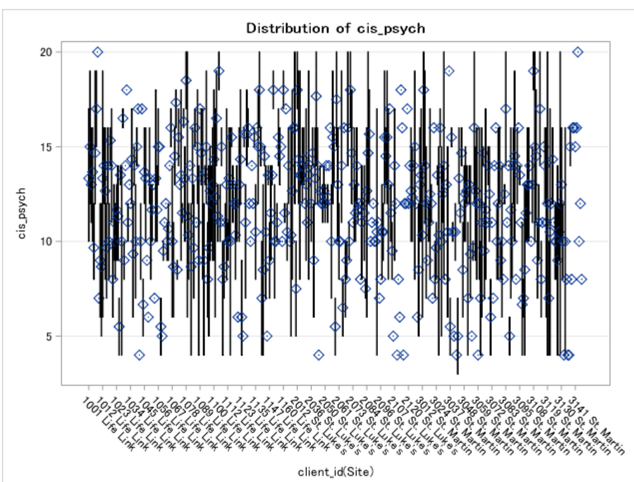

Figure A. 39: Time Series plot of the response of clients within sites- There is random noise in this plot of the outcome, meaning that we can assume there is independence in the variable.

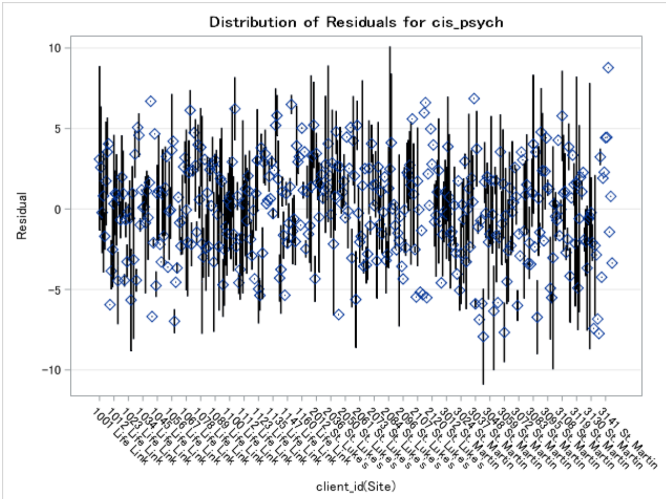

Figure A. 40 : Time Series plot of the residuals of clients within sites- There is random noise in this plot of the outcome residuals, meaning that we can assume there is independence in the residuals.

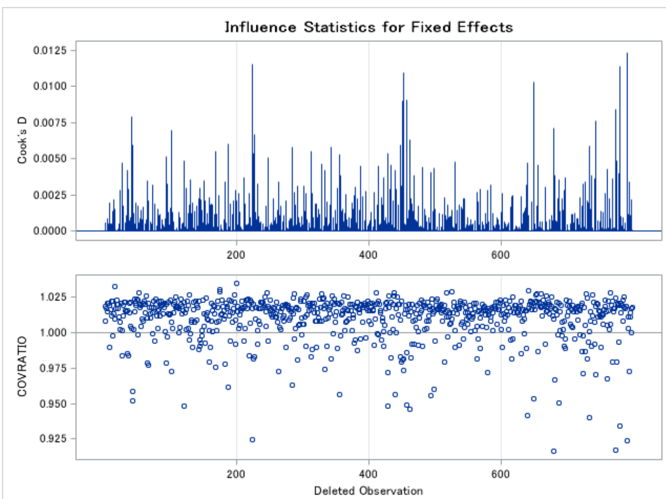

Figure A. 41: Cooks Distance For Fixed Effects and Influential points for Random Effects- There are no high influential observations in the fixed or random effects.

