Statistical Analyses of Longitudinal Observational Data from Adults Living in Marginalized Housing, with Particular Attention to Cognitive Function Characterization

by

Linwan Xu

B.Sc., Simon Fraser University, 2021

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Fall 2023

Declaration of Committee

Name: Linwan Xu

Degree: Master of Science

Thesis title: Statistical Analyses of Longitudinal Observational Data from Adults Living in Marginalized Housing, with Particular Attention to Cognitive Function Characterization

Ethics Statement

[illegible]

Abstract

An ongoing longitudinal study named by Hotel Study focuses on investigating the physical and mental health condition of people who lived in Vancouver's Downtown Eastside (DTES). The study's primary objective is to provide better healthcare services for the target population. Study participants could not follow closely the study's predetermined visit schedule and thus have caused their study data with nonignorable missing. That has motivated this project. We aim to understand the data missing mechanism and then account for it in analyses of the participants' cognitive test results. We begin with a descriptive analysis of all the study's available data. The analysis indicates that the study participants who were recruited from different sources represent different populations. That leads us to focus on the participants from single-room occupancy (SRO) hotels of the DTES in this project. We explore the frequency and rate of participants' study visits and their attendances to the cognitive tests over time by conducting regression analyses under various models. Participants who joined the study early appear to experience less missing visits. We conduct regression analyses of the study's available cognitive test scores with adjustment for the data missing. The test scores appear to be strongly associated with when the participants' joined the study. We also find that the participants' decision-making abilities are associated with their age and sex.

Keywords: cross-sectional analysis; generalized linear model; generalized estimating equation; linear mixed effects model; longitudinal analysis; missing data; visit frequency

Acknowledgements

The completion of this project would not have been possible without the guidance and support of my supervisor, Professor Joan Hu. Thank you for providing me with a USRA opportunity for me to identify my research interests when I was an undergraduate student, and thank you for providing me with a chance to pursue my Master's study. I am really

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Health issues are crucial to be concerned about around the world; it relates to the quality of life of local residents. The long-term homeless or low-income population is often associated

While reviewing the past hotel sub-studies, one problem received my attention. For example, the work of Baitz et al. (2021) measured how participants' decision-making abilities are related to health-risk behavior as well as learning and memory; study samples with invalid data or missing data were excluded from this work, which may lead to biased results and difficulties in interpreting. Some other sub-studies use the same strategies to ignore incomplete data; that is researchers directly assume the missing data is under the missing completely at random assumption. The fact that participants could not follow closely the study's predetermined visit schedules and have caused significant amounts of missing data motivates me to consider this project. Additionally, the study's chosen sample is more akin to a convenience sample than a random sample; particularly when the study sample is expanded, we should consider whether it is suitable to pool all of the study samples for analysis, which is the first question that my project seeks to address. The second question we would like to answer from the data is what are the reasons for participants not following the predetermined visit schedule in the longitudinal, and how should we use the available information to find out the factors that affect participants' cognitive test scores.

In psychology research, longitudinal studies play a role because they keep track of how important factors change over time for an individual; however, missing data could frequently arise. Following Little and Rubin (2019), missing data can be categorized into three main types of missing: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). After the identification of the missing data mechanism, based on Schlomer et al. (2010) as well as Baraldi and Enders (2010), the common methods for dealing with missing data issues in longitudinal studies can be summarized as deletion methods, imputation methods, and maximum likelihood estimation.

1.2 Study Objectives

With the missing data issue in our dataset, we first explore if it is appropriate to use the whole collected sample in our project. Then we focus on the frequency and rate of completing follow-up in the longitudinal study; this will allow us to understand the missing mechanism in this study. Later, we investigate the relationship between participants' test scores and characteristics.

This paper is organized as follows. In Chapter 2, we will introduce general notation as well as some cognitive tests in this project; provide descriptive analyses for the data. In Chapter 3, we will first conduct cross-sectional analyses of missing mechanisms on the count and rate of study visits using Poisson regression; then we will conduct longitudinal data analyses on the frequency of attending cognitive tests in each visit and the probability of visiting

Chapter 2

Descriptive Summary of the Cognitive Test Data from the Hotel Study

The neurocognitive data we obtained from Simon Fraser University Psychology Department's Professor Dr. Allen Thornton was used in this project. Data have been gathered since November 17, 2008, with the most recent update occurring on April 11, 2022. Key cognitive tests evaluate verbal learning and memory (measured by Hopkins Verbal Learning Test), inhibitory control (measured by Stroop Color Word Test), sustained attention and processing speed (measured by Rapid Visual Information Processing), mental flexibility (measured by Intradimensional/Extradimensional), decision-making ability (measured by Iowa Gambling Task) once per study year; premorbid IQ measured by the Wechsler Test of Adult Reading and additional data such as demographic variables, were gathered at the baseline. In this chapter, we will first define the notation for the project, and have a description of the cognitive tests. Following that, we will provide a descriptive analysis of the data.

2.1 Notation

Participants are required to visit the study annually, in each visit, at most 5 cognition tests will be given. We introduce basic notation in this section; let a_i be the calendar date of recruitment for participants i

2022 in the study time scale, and T_i be the last time point for participant i; then $C_i(t) = 0$ for t within the time interval $[0; a_i)$.

Let the covariate vector $X_i = (X_{i1}; :::; X_{i5})$ collect the available explanatory variables for participant i. X_{i1} is the age of participant i when recruited in the study; X_{i2} is the premorbid

September 20, 2011, cannot finish all follow-ups. Participants in other groups are not able to finish follow-ups, while almost all of the participants in the hotel group can.

2.3.2 Demographic Variables

The dataset contains 363 (75%) male participants and 121 (25%) female participants. The average age at recruitment was 41.68 years, with a standard deviation of 11.64 years. The youngest participant in the research was 20 years old while recruiting, and the oldest participant was 75 years old. When we analyze subgroups, based on Table 2.1, participants in the hotel group had a mean age at recruitment of around 44 and a standard deviation of 9.5; those in the hospital group had a mean age at recruitment of around 44 and a standard deviation of 13. The average age at recruitment for the participants in the community group was 39, and the standard deviation was 11. The youth cohort consisted of adult participants under the age of 30.

Participants in the hotel group had a minimum education level of two years and a maximum education level of sixteen years; on average, participants had received 10.5 years of education, with a standard deviation of 2.38. According to Table 2.1, participants in the hotel group, community group, and youth group had comparable distributions of educational levels, while those in the hospital group had relatively higher levels of education.

2.4 Cognitive Tests

2.4.1 Estimate of Premorbid IQ

The score is calculated based on the WTAR test result, age, sex, and education. 13 individuals had missing values, and 22 participants had scores but invalid. Participants with invalid scores may not have understood the test, may not have completed it, or may have had other reasons for the invalid result; the hotel research team reviews the test's validity. The participants' valid scores ranged from 73 to 122, with an average score of 98.44 and a standard deviation of 9.12 for all participants. Table 2.2 shows that participants in the youth group had relatively higher premorbid IQ scores than participants in other groups.

The distributions of recruitment time and demographic variables for participants with premorbid IQ scores missing as well as invalid are shown in Figures 2.3 - 2.6 and Table 2.3. From Figure 2.3, we can see that most of the participants with missing premorbid IQ scores or invalid scores were recruited from SRO hotels early in the study time. Table 2.3 shows more males had missing premorbid IQ scores or invalid scores than females. Figure 2.4 shows participants with age around 50 years old are more likely to have missing premorbid IQ scores or invalid premorbid IQ scores. We cannot see any relationship between participants' education level and missingness as well as invalidity of scores from Figures 2.5 and 2.6.

2.4.2 Verbal Learning and Memory

The available data recorded 12 years of verbal learning and memory scores for the hotel groups, 10 years of scores for the community group, 9 years of scores for the youth group, and 7 years of scores for the hospital group. Based on the available data, we could tabulate the numerical value of verbal learning and memory score change over time for each group in Table 2.4; we also count how many participants attended tests at each of their study years. The longitudinal score is relatively flat at around 20. Youth group participants seem to have higher test scores than other groups.

2.4.3 Inhibitory Control

The inhibitory control tests for the number of ink colours correctly specified when given inconsistent colour words. Tabulated results in Table 2.5 shows that the trend of the longitudinal score is flat; on average, participants get scores around 36. Noted that youth group to 100. Participants in the study gradually performed well with scores increasing over the years. Based on Table 2.8, youth group participants had relatively higher IGT net scores compared to participants in other groups. Participants in the hospital group only attended baseline test.

2.5 Conclusion of Descriptive Analyses

We first introduced basic notation in this chapter; then we conducted descriptive analyses of the Hotel neurocognitive data in terms of participants' recruitment time, demographic variables, and cognitive test scores. Participants from a specific subgroup were recruited during a particular time frame. The distributions of demographic variables or cognitive test scores are not that similar within subgroups. Four subgroups seem not from the same population based on our descriptive analyses. Therefore, we will focus on the original Hotel Study sample in the rest of the project by treating this group of people as a random sample from the target population.

Tables and Figures

Figure 2.1: Distribution of participants recruitment time

Figure 2.2: Distribution of participants' age at recruitment time

Figure 2.3: Distribution of recruitment time for participants with premorbid IQ score missing and premorbid IQ score invalid

Figure 2.4: Distribution of age at recruitment for participants with premorbid IQ score missing and premorbid IQ score invalid

Figure 2.5: Distribution of education level for participants with premorbid IQ score missing

Figure 2.6: Distribution of education level for participants with premorbid IQ score invalid

Figure 2.7: Histogram of education level for Hotel group participants

Figure 2.8: Hotel group participants' premorbid IQ score vs education level in year

Figure 2.9: Hotel group participants' recruitment time vs education level in year

Figure 2.10: Hotel group participants' recruitment time vs premorbid IQ score

Figure 2.11: Hotel group participants' total IGT test attendances vs education level in year

Figure 2.12: Hotel group participants' total IGT test attendances vs premorbid IQ score

