Attrition in Studies of Cognitive Aging

Maria Josefsson

Department of Statistics
Umeå School of Business and Economics
Umeå University 2013

Doctoral Dissertation
Department of Statistics
Umeå School of Business and Economics
Umeå University
SE-901 87 Umeå

Statistical Studies No. 47
ISBN: 978-91-7459-760-8
ISSN: 1100-8989
Cover: Håkan Gustafson
Electronic version available at http://umu.diva-portal.org/

Printed by: Print & Media
Umeå, Sweden 2013

As we grow older, we may grow wiser...

Contents

List of Papers

The thesis is based on the following papers:

I. Josefsson, M., de Luna, X., Pudas, S., Nilsson, L. G., and Nyberg, L. (2012). Genetic and Lifestyle Predictors of 15-Year Longitudinal Change in Episodic Memory. *Journal of the American Geriatrics Society*, 60(12), 2308-2312.

II. Pudas, S., Persson, J., Josefsson, M., de Luna, X., Nilsson, L. G., and Nyberg, L. (2013). Brain Characteristics of Individuals Resisting Age-Related Cognitive Decline over Two Decades. *The Journal of Neuroscience*, 33(20), 8668-8677.

III. Josefsson, M., de Luna, X., Daniels, M, and Nyberg, L. (2013). Causal inference with longitudinal outcomes and non-igorable drop-out: Estimating the effect of living alone on cognitive decline. *Manuscript*

IV. Josefsson, M., Lundquist, A., and Nyberg, L. (2013). Imputation of missing longitudinal fMRI data. *Manuscript*

Abstract

Longitudinal studies of cognition are preferred to cross-sectional studies, since they offer a direct assessment of age-related cognitive change (within-person change). Statistical methods for analyzing age-related change are widely available. There are, however, a number of challenges accompanying such analyzes, including cohort differences, ceiling- and floor effects, and attrition. These difficulties challenge the analyst and puts stringent requirements on the statistical method being used.

The objective of Paper I is to develop a classifying method to study discrepancies in age-related cognitive change. The method needs to take into account the complex issues accompanying studies of cognitive aging, and specifically work out issues related to attrition. In a second step, we aim to identify predictors explaining stability or decline in cognitive performance in relation to demographic, life-style, health-related, and genetic factors.

In the second paper, which is a continuation of Paper I, we investigate brain characteristics, structural and functional, that differ between successful aging elderly and elderly with an average cognitive performance over 15-20 years.

In Paper III we develop a Bayesian model to estimate the causal effect of living arrangement (living alone versus living with someone) on cognitive decline. The model must balance confounding variables between the two living arrangement groups as well as account for non-ignorable attrition. This is achieved by combining propensity score matching with a pattern mixture model for longitudinal data.

In paper IV, the objective is to adapt and implement available imputation methods to longitudinal fMRI data, where some subjects are lost to follow-up. We apply these missing data methods to a real dataset, and evaluate these methods in a simulation study.

Keywords: Attrition, missing data, age-related cognitive change, non-ignorable dropout, monotone missing pattern, mixture models, pattern-mixture models, imputation

Acknowledgements

This journey is soon to be over and it's with definitely no mixed feelings I'm about to submit this thesis - pure relief. There are however many people who have made this trip much less painful and I'm truly grateful for your support.

First and foremost, I would like to thank my main supervisor Xavier de Luna, without your constant support, encouragement and guidance this thesis would never have been written. I would also like to thank you for giving me freedom to pursue my own research interests while at the same time keeping me on the right track.

I want to thank my co-supervisors: Lars Nyberg, for your inspirational guidance, and expertise, as well as for introducing me to and guiding me through the world of neuroscience; Anders Lundquist and Ingrid Svensson, for your guidance, encouragement and care.

I also want to thank some of my co-authors: Lars-Göran, for being a great source of inspiration and for giving me access to such a high-quality data set; Sara, for insightful comments and your (sometimes annoying) eye for details - making our papers so much better; Mike, for your knowledgeable advice and patience. It has been a pleasure working with all of you.

I also would like to thank all of my colleagues at the Department of Statistics for your great support during these years. Especially Katarina for helping me keep track of things (including me) when I don't. Jenny, Emma, Marie (all of you), Henke and Mathias for our daily non-work-related conversations, get-togethers outside the office and for always knowing how to put a smile back on my face. Emma and Jenny - I think it's time for our 24-h-without-complaints some time soon. Also, a special thank to Jenny and Mathias for providing valuable feedback on my thesis.

A special gratitude to the Graduate School in Population Dynamics and Public Policy for providing a stimulating and friendly research environment as well as for their generous financial support. This support made it possible for me to participate in several conferences abroad, as well as spend a semester at University of California in Berkeley.

I am also grateful to all the talented and dedicated colleagues at UFBI for being open and friendly to newcomers like me.

To all my friends outside the department. Thank you for your love, support and friendship. And for giving me the distraction needed every now and then. You guys are amazing!

A very special thanks to my family. I wouldn't be the person I am today without the love and support from each and every one of you. Also for reminding me once in awhile that imputing is just about making up

numbers...

Above all I would like to thank Håkan, for putting up with me the past few months, for standing by me and having faith in me when I don't. I love you!

Umeå, October 2013
Maria Josefsson

1 Introduction

Normal aging is associated with decline in memory and other higher-order cognitive functions (Rönnlund et al., 2005; Schaie et al., 1994; Verhaeghen and Salthouse, 1997). There are, however, large discrepancies between individuals (Habib et al., 2007), where some elderly adults have atypically large decline, whereas others maintain high levels of cognitive functions in old age (Christensen et al., 1999). Accelerated decline has been much studied and linked to pathology such as dementia diseases (Bäckman et al., 2001; Palmer et al., 2008). By contrast, factors determining well-preserved cognition, "successful aging" (Rowe and Kahn, 1987), in older age remain less well characterized (Yaffe et al., 2009).