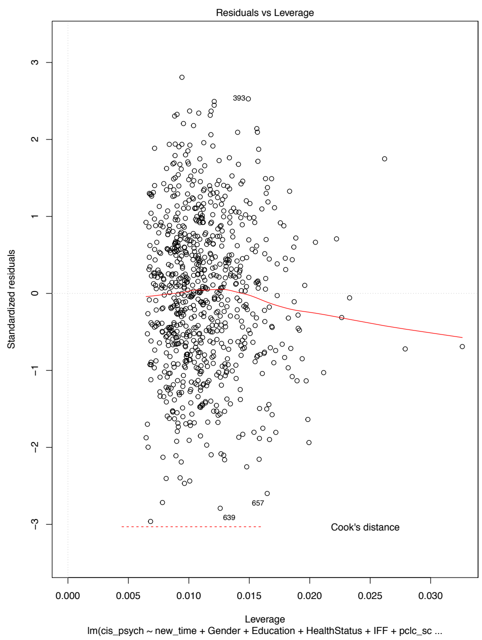

Figure A. 42: Full Model Residuals vs Leverage points- There are no high leverage points that cause concern for removing of an observation.

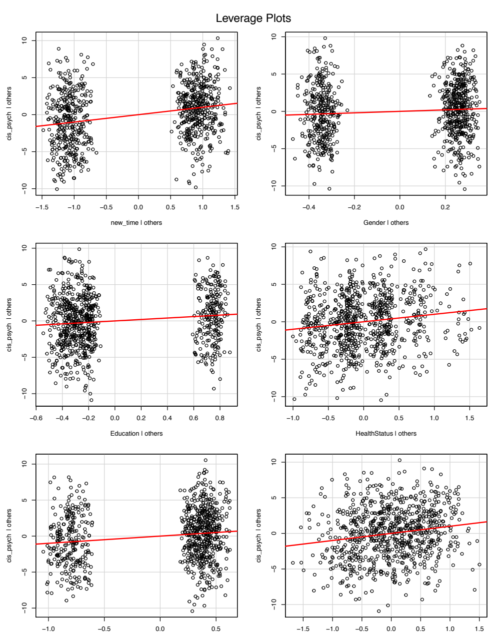

Figure A. 43: Full Model Leverage plots for all the predictors- This plot shows there are no linear dependencies in the variables for this model.

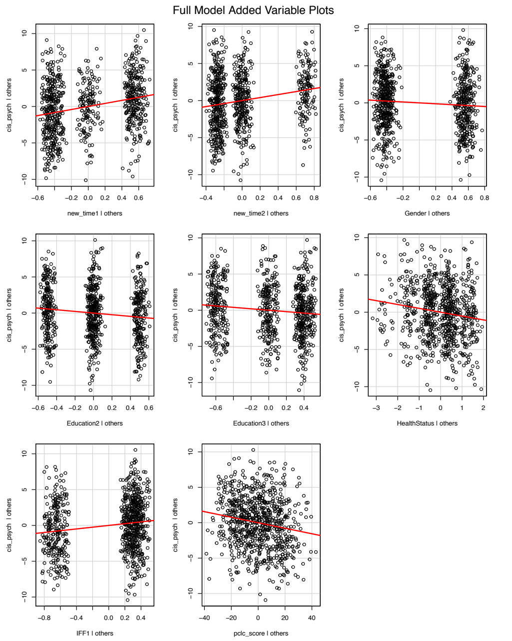

Figure A. 44: Added Variable plots for the Full Model- This plot shows that there are no linear strong dependencies among the outcome and covariates.

References

Austin, Peter C. “A Comparison of the Statistical Power of Different Methods for the Analysis of Cluster Randomization Trials with Binary Outcomes.” *Statistics in Medicine*, vol. 26, no. 19, Aug. 2007, pp. 3550–65. *DOI.org (Crossref)*, doi:10.1002/sim.2813.

Baumgartner, Joy Noel, and Daniel B. Herman. “Community Integration of Formerly Homeless Men and Women With Severe Mental Illness After Hospital Discharge.” *Psychiatric Services*, vol. 63, no. 5, May 2012, pp. 435–37. *DOI.org (Crossref)*, doi:10.1176/appi.ps.201100392.

Beck, Nathaniel. “Of Fixed-Effects and Time-Invariant Variables.” *Political Analysis*, vol. 19, no. 2, 2011, pp. 119–22. *DOI.org (Crossref)*, doi:10.1093/pan/mpr010.

Bell , Bethany, et al. “A Multilevel Model Primer Using SAS® PROC MIXED .” *SAS Global Forum* , 2013, https://support.sas.com/resources/papers/proceedings13/433-2013.pdf.

