Mathematical Statistics
Stockholm University

Latent variable models for longitudinal twin data with dropout and death

Annica Dominicus

Research Report 2003:17
Licentiate thesis

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2003:17**,
http://www.math.su.se/matstat

Latent variable models for longitudinal twin data with dropout and death

Annica Dominicus*

November 2003

Abstract

Latent growth curve modelling is the commonly used method for analyzing longitudinal twin data. Estimation is carried out using maximum likelihood under the assumption of multivariate normal outcomes. We relate these models to a larger framework of latent variable models and discuss extensions that relate to dropout, due to death or for other reasons than death. The standard procedure for handling incomplete data in latent growth models is to use full information maximum likelihood, which gives consistent estimates if values are missing at random (MAR). We discuss the implications of this assumption for making inference about of the importance of genes for different features of the underlying longitudinal process. Methods for assessing the importance of genes are compared, and a new measure of heritability of change is proposed.

KEY WORDS: latent variable models; latent growth curve models; longitudinal twin data; dropout; full information maximum likelihood.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: annicad@math.su.se.

Acknowledgements

I wish to thank my supervisor Juni Palmgren for encouragement, guidance and support. Juni, your enthusiasm is a great source of inspiration. I also thank my co-supervisor Nancy Pedersen for support and for introducing me to twin research. The opportunity to work with data from the Swedish Twin Registry from the very beginning of my PhD studies has been very important for opening my eyes to what can be accomplished by statistical research motivated by questions of scientific interest. I also thank all colleagues at the Department of Mathematical Statistics, Stockholm University, and the Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, for enlightening my working life. Finally I thank Magnus for invaluable support and for being a big source of happiness in my life.

Stockholm, November 2003,
Annica Dominicus

Contents

1 Introduction

Human phenotypes are affected by genes and/or environmental factors. The first step in establishing the importance of genes involves studies based on family data. The idea is to compare the phenotypic resemblance for different kinds of relatives. In twin studies, assuming that the effect of environment shared by twins is the same for identical and fraternal twins, a larger phenotypic resemblance among identical twins compared to fraternal twins is interpreted as an effect of genes.

Longitudinal twin data makes it possible to answer questions that are not possible with cross-sectional data. It also introduces a need for new statistical tools. Approaches for modelling the dependency between observations in longitudinal data can be classified into three families of models: marginal models, transition models and random effects models (Diggle, Liang, and Zeger, 1994). In some situations hybrids of these models are the most appropriate. In marginal models the dependency is accounted for by introducing a correlation structure between observations. Transition models assume observations to depend on earlier observations, and random effects models include latent, unobserved, variables shared by observations from the same unit to account for the dependency. The three types of models address different aspects of the underlying process and the choice of model framework will depend on what assumptions that can be made about the underlying process and what questions that are of scientific interest.

For twin research focus is on testing hypothesis about the effect of genes (often unobserved) and environmental factors (observed or unobserved). Therefore, the natural choice is to base the analysis on a model including latent, unobserved, variables. Classical twin models for cross-sectional data belong to the domain of structural equation models (SEM) (e.g. Bollen, 1989). For longitudinal twin data, these models are extended to incorporate the dependency between repeated measurements from the same individual and are then referred to the domain of latent growth curve models (LGCM)(e.g. McArdle and Hamagami, 2003).

The models for twin data, cross-sectional as well as longitudinal, all belong to a larger framework of latent variable models. This framework include latent variable models for (multivariate) responses of mixed type, including continuous responses, counts, survival data, dichotomous, ordered and unordered categorical responses and rankings. Other examples include multilevel models, mixed (random effects) models, factor models and latent class models. All these models belong to the same family of generalized latent variable models (Muthén, 2002; Skrondal and Rabe-Hesketh, 2004). In this work we focus on models for multivariate continuous responses, but the fact that these belong to a larger model framework will be important when extending the classical twin models to incorporate new features.