Figure 2.13: Hotel group participants' premorbid IQ score vs IGT score

Figure 2.14: Hotel group participants' premorbid IQ score vs age at recruitment

	Hotel (294)			Community (74)			Youth (43)			Hospital (73)		
	Count	Mean	SD	Count	Mean	SD	Count	Mean	SD	Count	Mean	SD
Y0	284	18.57	5.58	72	19.83	6.18	37	25.05	5.52	64	18.86	6.39
Y1	220	19.03	6.14	51	22.16	5.59	23	24.17	6.2	42	18.45	6.24
Y2	143	20.49	6.82	55	20.73	7.23	21	21.71	6.92	36	18	6.61
Y3	140	19.76	6.75	44	19.52	7.6	17	23.06	7.63	28	18.32	8.27
Y4	164	19.48	6.67	42	21.05	6.49	12	24.33	4.52	22	20.41	7.44
Y5	168	19.96	7.03	43	19.84	6.41	8	18	8.04	9	18.67	8.75
Y6	139	20.08	6.48	35	20.23	6.28	6	24.5	10.29	1	26	
Y7	144	20.47	6.15	31	21.68	5.4	3	19.33	10.97	0		
Y8	118	20.8	6.81	25	19.96	7.38	1	30		0		
Y9	102	19.61	6.14	14	18.93	6.08	0			0		
Y10	83	19.34	6.26	0			0			0		
Y11	61	18.75	6.49	0			0			0		

Count in tables is the number of participants who had test scores at their visit year j

The means and standard deviations (SD) were calculated based onns

	Hotel (294)			Community (74)			Youth (43)			Hospital (73)		
	Count	Mean	SD	Count	Mean	SD	Count	Mean	SD	Count	Mean	SD
Y0	272	34.95	10.32	71	36.1	10.34	41	44.71	11.96	61	33.98	12.31
Y1	208	35.5	11.04	52	38.35	9.96	25	41.52	12.07	37	32.32	9.87
Y2	141	35.93	11.15	54	38	10.82	21	41.29	11.97	39	31.21	12.68
Y3	131	36.08	10.72	45	38.73	11.64	19	43.32	12.77	28	30.32	13.49
Y4	158	37.08	11.95	41	39.63	10.97	12	42.25	14.78	21	31.38	9.5
Y5	158	37.46	11.4	42	38.5	10.08	9	44.11	14.74	9	33.44	11.66
Y6	131	38.32	11.83	37	37.27	10.95	6	38.67	13.6	1	41	
Y7	136	37.82	11.63	32	37.69	12.05	4	37.75	16.17	0		
Y8	113	35.96	12.04	25	31.28	13.26	1	53		0		
Y9	102	35.24	10.37	14	34.14	11.28	0			0		
Y10	81	34.42	9.89	0			0			0		
Y11	61	33.02	10.64	0			0			0		

Count in tables is the number of participants who had test scores at their visit year j

The means and standard deviations (SD) were calculated based on the available number of participants

Table 2.5: Summary statistics of inhibitory control test

	Hotel (294)			Community (74)			Youth (43)			Hospital (73)		
	Count	Mean	SD	Count	Mean	SD	Count	Mean	SD	Count	Mean	SD
Y0	257	-1.2	1.3	54	-1.05	1.03	27	-1.26	1.33	46	-1.46	1.46
Y1	188	-1.06	1.31	44	-0.77	1.18	15	-1.03	1.3	0		
Y2	126	-0.88	1.4	38	-0.4	1.12	10	-0.37	1.26	0		
Y3	112	-1.08	1.49	23	-0.51	1.17	1	1.6		0		
Y4	125	-0.82	1.37	23	-0.72	1.21	0			0		
Y5	130	-0.69	1.33	22	-0.55	1.15	0			0		
Y6	90	-0.75	1.22	8	-0.34	0.98	0			0		
Y7	110	-0.78	1.27	19	-0.63	1.63	0			0		
Y8	78	-0.61	1.29	1	-0.29		0			0		
Y9	48	-0.65	1.06	0			0			0		
Y10	36	-0.46	1.29	0			0			0		
Y11	14	-0.21	1.27	0			0			0		

Count in tables is the number of participants who had test scores at their visit year j

The means and standard deviations (SD) were calculated based on the available number of participants

Table 2.6: Summary statistics of sustained attention test

	Hotel (294)			Community (74)			Youth (43)			Hospital (73)		
	Count	Mean	SD	Count	Mean	SD	Count	Mean	SD	Count	Mean	SD
Y0	259	-5.95	32.41	48	-2.38	29.43	36	4.17	27.46	47	-2.72	30.19
Y1	171	-5.32	36.66	46	-0.61	39.8	14	8.14	36.91	0		
Y2	105	0.63	35.26	37	3.95	36.92	12	10.33	40.82	0		
Y3	101	2.2	32.65	31	6.97	30.97	1	8		0		
Y4	141	-5.52	35.06	25	5.12	37.87	0			0		
Y5	142	2.65	37.03	19	16.63	32.24	0			0		
Y6	109	-0.04	38.81	7	-6	26.1	0			0		
Y7	98	-1.71	40.42	14	24.43	27.63	0			0		
Y8	70	5.74	43.32	1	50		0			0		
Y9	36	-3.44	39.87	0			0			0		
Y10	24	19	41.46	0			0			0		
Y11	12	4	47.25	0			0			0		

Count in tables is the number of participants who had test scores at their visit year j

The means and standard deviations (SD) were calculated based on the available number of participants

Table 2.8: Summary statistics of decision-making ability test

Education (year)		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Y0	IGT visit	1	1	3	3	5	13	25	33	55	39	43	20	7	0	11
	Not visit	0	1	0	0	4	1	3	3	11	6	2	1	0	1	2
Y1	IGT visit	0	0	1	0	4	7	16	23	40	30	27	12	3	0	8
	Not visit	1	2	2	3	5	7	12	13	26	15	18	9	4	1	5
Y2	IGT visit	0	1	0	1	1	5	13	16	16	17	17	10	3	0	5
	Not visit	1	1	3	2	8	9	15	20	50	28	28	11	4	1	8
Y3	IGT visit	0	0	1	1	3	7	8	14	22	12	13	10	4	1	5
	Not visit	1	2	2	2	6	7	20	22	44	33	32	11	3	0	8
Y4	IGT visit	0	1	1	1	3	7	13	23	35	22	21	7	4	0	3
	Not visit	1	1	2	2	6	7	15	13	31	23	24	14	3	1	10
Y5	IGT visit	0	2	1	1	3	7	17	19	38	19	17	11	4	0	3
	Not visit	1	0	2	2	6	6	11	17	28	26	28	10	3	1	8
Y6	IGT visit	0	1	1	1	2	5	8	12	34	15	13	10	3	0	3
	Not visit	1	1	2	2	7	8	20	24	32	30	31	11	4	1	8
Y7	IGT visit	0	1	1	1	3	7	10	12	29	14	11	4	3	0	2
	Not visit	1	1	2	2	6	6	18	24	36	31	33	17	4	1	9
Y8	IGT visit	0	0	0	1	0	4	8	9	22	10	5	3	1	0	1
	Not visit	1	2	3	2	8	6	20	27	42	32	36	17	5	1	8
Y9	IGT visit	0	1	0	1	1	2	4	4	7	5	4	2	0	0	1
	Not visit	1	1	2	0	4	6	20	21	45	28	26	11	3	0	6
Y10	IGT visit	0	0	0	0	0	2	2	1	6	4	1	1	0	0	0
	Not visit	1	1	1	0	2	5	9	15	28	17	13	7	1	0	4

Table 2.9: Number of hotel participants with different education levels attend the IGT test every year

Chapter 3

Patterns of Study Visits of the Hotel Study Participants

Investigating the missing mechanism may aid in our comprehension of the missingness in the longitudinal study and point us in the right direction for conducting an analysis of the missing data. This chapter aims to explain how covariates for hotel group participants relate to the missingness of the study visit and test visits. We first consider individual time scale on the count and rate of study visits using the Poisson regression model. Then we explore the missing pattern in longitudinal study visits as well as tests attending using the Logistic regression model. Finally, we consider the count and rate of the study visit in the study time scale using the Logistic regression model and Poisson regression model.

3.1 Cross-sectional Analysis of Total Count and Rate of Study Visit

We conduct analyses of participants' total number and rate of visits in this section to explore any factors that could affect overall visits for participants in this longitudinal study.

Model

For participant $i = 1, \ldots, 294$ in the hotel group, $V_{i:}$ is the total number of visits made by participant i in this longitudinal study; n_i is the ideal number of visits that can be given by participant i based on participant i's recruitment time and the last updated time of the dataset. We consider the following Poisson regression models for participant i's total number and rate of visits:

$$\log[E(V_{i:}|X_i)] = {}^{s}_{0} + {}^{s}_{1}X_{i1} + {}^{s}_{2}X_{i2} + {}^{s}_{3}X_{i3} + {}^{s}_{4}X_{i4} + {}^{s}_{5}X_{i5}; \qquad (3.1a)$$

$$\log[E(\frac{V_{i:}}{n_i} jX_i)] = {}^{s}_{0} + {}^{s}_{1}X_{i1} + {}^{s}_{2}X_{i2} + {}^{s}_{3}X_{i3} + {}^{s}_{4}X_{i4} + {}^{s}_{5}X_{i5}; \qquad (3.1b)$$

$$\log[E(V_{i:} jX_i; n_i)] = {}^{s}_{0} + {}^{s}_{1}X_{i1} + {}^{s}_{2}X_{i2} + {}^{s}_{3}X_{i3} + {}^{s}_{4}X_{i4} + {}^{s}_{5}X_{i5} + {}^{s}_{6}\log(n_i); \qquad (3.1c)$$

where covariate vector X_i is defined in the previous chapter as for participant i. X_{i1} is the age at recruitment, X_{i2} is the premorbid IQ score at recruitment, X_{i3} is the education level, X_{i4} is sex, and X_{i5} is the time of recruitment; 30 participants who did not have premorbid IQ score information and will be excluded when modelling. $\frac{V_{i:}}{n_i}$ is the rate of visit made by participant i in this longitudinal study adjusting for the ideal number of visits. Models (3.1a) and (3.1b) are the reduced models of the model (3.1c) if the ${}^{s}_{6} = 0$ or ${}^{s}_{6} = 1$ in the model (3.1c), respectively.

Analysis results

analyzed. Third, for each of the cognitive tests, we consider the participants' probability of attending.