Longitudinal studies offer a direct assessment of age-related change (within-person change), in contrast to cross-sectional studies, where estimates rather show age differences (between-person change). Statistical techniques for analyzing longitudinal change are widely available (Laird and Ware, 1982; Zeger et al., 1988). However, a number of challenges accompany analyses of cognitive aging, such as cohort differences, ceiling and floor effects, and attrition. These difficulties challenge the analyst and puts stringent requirements on the statistical method being used. First, attrition - loss of participants to follow-up, is one of the major methodological problems in longitudinal studies. There is evidence of attrition causing severe positive bias (Cooney et al., 1988; Yaffe et al., 2010; Dufouil et al., 2004; Caracciolo et al., 2008), because those remaining in the study have systematically better cognitive performance than those dropping out. Furthermore, if participants with a more rapid decline in memory performance tend to drop out earlier from the study, the attrition is related to this progression and hence is informative (Little, 1995).

Second, cohort differences are group discrepancies that appears, when examined at the same age, later born cohorts are found to be cognitively fitter and experience less rapid cognitive decline than earlier born cohorts (Aldwin et al., 2006; Bäckman et al., 1999; Schaie, 2012; Gerstorf et al., 2011). It is important to separate the effects of cohorts and historical differences from those processes underlying aging.

A third consideration arises from the fact that when tests are performed repeatedly over time participant may perform better at subsequent test occasions because of practice effects.

Finally, cognitive tests are usually discrete and consists of sum-scores of different items, which gives a limited range of possible values. If the participant's performance level is outside the range of the test, this results in problematic distribution of the cognitive tests such as ceiling and floor effects (Morris et al., 1999; Jacqmin-Gadda et al., 1997;

Proust-Lima et al., 2011). Causing an artificial association between initial level and change, since individuals in the upper or lower end of the distribution only can change in one direction.

The aim of this thesis is to develop methods for studying age-related cognitive change. Importantly, these methods must take into account the complex issues accompanying longitudinal studies of cognitive aging, and specifically work out issues related to attrition.

2 The Betula Study

The Betula project (Nilsson et al., 1997, 2004) is an ongoing prospective cohort study of nearly 4500 adults. The objective is to study how memory functions change over time and identify risk factors for dementia. Data has so far been collected every 5th year since 1988-1990. At present, five waves of data collection have been completed and a sixth wave started during the fall of 2013. Participants were recruited using random selection from the population registry in Umeå, Sweden, comprising of individuals evenly distributed across 11 age cohorts, 35, 40, 45,... 85 years of age, and balanced on sex to reflect the distribution in the population within that age-cohort. At each visit participants underwent a health examination, questionnaires, and an extensive battery of cognitive tests.

Participants from several samples have been tested over the years, though not all were tested at each wave. Two samples have been followed longitudinally over 15-20 years; Sample 1 (S1) was recruited in 1988-1990 and Sample 3 (S3) in 1993-1995. Data from these two samples, and one additional sample, Sample 6 (S6), are used in this thesis. These are further described below.

In addition to cognitive data, structural and functional MRI (magnetic resonance imaging) was also available. First, we considered a longitudinal brain imaging study, where the initial objective was to study genetic variation in relation to brain function. Participants (n=60), mainly from S1 (additionally three individuals from Sample 2), were recruited in 2002-2003 which was followed by a second test wave in 2008-2009. Second, in connection with the fifth measurement wave, (2008-2010) in Betula, another subsample from S1 and S3 were also scanned with structural and functional MRI. Additional participants from a newly recruited sample (S6) were also included in the latter imaging study, adding up to 375 persons in total. Participation in this study was randomly offered to individuals, stratified by age and gender, who had completed cognitive testing at T5 until all slots were fille.

The study was approved by the Regional Ethical Vetting Board at Umeå University and all participants gave informed written consent.

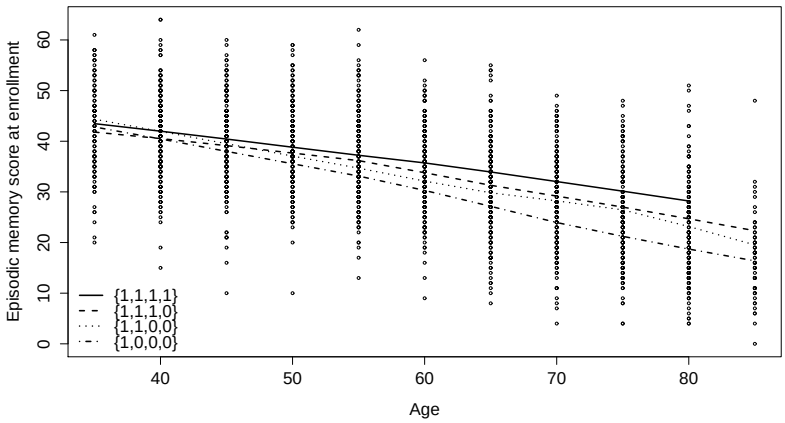

Figure 1: Lowess curves fitted to Episodic memory score at enrollment for approximately 2000 participants in the Betula study. Participants were stratified into four groups on the basis of their last available measurement; $\{1,0,0,0\}$,$\{1,1,0,0\}$,$\{1,1,1,0\}$ and $\{1,1,1,1\}$, where 0 indicates missing and 1 indicates observed for each of the four measurement waves.

2.1 Study participants: Paper I

The study participants consisted of 1954 individuals from S1 and S3. At enrollment the samples comprised of approximately 1000 individuals each, evenly distributed over 11 age-cohorts with 5-year intervals ranging from 35-85 years of age, and approximately balanced on sex. Cognitive, lifestyle and medical data from four test waves were used in this study.

Memory was assessed using number of correct answers from a composite of five validated episodic memory tasks sensitive to mild cognitive deficits (Nilsson et al., 1997). The composite consisted of: (1) immediate free recall of 16 imperative verb-noun sentences that were enacted by the participant, (2) delayed cued recall of nouns from the previously enacted sentences, (3) immediate free recall of 16 verbally and visually presented verb-noun sentences, (4) delayed cued recall of nouns from the previously presented sentences, and (5) immediate free recall of 12 verbally presented nouns. Testing procedures remained constant across measurement occasions. The range of the episodic memory composite scores is 0 to 76, with a higher score indicating better episodic memory.