Bell, Andrew, et al. “Fixed and Random Effects Models: Making an Informed Choice.” *Quality & Quantity*, vol. 53, no. 2, Mar. 2019, pp. 1051–74. *DOI.org (Crossref)*, doi:10.1007/s11135-018-0802-x.

Bickel, Robert. “Multilevel Analysis for Applied Research: It’s Just Regression!” *Curriculum Inquiry*, vol. 38, no. 4, Sept. 2007, pp. 433–34. *DOI.org (Crossref)*, doi:10.1111/j.1467-873X.2008.00418.x.

Bliese, Paul D. “Group Size, ICC Values, and Group-Level Correlations: A Simulation.” *Organizational Research Methods*, vol. 1, no. 4, Oct. 1998, pp. 355–73. *DOI.org (Crossref)*, doi:10.1177/109442819814001.

Bolker, Benjamin M., et al. "Generalized Linear Mixed Models: A Practical Guide for Ecology and Evolution." *Trends in Ecology & Evolution*, vol. 24, no. 3, Mar. 2009, pp. 127–35. *DOI.org (Crossref)*, doi:10.1016/j.tree.2008.10.008.

Bryk, Anthony S., and Stephen W. Raudenbush. "Application of Hierarchical Linear Models to Assessing Change." *Psychological Bulletin*, vol. 101, no. 1, 1987, pp. 147–58. *DOI.org (Crossref)*, doi:10.1037/0033-2909.101.1.147.

Centralized Intake for Helping People Experiencing Homelessness: Overview, Community Profiles, and Resources | The Homeless Hub. https://www.homelesshub.ca/resource/centralized-intake-helping-people-experiencing-homelessness-overview-community-profiles-and. Accessed 2 Oct. 2020.

Christensen, Ronald. *Analysis of Variance, Design, and Regression: Linear Modeling for Unbalanced Data*. Second edition, CRC Press, Taylor & Francis Group, 2016.

Collins, Susan E., et al. "Project-Based Housing First for Chronically Homeless Individuals With Alcohol Problems: Within-Subjects Analyses of 2-Year Alcohol Trajectories." *American Journal of Public Health*, vol. 102, no. 3, Mar. 2012, pp. 511–19. *DOI.org (Crossref)*, doi:10.2105/AJPH.2011.300403.

"Community Integration and Mental Health." *A4CIP*, 18 Dec. 2017, https://communityintegration.com/community-integration-mental-health/.

Crisanti, Annette, et al. "Community Integration Among an Ethnically Diverse Population with Mental Illness and/or Substance Use Disorders in a Permanent Supportive Housing Program." *In Review/Progress*, 2020.

Crisanti, Annette S., et al. "A Longitudinal Analysis of Peer-Delivered Permanent Supportive

Housing: Impact of Housing on Mental and Overall Health in an Ethnically Diverse Population." *Psychological Services*, vol. 14, no. 2, 2017, pp. 141–53. *DOI.org (Crossref)*, doi:10.1037/ser0000135.

De Oliveira, Victor, and Marco A. R. Ferreira. "Maximum Likelihood and Restricted Maximum Likelihood Estimation for a Class of Gaussian Markov Random Fields." *Metrika*, vol. 74, no. 2, Sept. 2011, pp. 167–83. *DOI.org (Crossref)*, doi:10.1007/s00184-009-0295-7.

Douma, Jacob C., and James T. Weedon. "Analysing Continuous Proportions in Ecology and Evolution: A Practical Introduction to Beta and Dirichlet Regression." *Methods in Ecology and Evolution*, edited by David Warton, vol. 10, no. 9, Sept. 2019, pp. 1412–30. *DOI.org (Crossref)*, doi:10.1111/2041-210X.13234.

Duchateau, Luc, et al. *Linear Mixed Models: An Introduction with Applications in Veterinary Research*. International Livestock Research Institute, 1998.

Fleiss, Joseph L. *The Design and Analysis of Clinical Experiments: Fleiss/The Design*. John Wiley & Sons, Inc., 1999. *DOI.org (Crossref)*, doi:10.1002/9781118032923.

Fox, Gordon A., et al., editors. *Ecological Statistics: Contemporary Theory and Application*. Oxford University Press, 2015.