3.2.1 Longitudinal Hotel Study Visits

Model

We consider following logistic models at each individual year to model the relationship between participant i's probability of visiting the study at each year and the covariate vector:

$$\text{logit}[P(V_{ij} = 1 j X_i)] = {}^{s}_{0j} + {}^{s}_{1j} X_{i1} + {}^{s}_{2j} X_{i2} + {}^{s}_{3j} X_{i3} + {}^{s}_{4j} X_{i4} + {}^{s}_{5j} X_{i5}; \qquad (3.2a)$$

$$\text{logit}[P(V_{ij} = 1 j X_i)] = {}^{s}_{0j} + {}^{s}_{1} X_{i1} + {}^{s}_{2} X_{i2} + {}^{s}_{3} X_{i3} + {}^{s}_{4} X_{i4} + {}^{s}_{5} X_{i5}; \qquad (3.2b)$$

$$\text{logit}[P(V_{ij} = 1 j X_i)] = {}^{s}_{0} + {}^{s}_{1} X_{i1} + {}^{s}_{2} X_{i2} + {}^{s}_{3} X_{i3} + {}^{s}_{4} X_{i4} + {}^{s}_{5} X_{i5}; \qquad (3.2c)$$

here time point j takes values from year1 to 11 since all participants visit the study at

3.2.2 Longitudinal Rate of Cognitive Tests Attending

Model

Participant i attended total $R_{ij:}$ cognitive tests at his or her the j^{th} year of visit, and the response variable we considered is $p_j(X_i)$ the probability of attending $R_{ij:}$ tests out of K_{ij} total available tests for participant i at year j; $R_{ij:}\, j K_{ij}\, ; X_i \quad B(K_{ij}\, ; p_j(X_i))$. We consider using logistic regression models to estimate $p_j(X_i)$:

$$\text{logit}[p$$

cognitive tests and covariate vector. However, all estimates of this model are close to 0; the reason for this result could be that we do not have enough possible values for the response variable in this model.

3.2.3 Longitudinal Cognitive Tests Attending

Model

In this section, we consider exploring the probability of test attending for each of the cognition tests separately at each year using Logistic regression. For $k = 1, \ldots, 5$ and $j = 0, \ldots, 11$, $R_{ijk} = 1$ if participant i attended the k^{th} cognitive test at the j^{th} year; the models are:

$$\text{logit}[P(R_{ijk} = 1|X_i)] = {}^{t}_{0jk} + {}^{t}_{1jk}X_{i1} + {}^{t}_{2jk}X_{i2} + {}^{t}_{3jk}X_{i3} + {}^{t}_{4jk}X_{i4} + {}^{t}_{5jk}X_{i5}; \quad (3.4a)$$

$$\text{logit}[P(R_{ijk} = 1|X_i)] = {}^{t}_{0jk} + {}^{t}_{1k}X_{i1} + {}^{t}_{2k}X_{i2} + {}^{t}_{3k}X_{i3} + {}^{t}_{4k}X_{i4} + {}^{t}_{5k}X_{i5}; \quad (3.4b)$$

$$\text{logit}[P(R_{ijk} = 1|X_i)] = {}^{t}_{0k} + {}^{t}_{1k}X_{i1} + {}^{t}_{2k}X_{i2} + {}^{t}_{3k}X_{i3} + {}^{t}_{4k}X_{i4} + {}^{t}_{5k}X_{i5}; \quad (3.4c)$$

where X_{i1} is the age at recruitment, X_{i2} is the premorbid IQ score at recruitment, X_{i3} is the education level, X_{i4} is sex, and X_{i5} is the recruitment time of participant i.

Model (3.4a) is a mixed-effect model with all effects varying every year; for $\forall j$ such that $(\hat{}^{t}_{0jk}, \ldots, \hat{}^{t}_{5jk})^T \quad MN(({}^{t}_{0k}, \ldots, {}^{t}_{5k})^T;)$. Model (3.4b) is a reduced model of model (3.4a) with only intercept varying every year; for $\forall j$ such that $\hat{}^{t}_{0jk} \quad N({}^{t}_{0k}; {}^{2}_{0k})$. Model (3.4c) is a reduced model of model (3.4b) that all effects are fixed at different year.

Analysis results

Missing of verbal learning and memory test ($k = 1$)

Based on Table 3.5, all models suggest that the probability of attending the verbal learning and memory test is significantly associated with the recruitment time; participants who joined the study earlier are more likely to attend this test. In model (3.4b), female participants are more likely to visit this test compared to male participants with odds ratio $\exp(0.1962) = 1.2168$.

Missing of inhibitory control test ($k = 2$)

Based on Table 3.6 in models 3.4b and 3.4c, the estimates of effects on premorbid IQ score, sex, and recruitment time are significantly different from 0 for all models. Participants with higher premorbid IQ scores are more likely to attend this test; female participants are more likely to attend the test.

first consider Logistic regression models to estimate $p_t(X_i)$ at each study year t such that $C_i(t)|n_i(t); X_i \sim B(n_i(t); p_t(X_i))$:

$$\text{logit}[p_t(X_i)] = \beta_{0t} + \beta_{1t}X_{i1} + \beta_{2t}X_{i2} + \beta_{3t}X_{i3} + \beta_{4t}X_{i4} + \beta_{5t}X_{i5}; \tag{3.5.1a}$$

$$\text{logit}[p_t(X_i)] = \beta_{0t} + \beta_{1}X_{i1} + \beta_{2}X_{i2} + \beta_{3}X_{i3} + \beta_{4}X_{i4} + \beta_{5}X_{i5}; \tag{3.5.1b}$$

$$\text{logit}[p_t(X_i)] = \beta_{0} + \beta_{1}X_{i1} + \beta_{2}X_{i2} + \beta_{3}X_{i3} + \beta_{4}X_{i4} + \beta_{5}X_{i5}: \tag{3.5.1c}$$

The mixed-effect model (3.5.1a) is the full model with all the effects varying at each study year; for $\forall t$ such that $(\hat{\beta}_{0t}; \ldots; \hat{\beta}_{5t})^T \sim MN((\beta_0; \ldots; \beta_5)^T; \Sigma)$. The mixed-effect model (3.5.1b) is the reduced model of model (3.5.1a) with only intercept term varying at each study year; $\hat{\beta}_{0t} \sim N(\beta_0; \sigma_0^2)$. The model (3.5.1c) is a reduced model of the model (3.5.1b) with all effects fixed.

Based on the strategy of the model specification for model (3.1c), we could also consider scores are more likely to visit the study. Models (3.5.1a) and (3.5.1b) agree that participants with later recruitment are more likely to visit the study. Compared to the conclusion in section 3.2.1 while considering the individual time scale, we obtain an inverse conclusion for participants' recruitment time. Based on Figure 3.19, this may be due to the reason

Tables and Figures

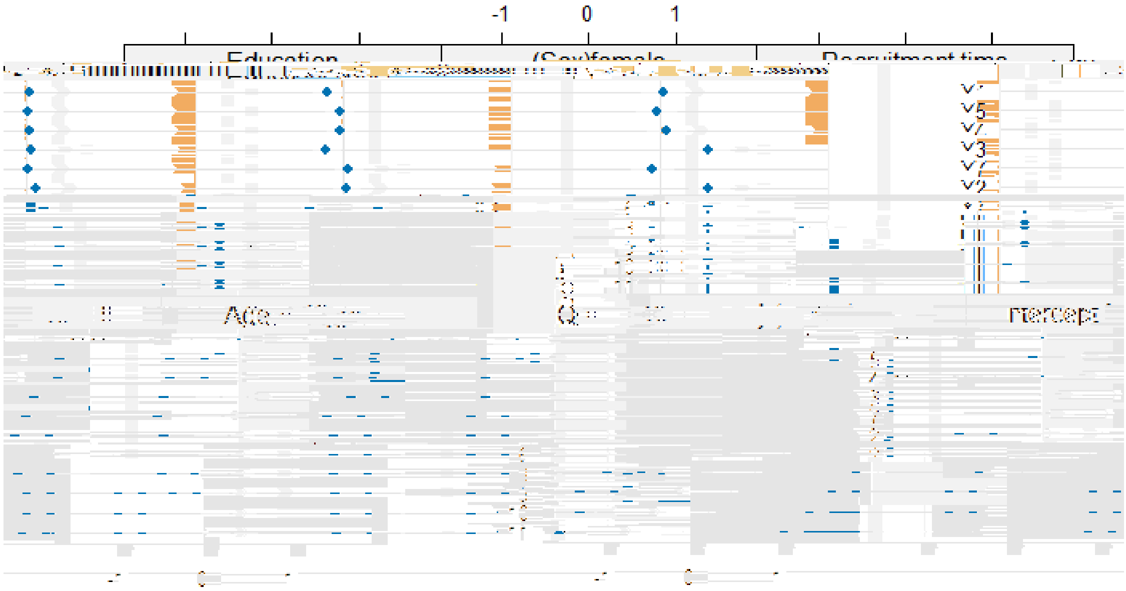

Figure 3.1: Estimate of random effects of the model (3.2a)

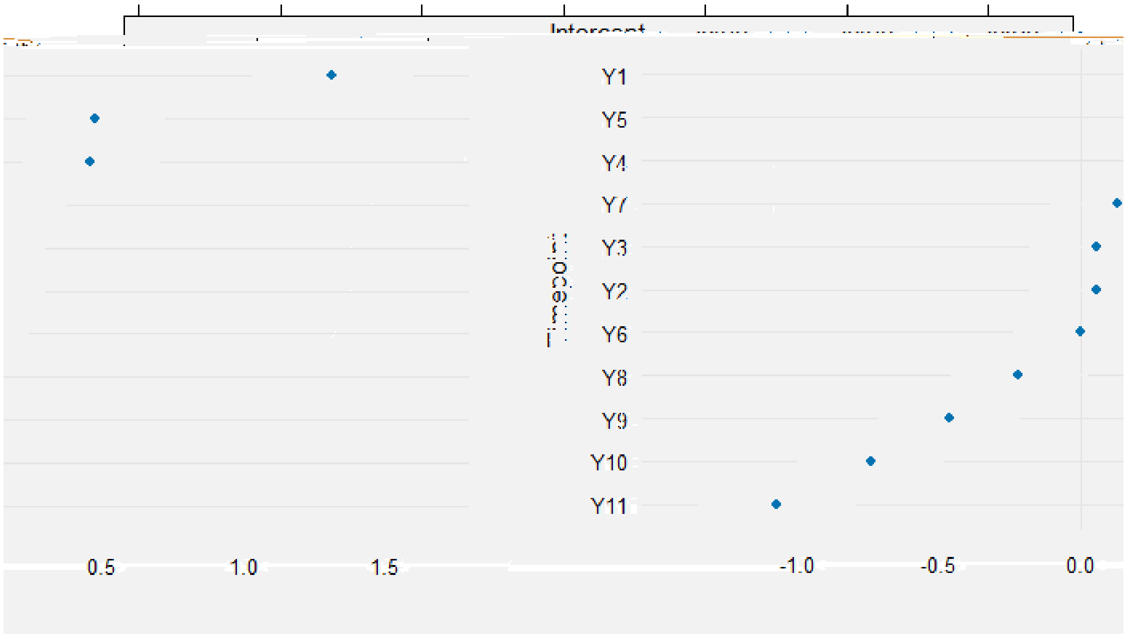

Figure 3.2: Estimate of random effect of the model (3.2b)