Only 45% of the participants from S1 and S3 completed the fourth test wave. Reasons for attrition included withdrawal, moving out of the area, sickness and dementia, death, no contact, and unknown/unspecified reasons. Although the study was based on a random sample, attrition

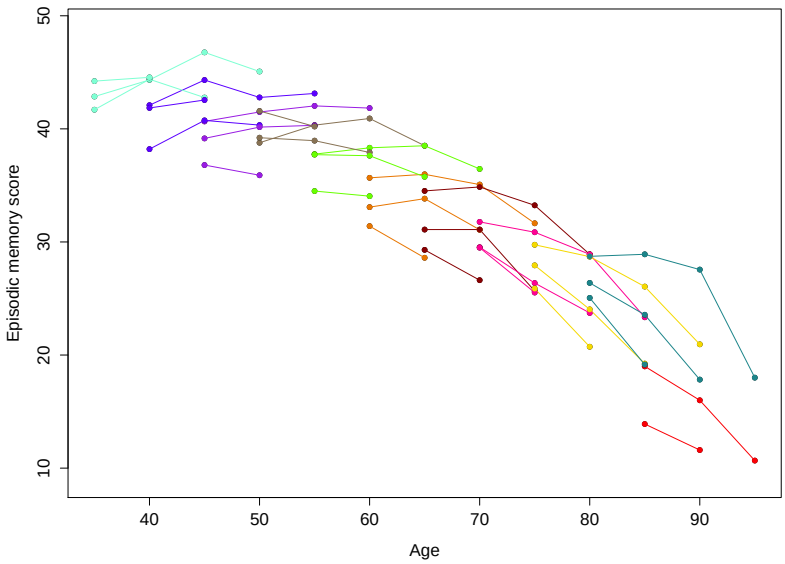

Figure 2: Longitudinal average episodic memory score for approximately 1500 participants in the Betula study with two or more measurements. Participants were stratified by age and dropout groups, where each line represents a dropout group stratified by age, and each color represents an age cohort.

was expected to be related to observed as well as unobserved memory performance. Participants were combined into four groups on the basis of their last available measurement. The performance level differ between the dropout groups, where individuals participating longer in the study perform better on the episodic memory composite test, both at baseline and over time, compared to early dropouts. When stratified by age, middle aged participants do not show this attrition bias. However among elderly participants the bias increases, see Figure 1 and 2.

2.2 Study participants: Paper II

The study participants consisted of individuals that underwent structural and functional MRI in the imaging study. Participants were subdivided into three groups. One group of successful-agers (n=51): all elderly participants in the imaging sample who had been classified as cognitive maintainers using the model presented in Study I, and one control group (n=51): participants classified as having average cognitive development over 15-20 years, matched on age (person by person). The classification is based on data up to the fifth test wave (when the

imaging study started), 20 years for S1 and 15 years for S3. Finally, one young reference group (n=45): 45 years of age or younger were included. The young reference group was used to interpret the group differences between the successful and average elderly.

Participants were excluded for the following reasons: problems with visual acuity, poor-quality structural T1 image preventing satisfactory normalization, not fulfilling pre-established performance criteria for the scanner task, and health-related issues, remarks from the radiologist screening the structural scans for abnormalities or outlier status across all voxels/clusters.

All included participants were in good general health, without major neurological impairments or diagnoses, and had normal, or corrected to normal, vision.

2.3 Study participants: Paper III

For Paper III the same samples as for Paper I were used. With the aim to study differences in longitudinal change of cognitive performance among individuals living alone versus living with someone, only individuals participating in two or more test waves and the variable "living arrangement" observed (at enrollment) were included (n=1552). Of these, 1213 individuals were living with someone and 339 living alone at enrollment. Many of the participants who dropped out were due to the event of death which further complicates analysis. In the event of death after a participant drops out, information regarding participant's date of death is available for all individuals including those who dropped out for other reasons.

2.4 Study participants: Paper IV

Participants included in Study IV were a subsample of 60 Betula participants (mainly from S1, although three subjects were from Sample 2). Longitudinal fMRI data from two test waves was available. Participants were originally recruited for a brain imaging study in 2002-2003. The initial objective was to study genetic variation in relation to brain function. For this reason, 50% of the participants were carriers of at least one APOE e4 allele. Of the 60 individuals who participated in the baseline study, 38 completed the follow-up study in 2008-2009. Reasons for nonparticipation included dementia, illness, death and other.

3 Brain Imaging - Functional MRI

Functional Magnetic Resonance Imaging (fMRI) is a noninvasive imaging technique that can be used to study neural activity in a subjects brain. The technique does not measure neural activity directly, instead, the vast majority of fMRI experiments measures the blood oxygen level-dependent (BOLD) signal, the ratio of oxygenated to deoxygenated hemoglobin (Ogawa et al., 1992).

In most fMRI research studies, subjects are given some task(-s) to perform inside the scanner. The goal is then to identify brain areas activated by the task under study. Commonly, fMRI experiments often use a so called block-design, where each session consists of a series of blocks. Within each block subjects perform a task of interest (e.g. a memory task) which is often alternated with blocks where the subjects rest. Each session runs for about 30-60 minutes creating an immense amount of data, and the sheer amount of data greatly contributes to the difficulties in data analysis. A typical fMRI scan of a whole-brain collects data in 30-40 slices, where each slice consists of a 64×64 array of voxels, dividing the brain into a set of cubes denoted voxels. The time required to collect all slices, called repetition time, typically ranges between 2-3 seconds if the whole brain is scanned.

3.1 Preprocessing

The preprocessing steps basically consist of transformations to prepare the data for the more interesting task-related analyses. The goal of preprocessing is to reduce the systematic variability in the data that comes directly from the scanner environment, and thus isolate or at least enhance the task-related signal. The major steps involved in fMRI preprocessing are briefly described below. For an extended overview see e.g. Ashby (2011, chapter 4) or Lazar (2008, chapter 3).

Usually *Slice-timing correction* is applied as a first step, since subsequent analysis steps usually assume that all voxels are measures simultaneously. The slice-timing correction accounts for the fact that slices are collected one at a time, rather than all at once, by shifting the observed time series for each voxel.