We study the process of cognitive decline in late life based on a sample of twins from the Swedish Adoption/Twin Study of Aging (SATSA), which is a longitudinal study investigating the aging process (Pedersen, McClearn, Plomin, Nessleroade, Berg, and

de Faire, 1991). The participants in the SATSA are a population-based sample from the Swedish Twin Registry (Lichtenstein, deFaire, Floderus, Svartengren, Svedberg, and Pedersen, 2002; Pedersen, Lichtenstein, and Svedberg, 2002). The sample we use include 840 twins that have participated in at least one in-person testing, including a battery of cognitive tests, and who reached the age of 50 years during their participation in the study. The maximum number of repeated measurements for an individual is four, spanning over 13 years.

Longitudinal aspects of the data have been examined based on different growth models, such as linear growth models (e.g. Reynolds, Finkel, Gatz, and Pedersen, 2002a; Reynolds, Gatz, and Pedersen, 2002b) and quadratic growth models (e.g. Reynolds, Finkel, McArdle, Gatz, and Pedersen, In press). Selective dropout from the study, due to death or for other reasons than death may, however, introduce bias in these models (Pedersen, Ripatti, Berg, Reynolds, Hofer, Finkel, Gatz, and Palmgren, 2003). Indeed, for any longitudinal study, especially of processes in late life, we expect some participants to drop out from the study. The standard procedure is to ignore the dropout mechanism, and base the estimation on full information maximum likelihood (FIML). Using the notation of Little and Rubin (2002), estimates obtained in this way are consistent only if values are missing completely at random (MCAR) or missing at random (MAR). This study sets out to investigate whether dropout, possibly correlated within twin pairs, affects estimates of heritability of change in cognition.

In the SATSA sample of 840 individuals 130 died before the fourth testing occasion. It has been suggested that cognitive decline in old age is related to impeding death, and that the dropout is informative and should be considered in the longitudinal data analysis (Pedersen et al., 2003). Another issue is the fundamental difference between dropout due to death and dropout for other reasons than death where the individual remain alive after dropping out (Zhang and Rubin, in press). In the former situation, values are not "missing" since cognitive measures after death are not a meaningful concept.

In section 2 classical twin models are described and measures of the importance of genes, such as the concept of heritability, is introduced. The framework of latent variable models for multivariate normal outcomes, including comments on model estimation, identification and selection, is presented in section 3. The special case of latent growth curve models for longitudinal twin data is treated in section 4. Section 5 is devoted to the issue of dropout, including general missing data theory and a discussion of the specific problem of truncation due to death. Methods for modelling the dropout process, as well as the methodological issues that may appear, are described in the manuscript in the Appendix, with an application to dropout from the SATSA. Latent growth curve models are fitted to data from the SATSA in section 6. Some remarks about the area of longitudinal twin modelling, as well as plans for future research, is given in section 7.

This study provides a framework for investigating under what conditions selective dropout of twins can introduce bias in estimates of heritability of change. Further

analytic work and simulation studies are planned in the PhD thesis in order to give a more comprehensive answer to this question.

2 Biometrical models

Biometrical models are models that incorporate familial resemblance. The aim with such models is to reveal the relative importance of genes and environment for a quantitative trait. The simplest twin model for a univariate trait, measured at one time point, is first introduced, followed by a discussion of extensions to twin data with repeated measurements of a trait.

2.1 Cross-sectional twin models

Twin models are based on knowledge about the fraction of genes shared by identical (monozygotic or MZ) and fraternal (dizygotic or DZ) twins. Assuming that the environmental effect is the same for MZ and DZ twins, a larger similarity within MZ compared to DZ twin pairs is interpreted as an effect of genes. The general idea is to model the trait value for each twin as a function of unobserved, latent, genetic factors, as well as observed and unobserved environmental factors. Genetic factors can have an *additive effect* on a trait value, or show a *dominance deviation*, reflecting the extent to which the effect of alleles at a locus do not simply "add up" (e.g. Plomin, DeFries, McClearn, and McGuffin, 2001). We use the notation η_{Aj} for an additive genetic effect and η_{Dj} for a dominant genetic effect acting on a univariate trait, denoted Y_j for twin j ($j = 1, 2$). Unobserved environmental factors can be either shared within twin pairs, denoted η_{Cj}, or individual-specific, denoted η_{Ej}. The model for twin j is

$$Y_j = \boldsymbol{\beta}\mathbf{x}_j + \lambda_A\eta_{Aj} + \lambda_D\eta_{Dj} + \lambda_C\eta_{Cj} + \lambda_E\eta_{Ej}, \tag{1}$$

where $\mathbf{x}_j$ is a vector of observed covariates and $\boldsymbol{\beta}$ the parameter vector for fixed effects. The latent variables η_{Aj}, η_{Dj}, η_{Cj} and η_{Ej} are all unobserved and assumed to be independent and normally distributed with mean 0 and variances σ^2_A, σ^2_D, σ^2_C and σ^2_E, respectively. The parameters λ_A, λ_D, λ_C and λ_E are referred to as factor loadings or path coefficients.