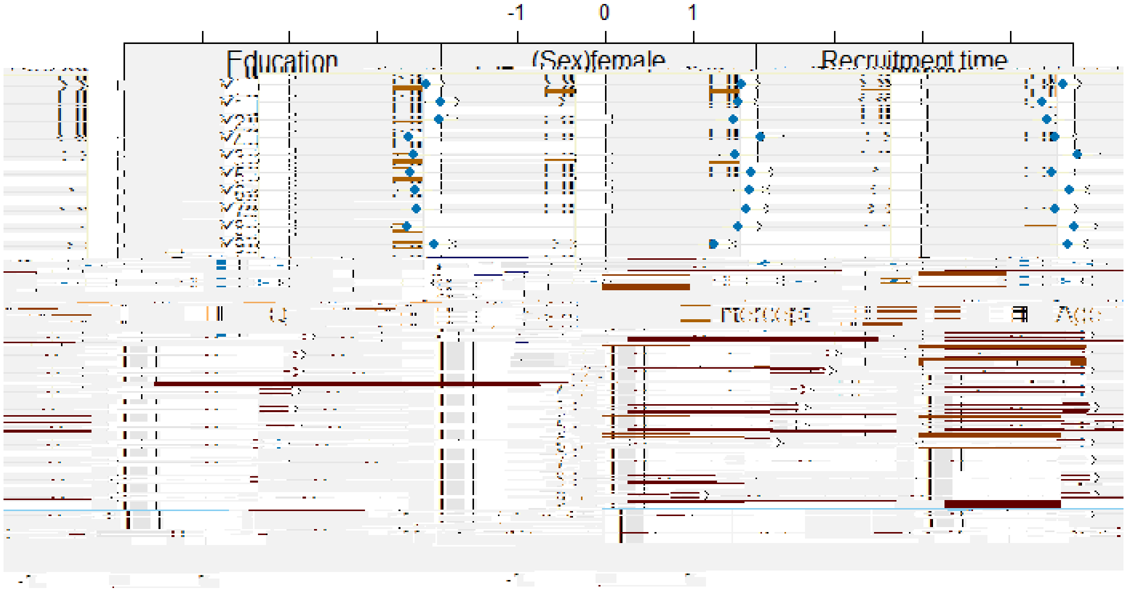

Figure 3.3: Estimate of random effects of the model (3.3a)

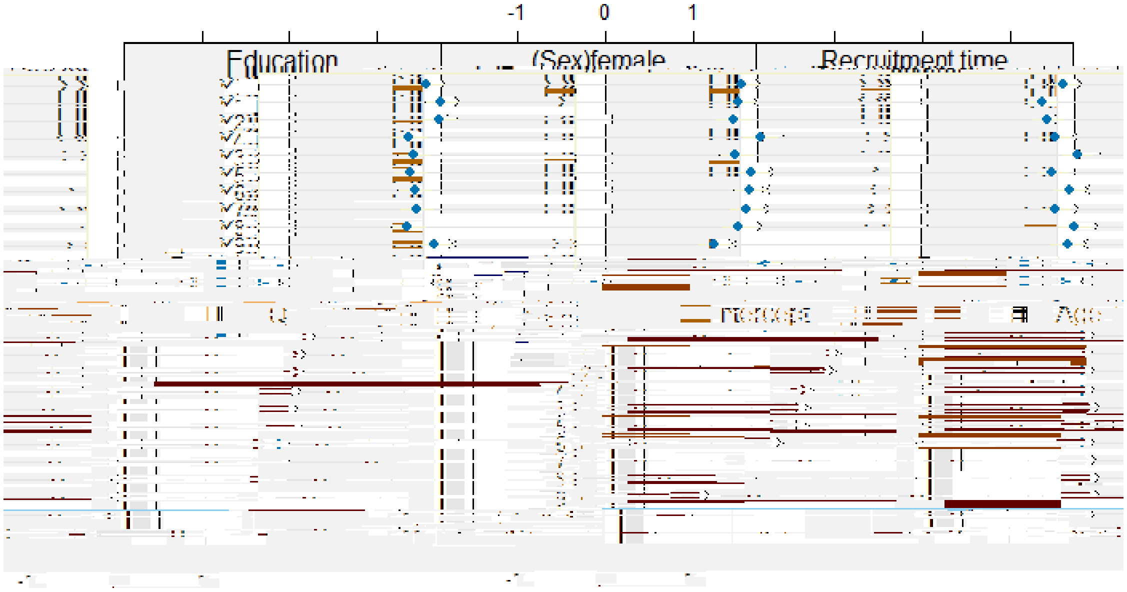

Figure 3.4: Estimate of random effect of the model (3.3b)

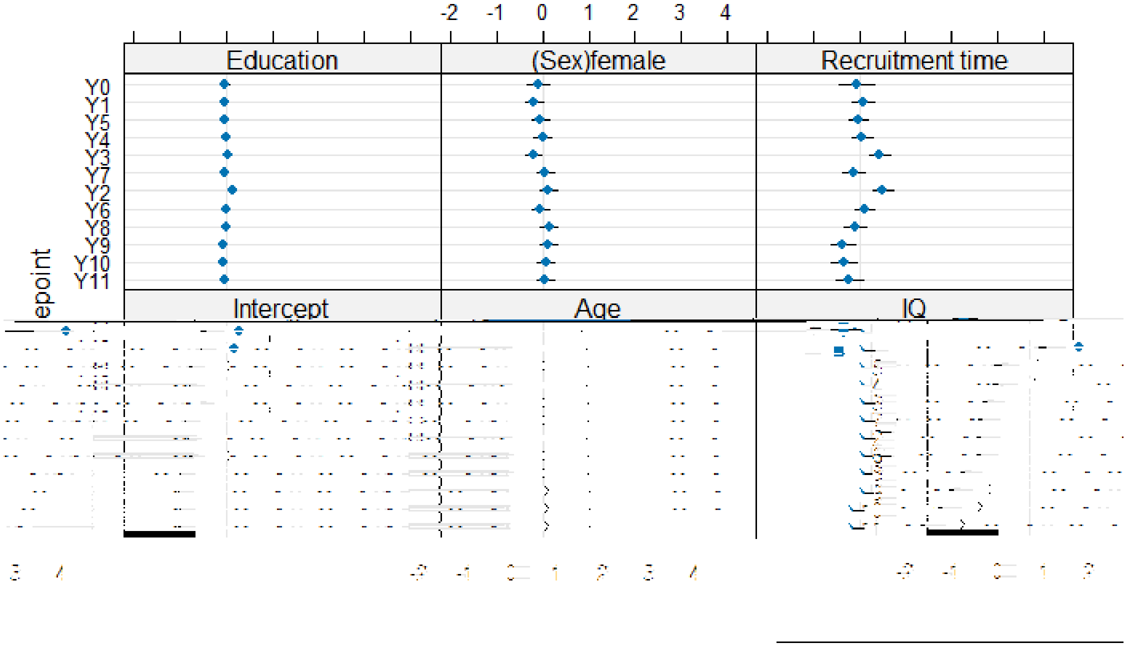

Figure 3.5: Estimate of random effects of the model (3.4a)

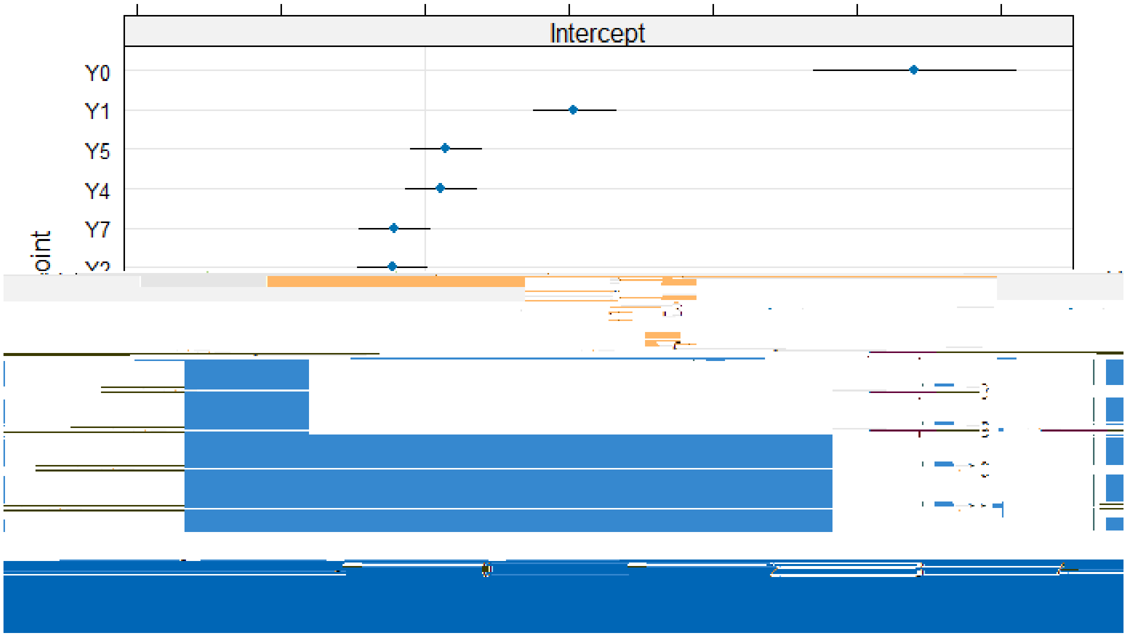

Figure 3.6: Estimate of random effect of the model (3.4b)

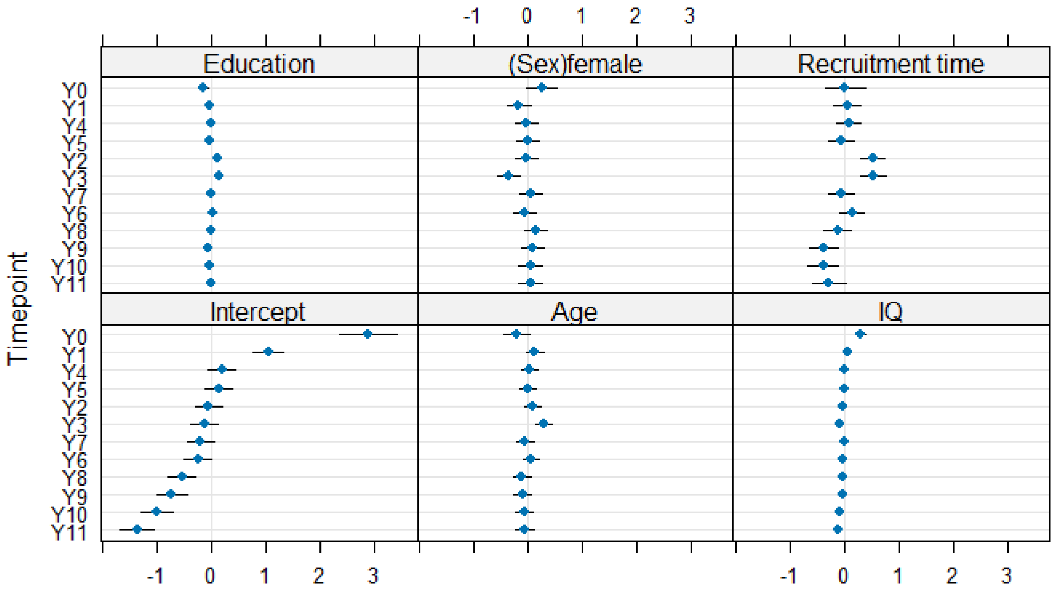

Figure 3.7: Estimate of random effects of the model (3.4a)