Even small head movements inside the scanner may create artificial changes in the BOLD signal, causing errors in latter statistical analyses. To correct for head-movements, each scan is aligned to a "reference" scan, e.g. the first image/scan. Most head motion algorithms assumes the head motion is "rigid body" motion, i.e. the head and the brain only change their position and orientation but not their shape.

There is, sometimes substantial, between-individual variation in the size and shape of the human brain, but also regularities shared among

individuals. To merge subjects in a group analysis, it is important that a specific voxel is assigned to the same brain structure for all individuals. The *Coregistration* procedure aligns the subjects structural and functional images, to improve spatial localization of the functional data , i.e the location of an activation is determined more precisely. *Normalization* is the process of registering each subjects images to a standardized space defined by a template, e.g. Talairach (Talairach and Tournoux, 1988) or Montreal Neurological Institute (MNI) atlas (Evans et al., 1993), further facilitating localization.

The next, and usually final, preprocessing step is *spatial smoothing*. The main advantages of smoothing are, first, to reduce noise, thus increasing the signal-to-noise ratio (unless the signal is very spatially narrow), and second, making the data distribution closer to the normal distribution to better satisfy the assumptions of many statistical models, perhaps most importantly those of Random Field Theory, for handling the multiple comparisons problem (further described below).

The most common approach is to spatially smooth the data using a Gaussian kernel function centred at the voxel "to be smoothed". The kernel is commonly characterized by the *full width half maximum* (FWHM), which is connected to the standard deviation σ of the Gaussian distribution by the expression $FWHM = \sigma \cdot 2\sqrt{2\log 2}$.

3.2 Data analysis

Statistical analysis of fMRI data involves working with data of massive size, complicated even further by the spatial and temporal dependence structure. The standard methods for group analysis uses two stages of analysis (since most modern fMRI studies have more than one subject, focus here is on group analysis). The goal with the first stage is a statistical parametric map for each subject and the second stage combines these maps into a single group map, to make group inference. Note that only brief outline of analysis is given here, for a more in-depth description see, e.g. Lazar (2008, chapter 5).

For the first-stage analysis, a time series regression model is often used for modeling the BOLD response observed over time in each voxel for each subject, individually.The subject-specific first-level model for a voxel is then written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is the timeseries of BOLD responses, $\mathbf{X}$ the design matrix, $\boldsymbol{\beta}$ a parameter vector, and $\boldsymbol{\varepsilon}$ the error term, normally, it is assumed that $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma^{\mathbf{2}})$. The design matrix, $\mathbf{X}$, is general and may contain any variable which is deemed of interest at subject level, e.g. a task/rest indicator. Another especially important regressor which is practically

always used in the first-stage model is the predicted hemodynamic response. Put briefly, it takes into consideration the fact that BOLD response does not start immediately upon presentation of a stimuli (the peak of the BOLD signal is usually seen 5-6 seconds after presentation), i.e. a change in experimental condition from baseline to task. The predicted hemodynamic response also accounts for an undershoot in the BOLD signal before returning to its baseline value.

The result of the first-stage analysis is an individual parametric map that specifies a statistic (e.g. a β-value) in every voxel.

The second-stage group analysis involves running a group-level analysis using the subject-specific values in each voxel, e.g. using Student's t-test, ANOVA, multiple regression, or linear mixed-effects models. The second-stage model can be written in a general way as:

$$\boldsymbol{\beta} = \mathbf{X}_G \boldsymbol{\beta}_G + \boldsymbol{\varepsilon}_G,$$

where $\mathbf{X_G}$ the second-level design matrix (e.g. indicators for cases and controls), $\boldsymbol{\beta}_G$ the second-level parameters and the error term is $\boldsymbol{\varepsilon}_G \sim N(0, \mathbf{I}\sigma_G^2)$. The result of the second-stage analysis is a group-level parametric map, again giving a statistic (at the group level, usually a t- or F-statistic) at every voxel.

3.3 The multiple comparisons problem

The result from a group analysis as described above is summarized in a statistical parametric map (SPM). The SPM:s are used for locating voxels who are e.g. activated by the experimental task across all groups or differentially activated by the task, between groups. Thus, hypothesis tests are carried out simultaneously for each voxel of the brain, typically hundreds of thousands, and it is crucial to correct for multiple comparisons.

A standard quantity to control is the *family-wise error rate* (FWER), the probability of having one or more false positives across all tests. The perhaps most common approach, considering all areas of applications and not just fMRI, for controlling the FWER is the "Bonferroni correction", where a p-value threshold of α/V, where V is the number of tests, is used instead of α in each individual test. The procedure becomes conservative if there is a correlation between adjacent voxels, which is expected here.

Another popular approach for controlling FWER in fMRI experiments is the Random Field Theory (RFT) approach (Worsley et al., 1996) where the spatial correlation between voxels is utilized when determining significance thresholds. Threshold given by RFT are generally lower (less conservative) than those given by the Bonferroni method if

the data is fairly smooth and sample sizes not-to-small (Nichols and Hayasaka, 2003). Following the development in computational power during recent years, permutation based methods (Nichols and Holmes, 2002) for FWER control are gaining popularity. The RFT and permutation approaches have in common that they take into consideration that true activation is likely to be spread over several contiguous voxels, since voxels are rather arbitrary divisions of the brain.

4 Missing Data in Longitudinal Studies

Participant attrition is one of the main methodological problems in longitudinal studies of cognitive aging. If participants who drops out of the study are systematically different from those who remain there is a potential risk of bias. In this dissertation, we consider a particular type of dropout, monotone dropout. Assume that the full-data response vector, for subject i possibly observed at timepoints $t = 1, \ldots, T$, is $Y_i = (Y_{i1}, \ldots, Y_{iT})^\intercal$. Then for monotone dropout, if Y_{it} is missing, then Y_{is} is missing for $s \geq t$.

Monotone patterns are common in longitudinal studies, because if a subject drops out of the study in a given time period, then his or her data will typically be missing in all subsequent time periods.

4.1 Missing data mechanisms

The mechanism that leads to missing data plays a crucial role in the choices of methods to analyse data that suffer from missing observations. Plausible, though untestable, assumptions about the mechanism are required from the data analyst to conduct analyses. Three main categories of missing data mechanisms are suggested, following the classification suggested by Rubin (1976); Little and Rubin (2002).