Based on some assumptions for the genetic mechanism, such as random mating and Hardy-Weinberg equilibrium in the population, the within-pair correlation for genetic factors can be derived. It can be shown that $\rho(\eta_{A1}, \eta_{A2}) = 1$ and $\rho(\eta_{D1}, \eta_{D2}) = 1$ for MZ twins, and $\rho(\eta_{A1}, \eta_{A2}) = 0.5$ and $\rho(\eta_{D1}, \eta_{D2}) = 0.25$ for DZ twins (e.g. Plomin et al., 2001). For the shared environment $\rho(\eta_{C1}, \eta_{C2}) = 1$ for both MZ and DZ twins. Genetic and environmental factors are assumed to be independent.

The normality assumption is motivated by assuming effects to consist of contributions from several different genetic and environmental sources. If each contribution is small

and independent of other effects, the central limit theorem gives normally distributed traits (Lange, 1978). This model, which assumes that many genes are acting on the trait, and that there is no major effect of a specific gene, is called a polygenic model. Hindsberger (2001) give an overview of the assumptions underlying the classical twin model, and propose an approach for analyzing twin data where the normality assumption is relaxed.

The cross-sectional twin model can be parameterized as a variance component model, fixing the factor loadings, $\lambda_A, \lambda_D, \lambda_C$ and λ_E, to one and estimating the variances, σ_A^2, σ_D^2, σ_C^2 and σ_E^2, for the latent variables. Alternatively, the model can be parameterized as a path coefficient model, fixing the variances for the latent variables and estimating the factor loadings. If we parameterize model (1) as a variance component model, the distribution of Y_j is given by

$$\begin{aligned} E[Y_j] &= \boldsymbol{\beta}\mathbf{x}_j \\ \mathrm{Var}(Y_j) &= \sigma_A^2 + \sigma_D^2 + \sigma_C^2 + \sigma_E^2 \\ \mathrm{Cov}_{MZ}(Y_j) &= \sigma_A^2 + \sigma_D^2 + \sigma_C^2 \\ \mathrm{Cov}_{DZ}(Y_j) &= 0.5\sigma_A^2 + 0.25\sigma_D^2 + \sigma_C^2. \end{aligned}$$

The impact of genes versus environment in the general population can be measured by the *broad-sense heritability*, defined as the fraction of the phenotypic variance due to genetic factors:

$$h_B^2 = \frac{\sigma_A^2 + \sigma_D^2}{\sigma_A^2 + \sigma_D^2 + \sigma_C^2 + \sigma_E^2}.$$

It is not possible to estimate the four variance components in model (1) from only three equations. One solution is to fit several constrained models, and compare them based on fit indicies such as the Akaike information criterion (AIC) (Akaike, 1987). Often the dominant genetic effect is excluded and the genetic effect assumed to be solely additive. This model is referred to as the ACE model. For complex traits the additive model has proven adequate based on empirical evidence. The model can be formulated in terms of the model equations or graphically by the corresponding path diagram in Figure 1. Following the conventions of path diagrams, circles represent latent variables, rectangles represent observed measurements and arrows represent linear relations.