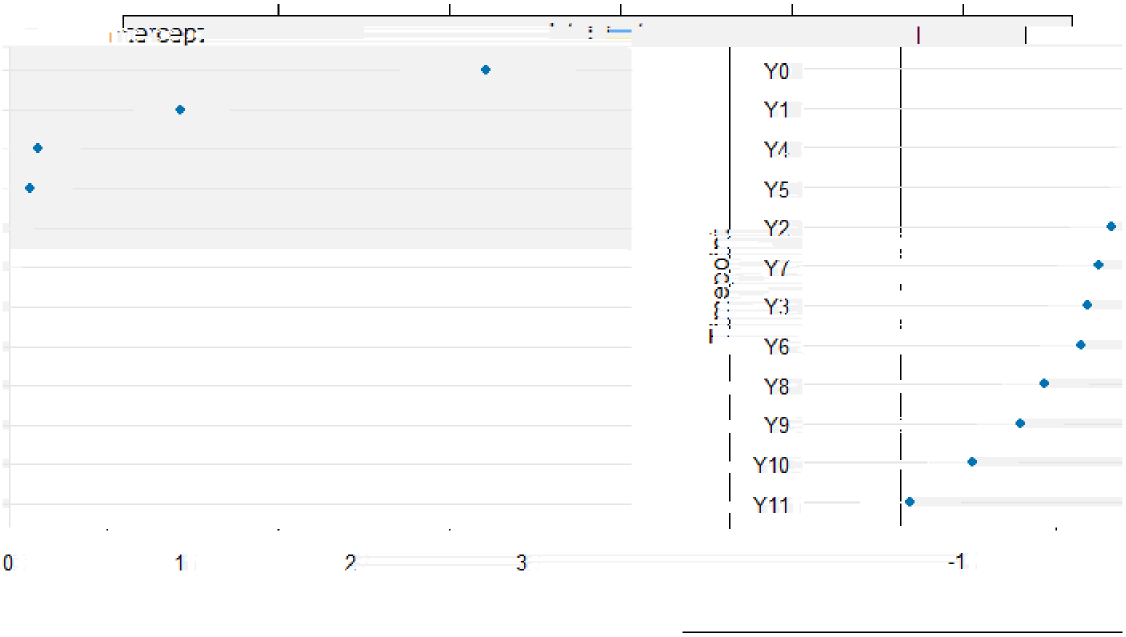

Figure 3.8: Estimate of random effect of the model (3.4b)

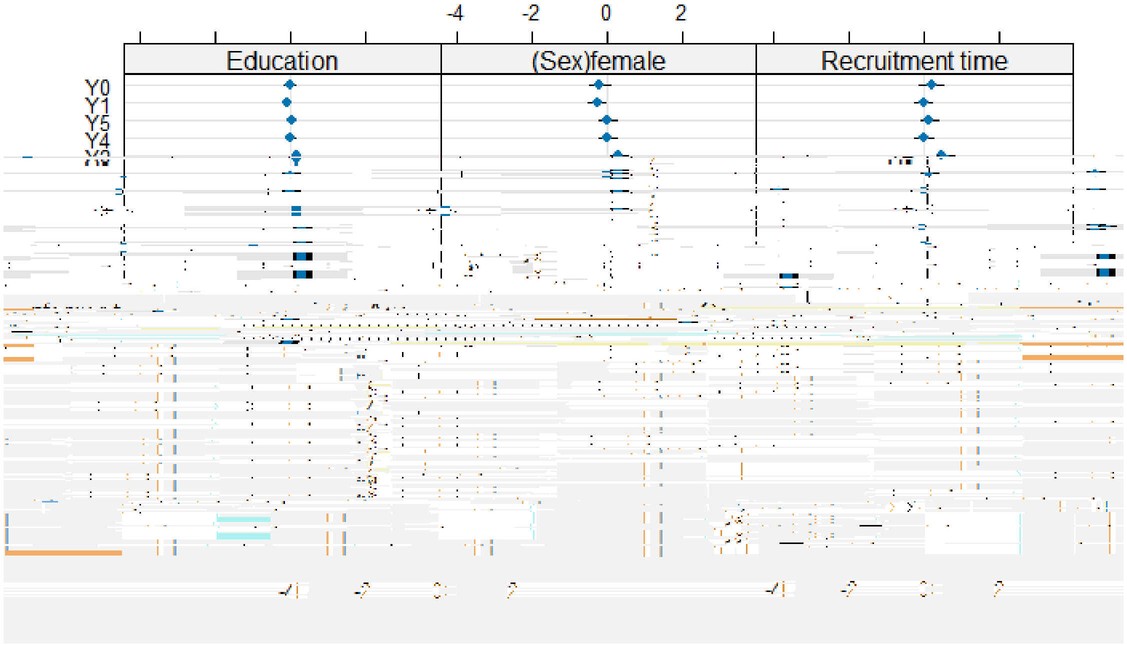

Figure 3.9: Estimate of random effects of the model (3.4a)

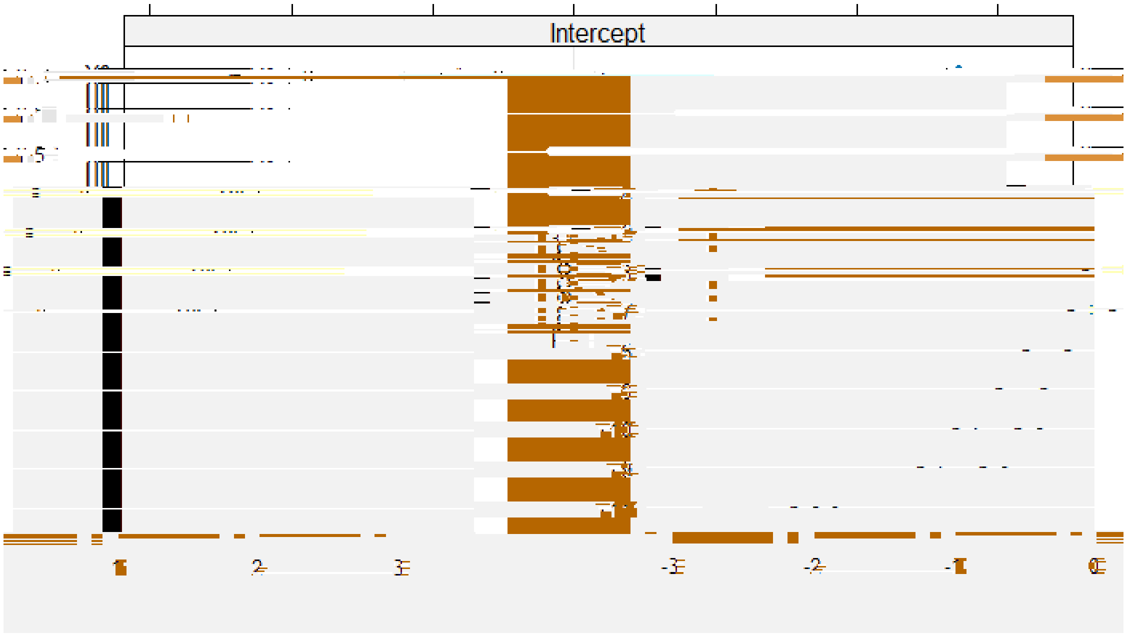

Figure 3.10: Estimate of random effect of the model (3.4b)

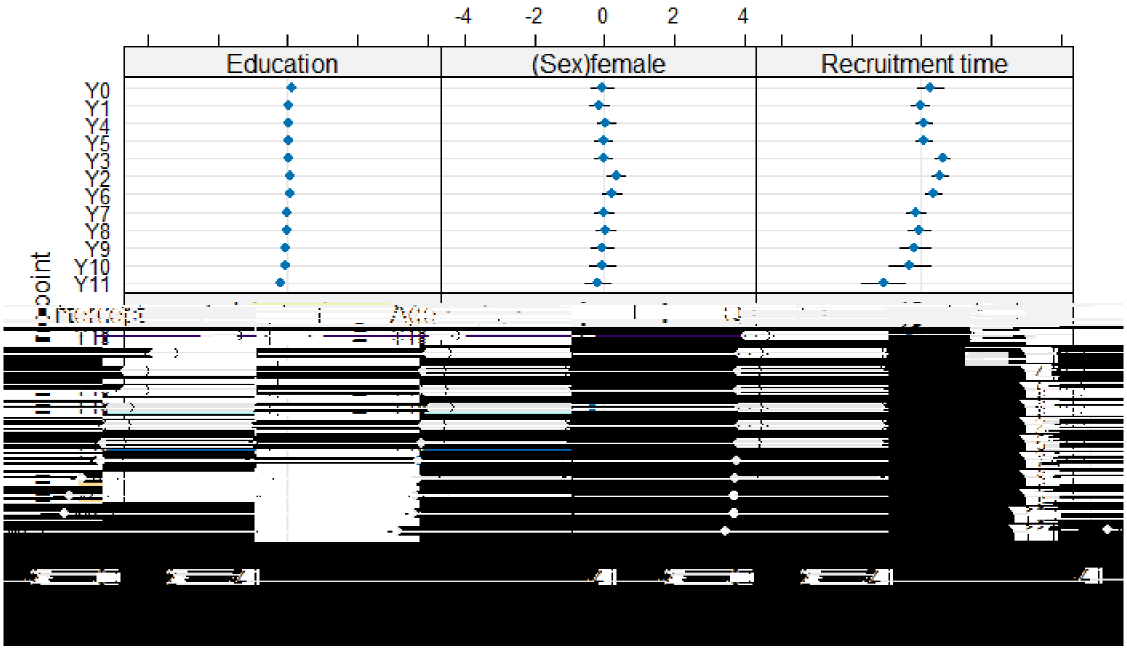

Figure 3.11: Estimate of random effects of the model (3.4a)