In general, for missing data, the response vector, Y_{it}, can be divided into two vectors, based on whether values are observed, Y_{it}^o, or missing, Y_{it}^m. Let M_i be a integer variable $(1, 2, ..., m)$ indicating the highest t for which the outcome is observed. Finally, let the conditional distribution of M given Y and covariates X, i.e. $f(M \mid Y, X)$, characterize the missing data mechanism.

If the missingness does not depend on the response vector, observed or unobserved, the missingness mechanism is called *missing completely at random* (MCAR),

$$f(M \mid Y, X) = f(M \mid X).$$

Some distinguishes between mechanisms that are completely random, $f(M \mid Y, X) = f(M)$, and mechanisms that depends on covariates,

$f(M \mid Y, X) = f(M \mid X)$. The latter is then called *covariate dependent missing.*

If the missingness depends on observed, but not unobserved, responses, the mechanism is called *missing at random* (MAR),

$$f(M \mid Y, X) = f(M \mid Y^o, X).$$

If the dropout mechanism depends on the unobserved responses,

$$f(M \mid Y, X) = f(M \mid Y^m, Y^o, X),$$

the mechanism is called *missing not at random* (MNAR). When the missing data are MNAR the mechanism is said to be non-ignorable. Then valid analysis can only be obtained by incorporating a model for the missingness mechanism.

4.2 Methods for missing data in longitudinal studies

The literature on missing data in longitudinal studies often covers *deletion methods*, discarding subjects or units from further analysis; *imputation methods*, where plausible values for the missing observations are filled in, or *model based approaches* (e.g., Little and Rubin, 2002; Schafer, 1997; Daniels and Hogan, 2008; Laird, 1988).

4.2.1 Deletion methods

Complete cases, or listwise deletion, is the most common approach in practice to handling missing data. The method omits those cases with missing observations (in the variables involved) and proceed with a, so called, complete cases analysis. All analyses are calculated with the same set of cases, which makes it possible to track individual changes over the entire study period. However, if missingness is not MCAR, analysis of these complete cases may be biased since they may not be representative of the population. Moreover, by excluding large number of individuals, the approach can lead to loss in power.

Another simple deletion approach is *available cases* analysis, where all available data are used. The main problem with this approach is that the parameter estimates will be based on different sets of cases, with different sample sizes and different standard errors.

In some non-MCAR situations, it is possible to reduce biases from a complete cases analysis by weighting remaining cases so that their distribution more closely resembles that of the full sample or population with respect to observed covariates. Weights are estimated from the probabilities of response, e.g. by logistic regression, see, e.g., Dufouil et al. (2000) for an application using weights to a longitudinal study of cognition.

4.2.2 Imputation methods

Another method of handling missing observations is to replace them by one or more plausible values, that is imputing them. Imputation methods are more efficient than deletion methods as there is no data loss, but are more difficult to implement and may distort relationships. Various imputation methods have been proposed in the literature (Rubin, 1987; Schafer, 1997; Raghunathan et al., 2001). The main difference between methods concerns the information being used for determining a plausible value to impute, e.g. population data, baseline data, the observed outcome and/or covariates. However, for missing repeated measurements data it is desirable to preserve the longitudinal properties of the data.

Single Imputation, replaces each missing observation with one plausible value. Once imputation is done, analysis is straightforward using available methods for complete data.

Methods using no subject-specific information are, e.g., mean or median imputation. These methods replaces missing observations with the mean/median of valid data for the variable in question. Because the same value is being substituted for each missing case, this method artificially reduces the variance of the variable in question, in addition to diminishing relationships with other variables.

"Observed outcome" methods are e.g. Baseline carried forward (BCF) and Last observation carried forward (LOCF). These methods replaces each missing outcome at follow-up with the same subjects baseline measure or last observed measure, respectively. This often tends to understate differences in estimated time-trends and is especially bad for outcomes that have high variation within a subject.

Hot-deck imputation, use information regarding the subject's observed outcome measures as well as observed covariates. For each missing subject, the method finds a subject with similar values on baseline measure and covariates and impute that subject's observed outcome.

Another method that uses both information on covariates and observed outcomes is regression imputation. The variable to impute, is regressed on available information of interest using all subjects that are fully observed. The value to impute is the predicted value from the regression model, with or without added noise. The error value is sampled from a $N(0, s^2)$, where s^2 is the residual variance from the regression model.

So far we have described imputation methods producing a single complete data set. Instead of replacing each missing value with a single value, Rubin's (1987) *multiple imputation* method, substitutes each missing value with a set of M values, typically M is $3 - 10$. The multiply imputed data sets represents the uncertainty about the missing data model. The imputed data sets are further analyzed, one by one, using

standard complete data methods. Resulting in M parameter estimates along with their estimated variances. Finally, results are combined to give a single point estimate and confidence intervals.

4.2.3 Mixture models

The linear mixed model (Laird and Ware, 1982) accounts for dependencies between repeated measurements on the same subject. Another significant feature is the allowance for unbalanced data. These features make it a natural candidate for modeling longitudinal cognitive data.

The model can be separated into two stages, where in the first stage group level associations are modeled as fixed effects, and in the second stage the within-subject dependencies are modeled as random effects.

More specifically, in the second stage one models subject-specific regression coefficients, e.g. individual intercept and slope. These individual longitudinal profiles are modeled to overcome problems with unbalanced data that occur with other multivariate regression techniques. The model considered is as follows,

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

where, $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iT})$ is the vector of longitudinal responses for subject i $(i = 1, ..., N)$. $\mathbf{X}_i$ is the corresponding matrix of covariates and $\boldsymbol{\beta}$ is the vector of fixed effect coefficients. $\mathbf{Z}_i$ is the random effects design matrix (e.g. intercept and time from enrollment) with subject-specific random effects $\mathbf{b}_i$. We assume $\mathbf{b}_i \sim N(\mathbf{0}, \Omega)$, and $\epsilon_i \sim N(\mathbf{0}, \sigma^2 I)$ are independently distributed.