Gardner, Tony, et al. "Centralized Intake for Helping People Experiencing Homelessness: Overview, Community Profiles, and Resources." *Www.Homelesshub.Ca*, Canadian Observatory on Homelessness, 2019, https://www.homelesshub.ca/resource/centralized-intake-helping-people-experiencing-homelessness-overview-community-profiles-and.

Garson, G. David, editor. *Hierarchical Linear Modeling: Guide and Applications*. Sage

Publications, 2013.

Gelman, Andrew. "Multilevel (Hierarchical) Modeling: What It Can and Cannot Do." *Technometrics*, vol. 48, no. 3, Aug. 2006, pp. 432–35. *DOI.org (Crossref)*, doi:10.1198/004017005000000661.

Gelman, Andrew, and Iain Pardoe. "Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models." *Technometrics*, vol. 48, no. 2, May 2006, pp. 241–51. *DOI.org (Crossref)*, doi:10.1198/004017005000000517.

Gormley, Todd A., and David A. Matsa. "Common Errors: How to (and Not to) Control for Unobserved Heterogeneity." *Review of Financial Studies*, vol. 27, no. 2, Feb. 2014, pp. 617–61. *DOI.org (Crossref)*, doi:10.1093/rfs/hht047.

GPRA Measurement Tools. https://www.samhsa.gov/grants/gpra-measurement-tools. Accessed 24 June 2020.

Gueorguieva, Ralitza, and John H. Krystal. "Move Over ANOVA: Progress in Analyzing Repeated-Measures Data AndIts Reflection in Papers Published in the *Archives of General Psychiatry*." *Archives of General Psychiatry*, vol. 61, no. 3, Mar. 2004, p. 310. *DOI.org (Crossref)*, doi:10.1001/archpsyc.61.3.310.

Hand, David J. "Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians by Ronald Christensen, Wesley Johnson, Adam Branscum, Timothy E. Hanson." *International Statistical Review*, vol. 79, no. 2, Aug. 2011, pp. 285–86. *DOI.org (Crossref)*, doi:10.1111/j.1751-5823.2011.00149_12.x.

Harville, David A. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association*, vol. 72, no. 358, June 1977, pp. 320–38. *DOI.org (Crossref)*, doi:10.1080/01621459.1977.10480998.

Hox, J. J. *Multilevel Analysis: Techniques and Applications*. Lawrence Erlbaum Associates, 2002.

Huang, Francis L. "Multilevel Modeling Myths." *School Psychology Quarterly*, vol. 33, no. 3, Sept. 2018, pp. 492–99. *DOI.org (Crossref)*, doi:10.1037/spq0000272.

Larson, Ronald B. "Controlling Social Desirability Bias." *International Journal of Market Research*, vol. 61, no. 5, Sept. 2019, pp. 534–47. *DOI.org (Crossref)*, doi:10.1177/1470785318805305.

LeBreton, James M., and Jenell L. Senter. "Answers to 20 Questions About Interrater Reliability and Interrater Agreement." *Organizational Research Methods*, vol. 11, no. 4, Oct. 2008, pp. 815–52. *DOI.org (Crossref)*, doi:10.1177/1094428106296642.

Lee, Y., and J. A. Nelder. "Hierarchical Generalized Linear Models." *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 4, 1996, pp. 619–56. *Wiley Online Library*, doi:10.1111/j.2517-6161.1996.tb02105.x.

Little, Roderick J. A., and Donald B. Rubin. "The Analysis of Social Science Data with Missing Values." *Sociological Methods & Research*, vol. 18, no. 2–3, Nov. 1989, pp. 292–326. *DOI.org (Crossref)*, doi:10.1177/0049124189018002004.

Maas, Cora, and Tom Snijders. "The Multilevel Approach to Repeated Measures for Complete and Incomplete Data." *Quality and Quantity*, vol. 37, no. 1, 2003, pp. 71–89, doi:10.1023/A:1022545930672.

McNeish, Daniel M., and Laura M. Stapleton. "The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration." *Educational Psychology Review*, vol. 28, no. 2, June 2016, pp. 295–314. *DOI.org (Crossref)*, doi:10.1007/s10648-014-

9287-x.