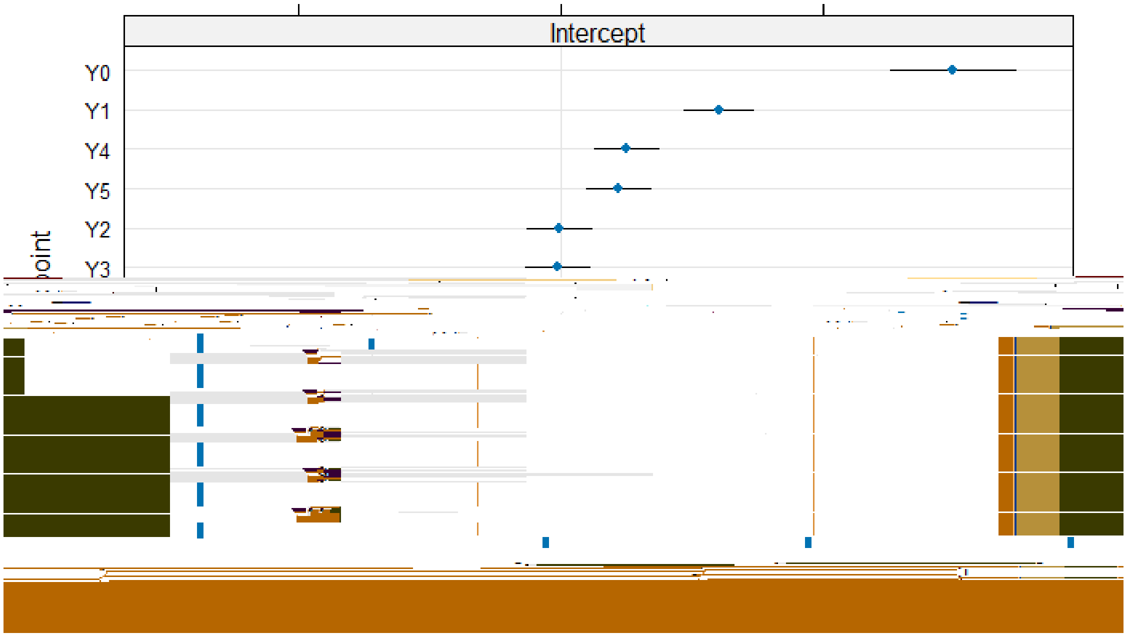

Figure 3.12: Estimate of random effect of the model (3.4b)

Figure 3.13: Estimate of random effects of the model (3.4a)

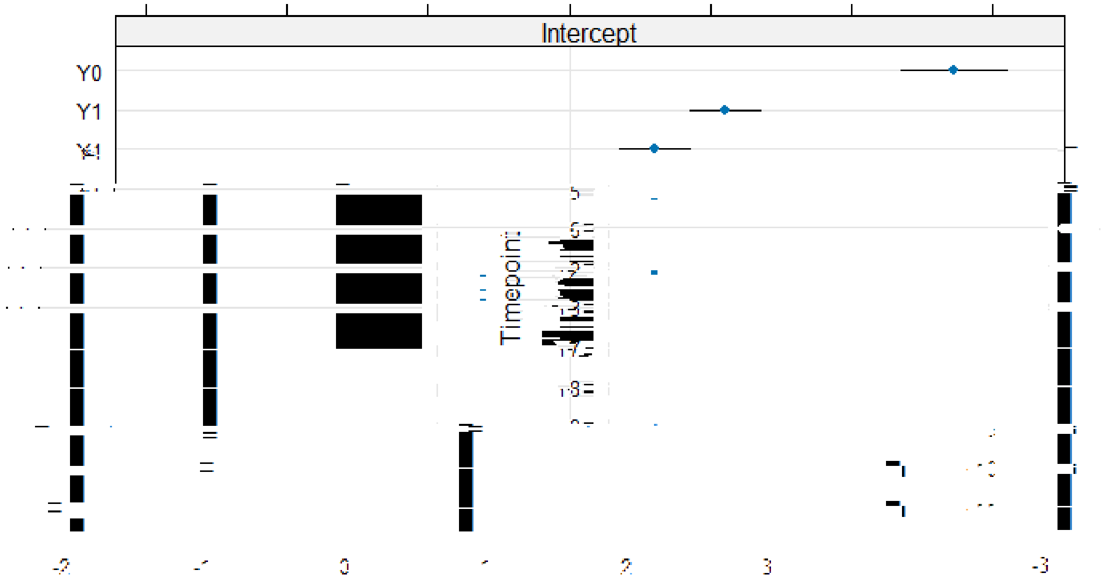

Figure 3.14: Estimate of random effect of the model (3.4b)

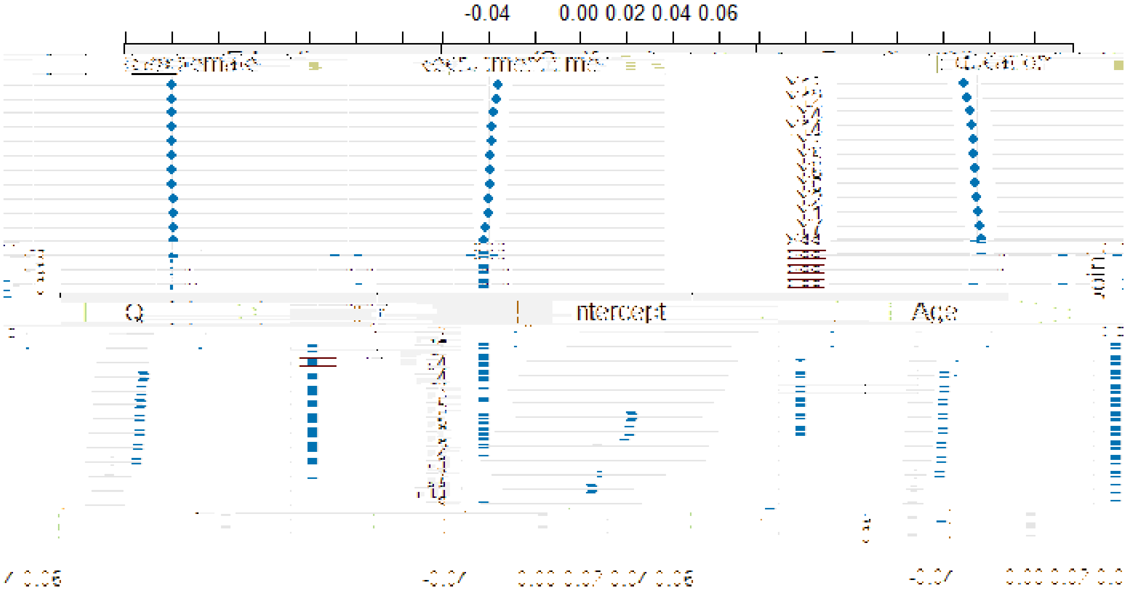

Figure 3.17: Estimate of random effects of the model (3.5.2a)

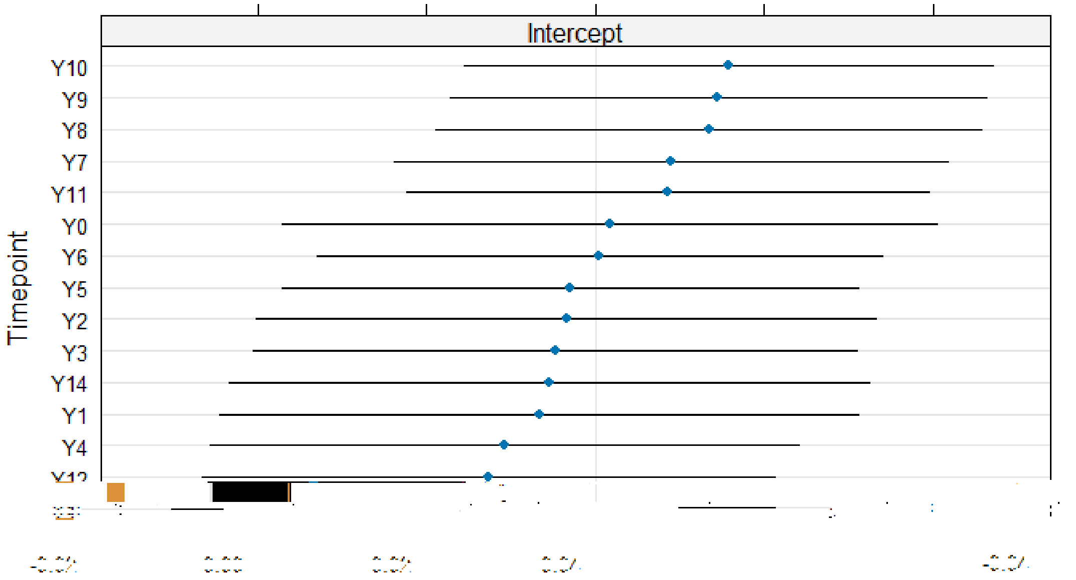

Figure 3.18: Estimate of random effect of the model (3.5.2b)