4.2.4 Mixture models for non-ignorable dropout

The linear mixed model provides valid inference if attrition is MAR. However, when the missing data mechanism is MNAR, one needs to specify a model for the missingness, which is incorporated in the full data distribution $f(Y, M \mid X)$.

There are two general classes of mixture models that account for non-ignorable dropout in longitudinal studies, *selection models* and *pattern mixture models* (PMM), based on different factorizations of the joint distribution of the outcome and the missingness.

In selection models, the joint distribution is factorized as the product of the outcome model and the missingness model (Little, 1995; Daniels and Hogan, 2008; Diggle and Kenward, 1994)

$$f(\mathbf{Y}_i, M_i, \mathbf{b}_i \mid \mathbf{X}_i) \;=\; f(Y_i \mid \mathbf{b}_i, \mathbf{X}_i) f(\mathbf{b}_i \mid \mathbf{X}_i) f(M_i \mid \mathbf{Y}_i, \mathbf{b}_i, \mathbf{X}_i).$$

In the pattern mixture model, the joint distribution is instead modeled as a mixture over dropout patterns (Little, 1995; Hogan and Laird,

1997; Wu and Bailey, 1989)

$$f(\mathbf{Y}_i, M_i, \mathbf{b}_i \mid \mathbf{X}_i) \quad = \quad f(Y_i \mid \mathbf{b}_i, \mathbf{X}_i, M_i) f(\mathbf{b}_i \mid M_i, \mathbf{X}_i) f(M_i \mid \mathbf{X}_i).$$

For selection models, the main advantage is that they directly specify the response model, $f(\mathbf{Y}_i \mid \mathbf{b}_i, \mathbf{X}_i)$, in which the analyst is usually interested. For pattern mixture models these parameters are not directly available. On the other hand, for pattern mixture models the main advantage is the straightforward implementation using standard procedures for linear mixed models.

In this dissertation, we focus on pattern mixture models to handle non-ignorable dropout. To account for non-ignorable dropout in the PMM, a common approach is to stratify the population by the subject's last observed measurement wave, as for a monotone missing data pattern. After we stratify on dropout time we assume that missing data, within a pattern, are MAR. The conditional distribution of the outcome is further assumed multivariate normal such that

$$(Y_i \mid \mathbf{b}_i, \mathbf{X}_i, M_i = m) \sim N(\mathbf{X}_i \boldsymbol{\beta}^{(m)} + \mathbf{Z}_i \mathbf{b}_i, \sigma^2 I),$$

a modification of the linear mixed model described in previous section, where $\boldsymbol{\beta}^{(m)}$ is the vector of fixed effects for subjects with dropout pattern m, $m = 1, ..., M$. $\sigma^2 I$ is the covariance matrix. Assume also the subject-specific random effects

$$\mathbf{b}_i \mid M_i = m \sim N(\mathbf{0}, \Omega^{(m)}).$$

The last term, the marginal distribution of M_i follows a multinomial distribution with parameters $\pi_m = \Pr(M_i = m)$, where π_m represents the proportion of the population in dropout pattern m, such that $\sum_{m=1}^{M} \pi_m = 1$.

The quantities of interest in this model are usually the population average regression parameters (e.g., population intercept and slope) which are weighted averages of the pattern-specific parameters across dropout patterns, $\sum_{m=1}^{M} \pi_m \beta^{(m)}$.

4.3 Principal stratification - Truncation by death

In longitudinal studies of cognitive aging, a significant proportion of the participants in the older cohorts die. Depending on the research question to solve, truncation due to death could complicate analysis, since, e.g., memory performance is not defined after an individual dies. It does not always make sense to impute such missing information using the previously mentioned missing data methods, neither is a mixture modeling approach useful, since a linear mixed model implicitly extrapolate beyond time of dropout.

Principal stratification (e.g., Frangakis and Rubin, 2002; Frangakis et al., 2007) is a framework that can be used to define the causal effect of a treatment on an outcome when the outcome is truncated by death and not defined thereafter. The parameter of interest is then the "principal stratum causal effect", that is, the effect on cognition of a treatment among the subpopulation of subjects that would have survived (during the follow up period) irrespective of treatment. The principal stratum causal effect is defined as

$$E(Y(1) - Y(0) \mid S(1) = S(0) = 1),$$

where $Y(0)$ and $Y(1)$ are assumed to be each individual's two underlying memory scores (often called potential outcomes), and $S(0)$ and $S(1)$, are the two potential survival times (during the time of study) respectively for each treatment status $A = 0$ and $A = 1$. However, this parameter is not identified since we do not know which individuals in our study are from this subpopulation.

To overcome this problem (in the setting of a randomized treatment) Chiba and VanderWeele (2011) showed that the treatment effect estimated by considering only survivors, $E(Y \mid A = 1, S = 1) - E(Y \mid A = 0, S = 1)$, is conservative in the sense that the estimate is attenuated relative to the principal stratum causal effect of interest under the following two assumptions. First monotonicity: for all individuals, $S(0) \geq S(1)$, i.e. their survival time under treatment cannot be longer than without treatment. The second assumption states that: $E(Y(1) \mid A = 1, S = 1) \geq E(Y(1) \mid A = 0, S = 1)$. This assumption holds if the subpopulation that has survived with treatment is healthier overall than the subpopulation that has survived without treatment. This assumption is reasonable if we believe that the treatment is harmful.

The two assumptions guarantee that $E(Y(1) - Y(0) \mid S(1) = S(0) = 1) \leq E(Y \mid A = 1, S = 1) - E(Y \mid A = 0, S = 1)$. This result is useful to understand under which conditions the estimated effects based on survivors have some validity.