2.2 Longitudinal twin models

For cross-sectional data it is clear how to measure the importance of genes, using the heritability measure defined above. For longitudinal data we first need to specify the shape of individual trajectories and decide which aspects that might be affected by genes. Often interest focuses on the genetic effect acting on the phenotypic level and

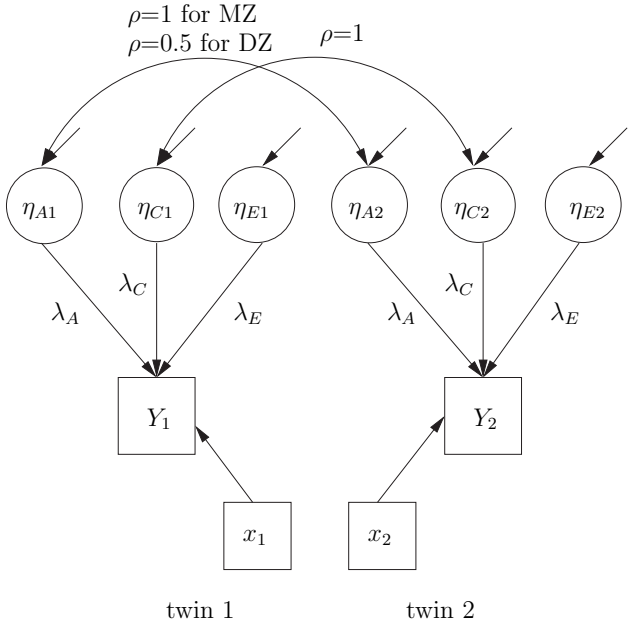

Figure 1: Path diagram for cross-sectional twin model for a univariate trait. The variable η_A is an unobserved additive genetic factor, η_C and η_E represent shared and non-shared environmental factors.

on the phenotypic change, respectively. The simplest model assumes the rate of change to be constant, but other models may be more appropriate. In section 4 different longitudinal models are discussed in more detail.

Conclusions about the importance of genes are usually based on the comparison of models with and without genetic factors included, using some measure of model goodness-of-fit. We propose a new measure of heritability of change and emphasize the need to explore its statistical properties.

3 Latent variable models

Latent growth curve models (LGCMs) typically used for analyzing longitudinal twin data belong to the larger framework of latent variable models (Muthén, 2002; Skrondal and Rabe-Hesketh, 2004). Models for various types of outcomes are included in this larger framework. Acknowledging LGCMs as a part of this larger framework will be important for model development, for data with non-normal outcomes, and possibly also for incorporation of the dropout process in the analysis of the longitudinal data.

The following presentation of the general model is restricted to models for multivariate normal outcomes. This is the scenario for quantitative traits where several genes and environmental factors act, and each of them has a small effect on the trait.

3.1 General model formulation

A general formulation of a latent variable model, which extends the basic twin model in (1), is

$$\begin{aligned} \mathbf{Y}_i &= \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\beta}\mathbf{x}_i + \boldsymbol{\epsilon}_i & (2) \\ \boldsymbol{\eta}_i &= \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i, & (3) \end{aligned}$$

where $\mathbf{Y}_i$ is a p-dimensional vector of outcomes, $\boldsymbol{\eta}_i$ an m-dimensional vector of latent random variables, $\mathbf{x}_i$ a q-dimensional vector of fixed and known covariates and $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\zeta}_i$ vectors of random errors, of dimension p and m respectively, with mean zero. The index i represent units. For twin data, twin pairs form the units. $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of measurement slopes or factor loadings, and $\boldsymbol{\beta}$ is a $p \times q$ parameter matrix of regression slopes. $\boldsymbol{\beta}$ corresponds to the direct influence of x variables on y variables. $\mathbf{B}$ is an $m \times m$ parameter matrix of slopes for regression of latent variables on other latent variables. $\mathbf{B}$ has zero diagonal elements and it is assumed that $\mathbf{I} - \mathbf{B}$ is non-singular. Furthermore, $\boldsymbol{\Gamma}$ is an $m \times q$ slope parameter matrix for regression of the latent variables on the known covariates. The first equation (2), which specifies the model for the vector of measured outcomes, is called the measurement model. The second equation (3) is called the structural model and specifies the model structure for the latent variables. Inserting the structural equation into the measurement model, the model for the vector of outcomes for unit i can be expressed as

$$\mathbf{Y}_i = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\beta}\mathbf{x}_i + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}_i + \boldsymbol{\epsilon}_i.$$