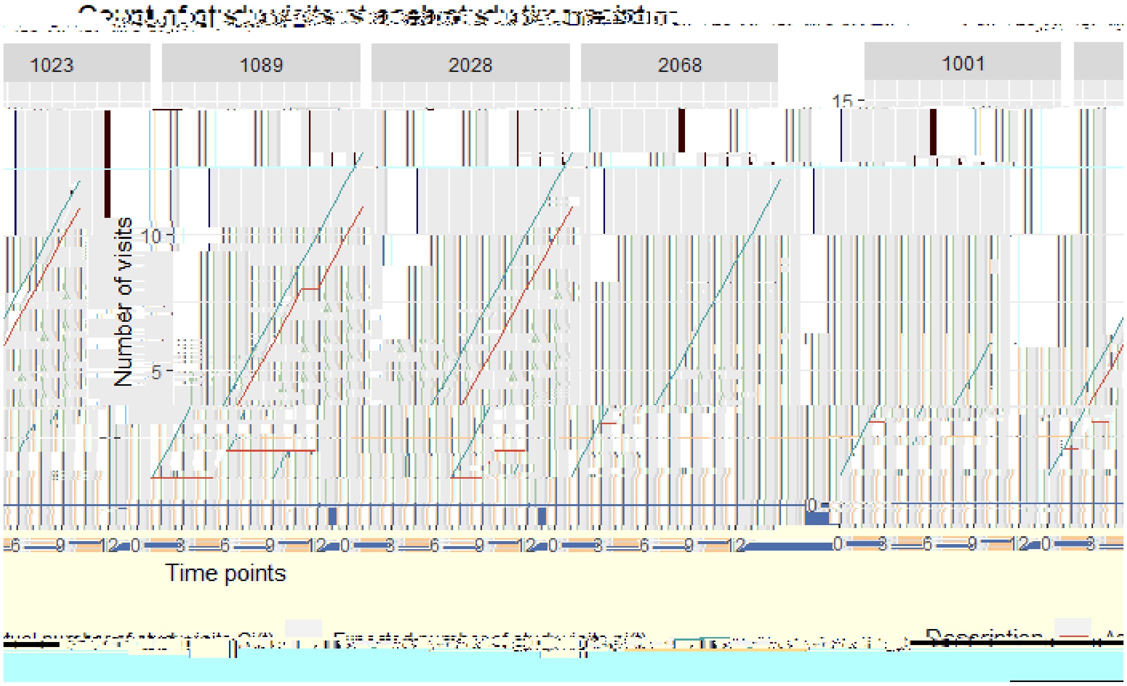

Figure 3.19: Example of number of study visits and ideal visits for participants

VIF	Age	Premorbid IQ score	Education	Sex	Recruitment time	log(n

	Model 3.2a		Model 3.2b		Model 3.2c	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	0:1291	0:2152	0:1064	0:1949	0:0981	0:0430
age at recruitment	0:0151	0:0725	0:0120	0:0406	0:0120	0:0390
premorbid IQ score	0:0822+	0:0481	0:0805+	0:0470	0:0727	0:0451
education	0:0518	0:0557	0:0390	0:0514	0:0315	0:0493
sex(Female vs. male)	0:1900	0:1035	0:1821	0:0936	0:1653	0:0897

enrolT 0 Td :j /R144 10.0213 Tf 0.07 TdTd ( )Tj /R167 10.0213 Tf 7.79453 0 Td (0)Tj /R144 10.0213 Tf 5.7625 0 Td (:)

	Model 3.3a		Model 3.3b		Model 3.3c	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	1:4493	0:2195	1:4411	0:2141	1:5665	0:0339
age at recruitment	0:0789	0:0344	0:0779	0:0321	0:0288	0:0306
premorbid IQ score	0:2022	0:0587	0:1949	0:0367	0:1566	0:0354

	Model 3.4a		Model 3.4b		Model 3.4c	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	0:1262	0:3710	0:1365	0:3530	0:0088	0:0412
age at recruitment	0:0276	0:0772	0:0006	0:0405	0:0006	0:0373
premorbid IQ score	0:0827	0:0531	0:0894	0:0469	0:0744	0:0431
education	0:0813	0:0601	0:0656	0:0513	0:0501	0:0471
sex(Female vs. male)	0:1935	0:1119	0:1962	0:0932	0:1642	0:0861
enroll.date	0:2636	0:0981	0:2146	0:0439	0:1670	0:0403
Num.Obs.	3155		3155		3155	
Num. groups: Timepoint	12		12			
AIC	3846:5619		3856:4567		4359:5942	
BIC	4010:0854		3898:8517		4395:9327	
Log Likelihood	1896:2809		1921:2283		2173:7971	
Deviance					4347:5942	
intercept	1:2699	0:3666	1:2080	0:3487		
age	0:2140	0:0618				
premorbidIQ	0:0437	0:0126				
education	0:0641	0:0185				
sex	0:1515	0:0437				
recruitment	0:2928	0:0845				

Model 3.4a - generalized linear mixed model; all effectsall effectsall effectsll AIC 12

	Model 3.4a		Model 3.4b		Model 3.4c	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	0:0086	0:3152	0:0062	0:2990	0:0625	0:0412
age at recruitment	0:1157	0:0619	0:0964	0:0403	0:0838	0:0373
premorbid IQ score	0:1448	0:0580	0:1227	0:0466	0:1043	0:0432
education	0:0686	0:0577	0:0471	0:0509	0:0356	0:0471
sex(Female vs. male)	0:2828	0:1135	0:2539	0:0925	0:2169	0:0861
enroll.date	0:2130	0:1032	0:1643	0:0433	0:1278	0:0401

	Model 3.4a		Model 3.4b		Model 3.4c	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	0:6355	04238	0:4991	03440	0:3600	0:0438
age at recruitment	0:0556	00735	0:0174	00440	0:0167	00395
premorbid IQ score	0:1465	00836	0:1694	0:0510	0:1347	0:0457
education	0:1600	0:0682	0:			

	Model 3.4a		Model 3.4b		Model 3.4c	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	0:4785	0:4542	0:3219	0:3691	0:2220	0:0434
age at recruitment	0:0174	0:0956	0:0245	0:0439	0:0164	0:0391
premorbid IQ score	0:0449	0:0695	0:0700	0:0506	0:0537	0:0450
education	0:1266	0:0654	0:0934	0:0557	0:0722	0:0495
sex(Female vs. male)	0:0825	0:1244	0:0535	0:1009	0:0466	0:0898
enroll.date	0:4788	0:1612	0:2863	0:0508	0:1456	0:0436
Num.Obs.	2914		2914		2914	
Num. groups: Timepoint	12		12			
AIC	3350:2417		3367:6686		4001:6112	
BIC	3511:6284		3409:5096		4037:4749	
Log Likelihood	1648:1289-12h.988(o.Td [(28T147(6)43426 0 Td (:)Tj /R148 10.021 Tf 2.78437 0 Td (10					

	Model 3.4a		Model 3.4b		Model 3.4c	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	0:8217	0:4387	0:6666	0:3606	0:4472	0:0442
age at recruitment	0:0037	0:0598	0:0259	0:0445	0:0182	0:0397
premorbid IQ score	0:0902	0:0773	0:1261	0:0515	0:0990	0:0459
education	0:1572	0:0674	0:1186	0:0565	0:0943	0:0505
sex(Female vs. male)	0:0343	0:1160	0:0169	0:1025	0:0176	0:0916
enroll.date	0:4419	0:1976	0:2400	0:0508	0:0936	0:0443
Num.Obs.	2914		2914		2914	
Num. groups: Timepoint	12		12			
AIC	3245:9149		3294:8040		3898:3332	
BIC	3329:5969		3336:6450		3934:1969	
Log Likelihood	1608:9575		1640:4020		1943:1666	
Deviance					3886:3332	
intercept	1:4847	0:4286	1:2340	0:3562		
age	0:1215	0:0351				
premorbidIQ	0:1799	0:0519				
education	0:0796	0:0230				
sex	0:1042	0:0301				
recruitment	0:5957	0:1720				

Model 3.4a - generalized linear mixed model; all effects are arbitrary change over time

Model 3.4b - generalized linear mixed model; intercept term is arbitrary change over time

Model 3.4c - generalized linear model

Table 3.9: Estimates and Standard Errors for models (3.4a - c) with k = 5 test

	Model 3.5.1a		Model 3.5.1b		Model 3.5.1c	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	0:7242	0:2580	0:7021	0:2409	0:3900	0:0164

	Model 3.5.2a		Model 3.5.2b		Model 3.5.2c	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	0:0474	00383	0:0449	00389	0:0521	00348
age at recruitment	0:0225	0:0097	0:0219	0:0094	0:0220	0:0094
premorbid IQ score	0:0221	00113	00208	00109	00207	00109
education	0:0079	00122	0:0068	00119	0:0067	00119
sex(Female vs. male)	0:0518	0:0219	0:0515	0:0214	0:0513	0:0214
enroll.date	0:0029	00115	0:0016	00106	0:0037	00101
log(nit)	0:7613	0:0206	0:7614	0:0200	0:7670	0:0169
Num.Obs.	3160		3160		3160	
Num. groups: Timepoint	15		15			
AIC	123429435		12290:4644		12289:0010	
BIC	125549849		12338:9310		12331:4093	
Log Likelihood	6136:4717		6137:2322		6137:5005	
Deviance					2715:1024	
intercept	0:0254	0:0066	0:1024 [illegible]			

Chapter 4

Analyses the Cognitive Test Scores of the Hotel Study Participants

We are now interested in how factors affect participants' scores in the Iowa gambling task (IGT) with missing data issue present. Participants are expected to visit the study once a year and take five cognition tests during each visit. The previous chapter suggested that the missing mechanism for the test of IGT is missing not completely at random; the probability of participant i had an IGT score at time point j is associated with the participant's premorbid IQ score, education, and recruitment time in the study time scale for $j = 0; \ldots; 11$. In this chapter, we assume all participants had baseline scores of IGT Y_{i05}, and first analyze the relationship between IGT net score and factors every year under the assumption that missing completely at random. In Chapter 3, we showed the evidence against the missing completely at random assumption, we then assume the missing mechanism is not missing completely at random, and use the weighted generalized estimating equations (Lin et al., 2015; Robins et al., 1995) to explore the relationship between IGT score and covariates. For convenience, we let Y_{ij} toi(Lin 2.8832 0 Td6 0 Td 94(4.937(ge)-0.0 (Y)Tj /R146 7.9701 T.494.4.4.4.4.4.05

with $_{ij} \quad N(0; \; ^2)$.

Analysis results

Under the setting that the IGT scores are independent within participants and the assumption that missing is completely at random, we used the available data to explore the relationship between IGT net score and factors. Table 4.1 shows that participants' IGT net score could be associated with premorbid IQ score, participants with higher premorbid IQ score would have higher IGT score.

4.1.2 Assuming IGT Scores are Correlated

We consider under the assumption that missing completely at random, the within-participants correlation is taken into account.

Model

- Linear mixed effect model with the random effect associated with the i^{th} participant:

$$Y_{ij} = {}_{0i} + {}_{1}X_{i1} + {}_{2}X_{i2} + {}_{3}X_{i3} + {}_{4}X_{i4} + {}_{5}X_{i5} + {}_{ij} \qquad (4.2a)$$

  with $_{0i} \quad N(\;_0; \;^2_0)$ and $_{ij} \;\_\, N(0; \;^2)$. $_{0i}$ and $_{ij}$ are assumed to be independent

- GEE method (Hedeker and Gibbons, 2006) with model as follows:

$$Y_{ij} = {}_{0} + {}_{1}X_{i1} + {}_{2}X_{i2} + {}_{3}X_{i3} + {}_{4}X_{i4} + {}_{5}X_{i5} + {}_{ij} : \qquad (4.2b)$$

  The within-participants covariance structures R we considered are AR(1) working correlation structure that the correlation over time is a function of lag:

$$\mathrm{Corr}(Y_{ij}; Y_{i;j+l}) = \;^{t} \text{ for } l = 0; 1; 2; :::; 11 \quad l$$

  as well as the exchangeable working correlation structure that all correlations are equal:

$$\mathrm{Corr}\; Y_{ij}; Y_{ij'} = \begin{cases} 1 & j = j^0 \\ & j \neq j^0 \end{cases}$$

  Thus, the variance-covariance matrix for Y_i is

$$V_i = \; R_i(\;)$$

  in the normal case. The unknown correlation parameter and scale parameter will be estimated in analyses.