Mundlak, Yair. “On the Pooling of Time Series and Cross Section Data.” *Econometrica*, vol. 46, no. 1, Jan. 1978, p. 69. *DOI.org (Crossref)*, doi:10.2307/1913646.

Musca, Serban C., et al. “Data with Hierarchical Structure: Impact of Intraclass Correlation and Sample Size on Type-I Error.” *Frontiers in Psychology*, vol. 2, 2011. *DOI.org (Crossref)*, doi:10.3389/fpsyg.2011.00074.

O’Neill, Mick. *The Mathematics of REML*. Universitas Brawijaya, Malang, Indonesia, 2013, https://www.stats.net.au/Maths_REML_manual.pdf.

Patterson, H. D., and R. Thompson. “Recovery of Inter-Block Information When Block Sizes Are Unequal.” *Biometrika*, vol. 58, no. 3, 1971, pp. 545–54. *DOI.org (Crossref)*, doi:10.1093/biomet/58.3.545.

Preacher, Kristopher J., et al. “Computational Tools for Probing Interactions in Multiple Linear Regression, Multilevel Modeling, and Latent Curve Analysis.” *Journal of Educational and Behavioral Statistics*, vol. 31, no. 4, Dec. 2006, pp. 437–48. *DOI.org (Crossref)*, doi:10.3102/10769986031004437.

Raudenbush, Stephen W., and Anthony S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed, Sage Publications, 2002.

Raudenbush, Stephen W., and Wing-Shing Chan. “Application of a Hierarchical Linear Model to the Study of Adolescent Deviance in an Overlapping Cohort Design.” *Journal of Consulting and Clinical Psychology*, vol. 61, no. 6, 1993, pp. 941–51. *DOI.org (Crossref)*, doi:10.1037/0022-006X.61.6.941.

Rutterford, Clare, et al. “Methods for Sample Size Determination in Cluster Randomized Trials.” *International Journal of Epidemiology*, vol. 44, no. 3, June 2015, pp. 1051–

67. *DOI.org (Crossref)*, doi:10.1093/ije/dyv113.

Schafer, J. L. *Analysis of Incomplete Multivariate Data*. 0 ed., Chapman and Hall/CRC, 1997. *DOI.org (Crossref)*, doi:10.1201/9781439821862.

Searle, Shayle R., et al., editors. *Variance Components*. John Wiley & Sons, Inc., 1992. *DOI.org (Crossref)*, doi:10.1002/9780470316856.

Stalter, Lily. "The Compensation for Few Clusters in Clustered Randomized Trials with Binary Outcomes." *Mathematics & Statistics ETDs*, Nov. 2018, https://digitalrepository.unm.edu/math_etds/129.

Umberson, Debra, and Jennifer Karas Montez. "Social Relationships and Health: A Flashpoint for Health Policy." *Journal of Health and Social Behavior*, vol. 51, no. 1_suppl, Mar. 2010, pp. S54–66. *DOI.org (Crossref)*, doi:10.1177/0022146510383501.

Wang, Jichuan, et al. *Multilevel Models: Applications Using SAS*. De Gruyter ; Higher Education Press, 2012.

White, Halbert. "Using Least Squares to Approximate Unknown Regression Functions." *International Economic Review*, vol. 21, no. 1, Feb. 1980, p. 149. *DOI.org (Crossref)*, doi:10.2307/2526245.

Wong, Y.I., Solomon, P.L. *Community Integration of Persons with Psychiatric Disabilities in Supportive Independent Housing: A Conceptual Model and Methodological Considerations.* Ment Health Serv Res **4,** 13–28 (2002). https://doi.org/10.1023/A:1014093008857

Wood, Simon N. "Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models: Estimation of

Semiparametric Generalized Linear Models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 1, Jan. 2011, pp. 3–36. *DOI.org (Crossref)*, doi:10.1111/j.1467-9868.2010.00749.x.

Yanos, Philip. "Beyond 'Landscapes of Despair': The Need for New Research on the Urban Environment, Sprawl, and the Community Integration of Persons with Severe Mental Illness." *Health and Place*, vol. 13, no. 3, Sept. 2007, pp. 672–76.