5 Summary of Papers

5.1 Paper I: Genetic and lifestyle predictors of 15-year longitudinal change in episodic memory

The objective of the first paper is to reveal distinct longitudinal trajectories in episodic memory over 15 years. And further to identify demographic, life-style, health-related, and genetic predictors of stability or decline in memory performance. For this purpose we use 1954 healthy

participants aged 35 to 85 at enrollment from the Betula prospective cohort study, see Section 2.1. In the first step, we analyze data using a pattern-mixture model that consider the impact of attrition over 2-4 longitudinal sessions. Participants are further classified into three groups, high-, average- or low performing, based on initial level and rate of change, in contrast to a attrition-corrected average development in their age cohort. In a second step, logistic regression is used to determine significant predictors of stability or decline relative to average change in episodic memory. The results show that, of 1558 individuals participating in two or more test waves, 18% were classified as high performing, 13% as low performing, and 68% show age-typical average change. More educated, physically active, females and those living with someone are more likely to be classified as high performing, as well as carriers of the met allele of the COMT-gene. Less educated, non labor-force active and male participants are more likely to be classified as low performing, and the APOE 4 allele is more frequent among low performing participants. To conclude, quantitative, attrition-corrected assessment of longitudinal changes in memory can reveal substantial heterogeneity in aging trajectories, and such heterogeneity is predicted by genetic and life-style factors.

5.2 Paper II: Brain characteristics of individuals resisting age-related cognitive decline over two decades

This paper is a continuation of paper I. Older participants from the Betula study were classified as successful or average based on initial level and rate of change in episodic memory scores across 15-20 years, using the method developed in Study I. Successful elderly (n=51) and age-matched average elderly underwent functional magnetic resonance imaging while performing an episodic memory face-name paired associated task. Results showed that successful older participants had higher BOLD-signal during encoding than average performing participants, in the bilateral prefrontal cortex, and the left hippocampus (HC). The HC activation of the young reference groups was higher than that of the average but not the successful older group. Task performance correlated with HC activation, thus likely contributing to the superior memory performance of successful older participants. The pattern might reflect individual differences present from young age. Further, the differences between the older groups could not be accounted for by differences in brain structure.

The results of this study suggest that a mechanism behind successful cognitive aging might be preservation of HC function combined with a high frontal responsivity. These findings highlight sources for hetero-

geneity in cognitive aging, and may hold useful information for cognitive intervention studies.

5.3 Paper III: Causal inference with longitudinal outcomes and non-ignorable dropout: Estimating the effect of living alone on cognitive decline

In this paper we develop a model to estimate the causal effect of living arrangement (living alone versus living with someone) on cognitive decline. For this purpose we use 1552 healthy participants aged 35 to 85 at baseline from the Betula study. Episodic memory function is measured every five years over a 15-year period. The model must both balance confounding variables between the two living arrangement groups, as well as account for non-ignorable attrition. This is achieved by combining propensity score matching with a pattern mixture model for longitudinal data.

A fully Bayesian approach allows us to convey the uncertainty in the estimation of the propensity score and subsequent matching in the inference of the causal effect of interest. The analysis conducted here adds to previous studies in the literature concerning the protective effect of living with someone, by proposing a modeling approach treating living arrangement as a causal agent.

The results of the current study for individuals 65 and older provide some evidence of a negative effect of living alone on episodic memory at enrollment and a negative effect on the rate-of-change. For middle aged individuals no effect of living arrangement on cognition was found.

5.4 Paper IV: Imputation of missing longitudinal fMRI data

In this paper we consider the problem with dropout in longitudinal fMRI (L-fMRI) studies. One disadvantage with standard available methods for analyzing fMRI data is the exclusion of subjects with missing information, which may result in biased estimates and loss of power. One approach to this problem is to impute the missing information to obtain a complete dataset. We propose to adapt available imputation methods to a setting with L-fMRI data, where some subjects had missing observations at follow-up. We apply the approaches to a real dataset, and further evaluate the methods in a simulation study.

The current study suggest that several imputation approaches described in this paper are feasible for L-fMRI analyses. Our results from the simulation study indicate that multiple imputation yielded an improvement of the validity of the results obtained from datasets with

missing values, while other simple imputation methods greatly increased the risk of finding significant activity changes in areas where there were no actual differences. Results from the analysis of the real data showed an overlap consisting of four regions, all found by a majority of the missing data methods including those who performed best in the simulation study. These findings suggests that imputation may greatly aid L-fMRI analyses.

6 Final Remarks and Further Research

In this thesis we have proposed methods to study age-related change in longitudinal studies. These methods were adapted to take into account the complex issues accompanying studies of cognitive aging, and specifically work out issues related to attrition.

We further developed a model to estimate the causal effect of a causal agent (living alone versus living with someone) on cognitive change in paper III. There are several aspects of our approach that allow for sensitivity analyses. This would be interesting to study in future work, e.g. sensitivity to the existence of unobserved confounders, or sensitivity to changes in slope after dropout.

Moreover, in paper IV, the objective was to adapt and implement available imputation methods to L-fMRI data, where some subjects are lost to follow-up. In this study we made rather strong assumptions on the missing data mechanism to be MCAR. Therefore, it would be of great interest to study how relaxations of these assumptions would effect results from imputing L-fMRI data.

References

Aldwin, C., A. I. Spiro, and C. L. Park (2006). *Health, behavior, and optimal aging: A lifespan developmental perspective* (6th ed.)., pp. 85–104. San Diego: Academic Press.

Ashby, F. G. (2011). *Statistical analysis of fMRI data.* Cambridge, MA: The MIT Press.

Bäckman, L., B. J. Small, and L. Fratiglioni (2001). Stability of the preclinical episodic memory deficit in alzheimer's disease. *Brain 124*, 96–102.

Bäckman, L., B. J. Small, Å. Wahlin, and M. Larsson (1999). Cognitive functioning in very old age. *Handbook of aging and cognition (2nd ed)*, 499–558.

Caracciolo, B., K. Palmer, R. Monastero, B. Winblad, L. Bäckman, and L. Fratiglioni (2008). Occurrence of cognitive impairment and dementia in the community: A 9-year-long prospective study. *Neurology 70*, 1778–1785.

Chiba, Y. and T. J. VanderWeele (2011). A simple method for principal strata effects when the outcome has been truncated due to death. *American journal of epidemiology 173*, 745–751.

Christensen, H., A. Mackinnon, A. Korten, A. Jorm, A. Henderson, P. Jacomb, and B. Rodgers (1999). An analysis of diversity in the cognitive performance of elderly community dwellers: individual differences in change scores as a function of age. *Psychology and aging 14*, 365.