Analysis results

Based on Table 4.2, we obtain a similar conclusion that participants' IGT score is significantly associated with premorbid IQ score; participants with higher

Weights are calculated based on $\hat{w}_{ij}$
$(R_{i0}, \ldots, R_{i,j-1})$.

- Missingness based on participants'
most recent attendance data of the I

$$\mathrm{logit}\,[P(R_{i1} = 1 \mid R_{i1}; X$$

$$\mathrm{logit}\,[P(R_{i2} = 1 \mid R_{i2}; X_{i2}; X$$

$$\mathrm{logit}\,[P(R_{i,11} = 1 \mid R_{i,11}; X_{i2};$$

Weights are calculated based on $\hat{w}_{ij}$
$(R_{i,j-1})$.

We will assign different weights to participa
independent over time in section 4.2.1 an
4.2.2.

4.2.1 Assuming IGT Scores are Inde

To begin with, we assume the observatio
weighted GEE method with the independ
Based on Table 4.3, we cannot find any as
factors.

4.2.2 Assuming IGT Scores are Cor

The result for the WGEE method with we
morbid IQ score, recruitment time, and all
we conclude that older participants would h
exchangeable working correlation structur
association of participants' IGT score with
study late could have higher IGT scores.

While we consider the WGEE method wit
morbid IQ score, recruitment time, and th
4.5, older participants would have higher I

between participants' premorbid IQ scores and IGT scores. Namely, under the missing completely at random assumption, participants with higher premorbid IQ scores would have higher IGT scores. However, based on Table 3.9, we also conclude that participants with lower premorbid IQ scores are less likely to attend the IGT test. It is possible that participants who missed the IGT tests with lower premorbid IQ scores can still achieve higher scores on the IGT test. Thus, it is reasonable that the IGT score is not significantly associated with premorbid IQ after we handling the missing data issue in the IGT test.

Tables and Figures

	Model 4.1	
	Est.	S.E.
(Intercept)	0.0159	0.0342
age at recruitment	0.0175	0.0310
premorbid IQ score	0.1712	0.0365
education	0.0512	0.0407
sex(Female vs. male)	0.0469	0.0708
enroll.date	0.0102	0.0331
Num.Obs.	1151	
AIC	3269.7720	
BIC	3305.1107	
Log.Lik.	1627.8860	
Deviance	1140.4244	

IGT scores are independent

Regression based on available data

Table 4.1: Estimates and Standard Errors for model (4.1)

	Model 4.2a		Model 4.2.1b		Model 4.2.2b	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
(Intercept)	0:0176	0.0524	0:0278	0.0500	0:0206	0.0498
age at recruitment	0:0257	0.0473	0:0286	0.0471	0:0270	0.0467
premorbid IQ score	0:2064	0:0555	0:2079	0:0556	0:2109	0:0544
education	0:0901	0.0612	0:0860	0.0576	0:0931	0.0574
sex(Female vs. male)	0:0485	0.1100	0.0427	0.1153	0.0496	0.1062
enroll.date	0:0089	0.0494	0.0200	0.0467	0.0111	0.0442
Num.Obs.	1151		1151		1151	
Num. groups: SubjectNumber	258					
AIC	3056.7		3275.1		3273.9	
BIC	3097.1		3310.5		3309.2	
Log.Lik.			1630.568		1629.947	
intercept	0:5942	0:0370				
Scale parameter: gamma			0:9950	0.0785	0.9940	0.0786
Correlation parameter: alpha			0:7050	0.0479	0.4440	0.0603
Num. clust.			258		258	

Regression based on available data

Model 4.2a - mixed linear effect model; intercept term is arbitrary change across different participant

Model 4.2.1b - generalized linear model with generalized estimating equations method; AR(1) working correlation matrix

Model 4.2.2b - generalized linear model with generalized estimating equations method; exchangeable working correlation matrix

Table 4.2: Estimates and Standard Errors for models (4.2a), (4.2.1b), (4.2.2b)

	Model 4.3.1a		Model 4.3.1b	
	Est.	S.E.	Est.	S.E.
(Intercept)	0:2137	02059	00207	01016
age at recruitment	0:4674	02965	01191	01238
premorbid IQ score	0:0359	01921	01435	01073
education	0:1592	01501	00053	00790
sex(Female vs. male)	0:3030	02817	02021	02278
enroll.date	0:1374	01218	0:0443	00707
Num.Obs.	1053		1053	
Num. clust.	235		235	
Scale parameter: gamma	1:4900	01990	11700	01520

	Model 4.3.2a		Model 4.3.2b	
	Est.	S.E.	Est.	S.E.
(Intercept)	0:4345+	0:2371	0:3492	0:2335
age at recruitment	0:6803	0:2844	0:7053	0:2949
premorbid IQ score	0:2729	0:2894	0:1693	0:2714
education	0:1751	0:2679	0:1905	0: 27140

	Model 4.3.3a		Model 4.3.3b	
	Est.	S.E.	Est.	S.E.
(Intercept)	0:0388	01739	00214	01669
age at recruitment	0:3856+	0:2069	0:2874	02069
premorbid IQ score	0:1257	01832	01756	01734
education	0:0208	01375	0:0104	01288
sex(Female vs. male)	0:6951+	0:3648	0:4474	03632
enroll.date	0:0811	01244	0:0567	01136
Num.Obs.	1053		1053	
Num. clust.	235		235	
Scale parameter: gamma	1:2700	02040	12000	01760
Scale parameter: alpha	0:8290	00626	06200	01160

Model4.3.3a - QIC = 1515 for AR(1) working correlation structure

Model4.3.3b - QIC = 1523 for exchangeable working correlation structure

Table 4.5: Estimates and Standard Errors for models (4.3.3a-b)

Chapter 5

Discussion

5.1 Summary

The data we obtained from the Hotel Study research team contains information for sample participants who lived on DTES; more specifically, sample participants were collected from three different sources on DTES: the SRO hotels, a community court, and the Emergency Room at St. Paul's hospital; additional youth group to collect any participants if they were below 30 years old when recruiting. We first performed descriptive analyses to compare the similarity of participants in four groups. We found the distributions of age at recruitment time, premorbid IQ score, education level received in years, and cognitive test scores are not that similar in the four groups; participants from different groups may come from different populations. Furthermore, the data contains 272 participants from the hotel group, 71 participants from the community group, 41 participants from the youth group, and 61 participants from the hospital group; due to the large sample size of the hotel group, when we pool all groups together, the hotel group would dominate the analyses. Thus, we choose the original hotel sample to be the random sample of the target population; in the rest of the project, we focused on the hotel group participants only.

Next, through cross-sectional analyses and longitudinal analyses, we investigated the efpants who joined the study late are less likely to attend any of the tests. Besides that, other factors also affect the probability of test attending:

- k = 1 test: The probability of attending verbal learning and memory test is significantly associated with sex in the mixed-effect model; female participants are more likely to visit this test compared to male participants.
- k = 2 test: All models suggest that the probability of attending an inhibitory control test is associated with premorbid IQ score and sex. Participants with higher measured premorbid IQ scores at baseline are more likely to visit this test; female participants are more likely to visit this test compared to male participants. The probability of attending this test could also be associated with age at recruitment; younger participants are more likely to attend the test.
- k = 3 test: All models suggest that the probability of attending sustained attention and processing speed test is associated with education level; participants with lower education level are more likely to attend this test. The mixed-effect model with intercept term varying over time and the generalized linear model indicate that participants

5.2.5 Study Sample

The Hotel Study contains four study samples that were collected in different time frames. With the time changes, the standards for selecting participants would change; we could test if two samples are from the same population (Wald and Wolfowitz, 1940). We chose to use the original hotel sample as a random sample for the target population in this project. However, we wasted 40% of the total participants we have. Future investigation can consider making use of the whole study sample with some weight adjustment.

5.2.6 Relationship between Variables

Based on Figure 2.8, we found that participants with higher premorbid IQ scores would have higher education levels. Based on Tables 2.9 and 2.10, we found trends that participants who joined the study late may have higher education levels or premorbid IQ scores. Future studies can investigate how these variables relate to each other, and why participants with later recruitment time tend to have higher education levels or premorbid IQ scores.

Bibliography

Intra-extra dimensional set shift (ied). https://www.cambridgecognition.com/cantab/cognitive-tests/executive-function/intra-extra-dimensional-set-shift-ied/. Accessed: 2023-03-19.

Rapid visual information processing (rvp). https://www.cambridgecognition.com/cantab/cognitive-tests/attention/rapid-visual-information-processing-rvp/. Accessed: 2023-03-19.

Aho, K., Derryberry, D., and Peterson, T. Model selection for ecologists: the worldviews of aic and bic. Ecology, 95(3):631 636, 2014.

Baitz, H. A., Jones, P. W., Campbell, D. A., Jones, A. A., Gicas, K. M., Giesbrecht, C. J., Loken Thornton, W., Barone, C. C., Wang, N. Y., Panenka, W. J., et al. Component processes of decision making in a community sample of precariously housed persons: Associations with learning and memory, and health-risk behaviors. Frontiers in Psychology, 12:571423, 2021.

Baraldi, A. N. and Enders, C. K. An introduction to modern missing data analyses. Journal of school psychology, 48(1):5 37, 2010.

Baranyi, G., Scholl, C., Fazel, S., Patel, V., Priebe, S., and Mundt, A. P. Severe mental illness and substance use disorders in prisoners in low-income and middle-income countries: a systematic review and meta-analysis of prevalence studies. The Lancet Global Health, 7(4):e461 e471, 2019.

Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. W. Insensitivity to future study clinical and health service effectiveness in a cohort of homeless or marginally housed persons.The Canadian Journal of Psychiatry, 62(7):482 492, 2017.

Jones, A. A., Cho, L. L., Kim, D. D., Barbic, S. P., Leonova, O., Byford, A., Buchanan, T., Vila-Rodriguez, F., Procyshyn, R. M., Lang, D. J., et al. Pain, opioid use, depressive symptoms, and mortality in adults living in precarious housing or homelessness: a longitudinal prospective study. Pain