Cooney, T. M., K. W. Schaie, and S. L. Willis (1988). The relationship between prior functioning on cognitive and personality dimensions and subject attrition in longitudinal research. *Journal of Gerontology 43*, P12–P17.

Daniels, M. J. and J. W. Hogan (2008). *Missing data in longitudinal studies: strategies for bayesian modeling and sensitivity analysis*. Boca Raton: Chapman and Hall/CRC.

Diggle, P. and M. G. Kenward (1994). Informative drop-out in longitudinal data analysis. *Applied statistics*, 49–93.

Dufouil, C., C. Brayne, and D. Clayton (2004). Analysis of longitudinal studies with death and drop-out: a case study. *Statistics in medicine 23*, 2215–2226.

Dufouil, C., D. Clayton, C. Brayne, L. Chi, T. Dening, E. Paykel, D. OConnor, A. Ahmed, M. McGee, and F. Huppert (2000). Population norms for the MMSE in the very old estimates based on longitudinal data. *Neurology 55*, 1609–1613.

Evans, A. C., D. L. Collins, S. Mills, E. Brown, R. Kelly, and T. M. Peters (1993). 3d statistical neuroanatomical models from 305 mri volumes. In *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.*, pp. 1813–1817. IEEE.

Frangakis, C. E. and D. B. Rubin (2002). Principal stratification in causal inference. *Biometrics 58*, 21–29.

Frangakis, C. E., D. B. Rubin, M.-W. An, and E. MacKenzie (2007). Principal stratification designs to estimate input data missing due to death. *Biometrics 63*, 641–649.

Gerstorf, D., N. Ram, C. Hoppmann, S. L. Willis, and K. W. Schaie (2011). Cohort differences in cognitive aging and terminal decline in the Seattle longitudinal study. *Developmental psychology 47*, 1026–1041.

Habib, R., L. Nyberg, and L.-G. Nilsson (2007). Cognitive and non-cognitive factors contributing to the longitudinal identification of successful older adults in the Betula study. *Aging, Neuropsychology, and Cognition 14*, 257–273.

Hogan, J. W. and N. M. Laird (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in medicine 16*, 259–272.

Jacqmin-Gadda, H., C. Fabrigoule, D. Commenges, and J.-F. Dartigues (1997). A 5-year longitudinal study of the mini-mental state examination in normal aging. *American Journal of Epidemiology 145*, 498–506.

Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in medicine 7*, 305–315.

Laird, N. M. and J. H. Ware (1982). Random-effects models for longitudinal data. *Biometrics*, 963–974.

Lazar, N. (2008). *The statistical analysis of functional MRI data*. New York, NY: Springer.

Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association 90*, 1112–1121.

Little, R. J. and D. B. Rubin (2002). *Statistical analysis with missing data* (second ed.). New York: Wiley.

Morris, M. C., D. A. Evans, L. E. Hebert, and J. L. Bienias (1999). Methodological issues in the study of cognitive decline. *American journal of epidemiology 149*, 789–793.

Nichols, T. and S. Hayasaka (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical methods in medical research 12*, 419–446.

Nichols, T. E. and A. P. Holmes (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping 15*, 1–25.

Nilsson, L.-G., R. Adolfsson, L. Bäckman, C. M. de Frias, B. Molander, and L. Nyberg (2004). Betula: A prospective cohort study on memory, health and aging. *Aging Neuropsychology and Cognition 11*, 134–148.

Nilsson, L.-G., L. Bäckman, K. Erngrund, L. Nyberg, R. Adolfsson, G. Bucht, S. Karlsson, M. Widing, and B. Winblad (1997). The Betula prospective cohort study: Memory, health, and aging. *Aging, Neuropsychology, and Cognition 4*, 1–32.

Ogawa, S., D. W. Tank, R. Menon, J. M. Ellermann, S.-G. Kim, H. Merkle, and K. Ugurbil (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences 89*, 5951–5955.

Palmer, K., L. Bäckman, B. Winblad, and L. Fratiglioni (2008). Mild cognitive impairment in the general population: occurrence and progression to alzheimer disease. *American Journal of Geriatric Psych 16*, 603–611.

Proust-Lima, C., J.-F. Dartigues, and H. Jacqmin-Gadda (2011). Misuse of the linear mixed model when evaluating risk factors of cognitive decline. *American journal of epidemiology 174*, 1077–1088.

Raghunathan, T. E., J. M. Lepkowski, J. Van Hoewyk, and P. Solenberger (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey methodology 27*, 85–96.

Rönnlund, M., L. Nyberg, L. Bäckman, L.-G. Nilsson, et al. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and aging 20*, 3–18.

Rowe, J. W. and R. L. Kahn (1987). Human aging: usual and successful. *Science 237*, 143–149.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*, 581–592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.

Schaie, K. W. (2012). *Developmental influences on adult intelligence: The Seattle longitudinal study*. Oxford University Press.

Schaie, K. W. et al. (1994). The course of adult intellectual development. *American psychologist 49*, 304–304.

Talairach, J. and P. Tournoux (1988). *Co-planar Stereotaxic Atlas of the Human Brain. 3-dimensional Proportional System: An Approach to Cerebral Imaging*. Thieme Medical Publishers, New York.

Verhaeghen, P. and T. A. Salthouse (1997). Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological bulletin 122*, 231.

Worsley, K. J., S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, A. C. Evans, et al. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping 4*, 58–73.

Wu, M. C. and K. R. Bailey (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, 939–955.

Yaffe, K., A. Fiocco, K. Lindquist, E. Vittinghoff, E. Simonsick, A. Newman, S. Satterfield, C. Rosano, S. Rubin, H. Ayonayon, et al. (2009). Predictors of maintaining cognitive function in older adults the health abc study. *Neurology 72*, 2029–2035.

Yaffe, K., K. Lindquist, E. Vittinghoff, D. Barnes, E. M. Simonsick, A. Newman, S. Satterfield, C. Rosano, S. M. Rubin, H. N. Ayonayon, et al. (2010). The effect of maintaining cognition on risk of disability and death. *Journal of the American Geriatrics Society 58*, 889–894.

Zeger, S. L., K.-Y. Liang, and P. S. Albert (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 1049–1060.