The mean vector $\boldsymbol{\mu}_i$ and variance-covariance matrix $\boldsymbol{\Sigma}_i$ for $\mathbf{Y}_i$ are thus

$$\begin{aligned} \boldsymbol{\mu}_i &= \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\beta}\mathbf{x}_i & (4) \\ \boldsymbol{\Sigma}_i &= \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}((\mathbf{I} - \mathbf{B})^{-1})^T\boldsymbol{\Lambda}^T + \mathbf{S}, & (5) \end{aligned}$$

where $\boldsymbol{\Psi} = \mathrm{Var}(\boldsymbol{\zeta}_i)$ is the variance-covariance matrix for the disturbance errors for latent variables and $\mathbf{S} = \mathrm{Var}(\boldsymbol{\epsilon}_i)$ is the variance-covariance matrix for the error terms for observed outcomes. If we set $\mathbf{S}$ to be a diagonal matrix this corresponds to assuming observed outcomes to be conditionally independent given the latent variables. The parameters to be estimated are the parameters appearing in $\boldsymbol{\Lambda}$, $\boldsymbol{\beta}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}$ and $\mathbf{S}$, which we jointly denote $\boldsymbol{\theta}$. The type of model investigated determines which parameters that are fixed and which are free to be estimated.

In some situations the matrices $\boldsymbol{\Lambda}$, $\boldsymbol{\beta}$, $\mathbf{B}$ and $\boldsymbol{\Gamma}$ will be different for different units. For example, in latent growth curve models the measurement scores in $\boldsymbol{\Lambda}$ will depend on the time points at which outcomes were observed, and can be different for different units if the data is unbalanced. In this case an index i is added to the matrices.

3.2 Implicit assumption about variance structure

From the expression for the variance-covariance matrix, $\mathbf{\Sigma}_i$, in (5) it is clear that this latent variable model imposes a specific structure on the variances and covariances for the outcomes. In latent growth modelling inclusion of a latent random slope on some time scale is common. For example, the most simple random slope model has the form

$$Y_{jt} = \beta_0 + \beta_1 t + \eta_{0j} + \eta_{1j} t + \epsilon_{jt},$$

where Y_{jt} is the trait value for individual j at time t. The parameters β_0 and β_1 are the population mean intercept and slope and η_{0j} and η_{1j} represent the random intercept and random slope for individual j, measured as a deviation from the mean. ϵ_{jt} is an error term. In the general formulation (2) this model corresponds to setting all the elements in $\mathbf{B}$ and $\mathbf{\Gamma}$ equal to zero. $\mathbf{\Lambda}$ is a matrix with a column of ones and a column of t's. Assuming the vector of latent variables $\boldsymbol{\eta}_j = (\eta_{0j}, \eta_{1j})^T$ to follow a multivariate normal distribution $N(0, \mathbf{\Psi})$ and the error terms ϵ_{jt} to be independent and normally distributed $N(0, \sigma^2)$ and independent of $\boldsymbol{\eta}_j$, the vector of trait values $\mathbf{Y}_j$ follow a multivariate normal distribution with mean $\boldsymbol{\mu}_j$ and variance-covariance matrix $\mathbf{\Sigma}_j$ with elements

$$\begin{aligned}
\mu_{jt} &= E[Y_{jt}] = \beta_0 + \beta_1 t \\
\Sigma_{j(tt)} &= \mathrm{Var}(Y_{jt}) = \psi_{11} + 2\psi_{12} t + \psi_{22} t^2 + \sigma^2 \\
\Sigma_{j(tt')} &= \mathrm{Cov}(Y_{jt}, Y_{jt'}) = \psi_{11} + 2\psi_{12}(t + t') + \psi_{22} t t'.
\end{aligned}$$

The time t appears in the expression for the variance, which implies a heteroscedastic model for the variance. In the discussion of different shapes of individual trajectories in latent growth models this is important to bear in mind.

3.3 Maximizing the likelihood

Estimates of the parameter vector $\boldsymbol{\theta}$ can be obtained by maximum likelihood estimation based on the model

$$\mathbf{Y}_i = N(\boldsymbol{\mu}_i(\boldsymbol{\theta}), \mathbf{\Sigma}_i(\boldsymbol{\theta})),$$

where $\mathbf{Y}_i$ is the outcome for unit i. If there are n observed units, i.e. $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_n)$, the log-likelihood is

$$l(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} -\frac{1}{2} \ln |\mathbf{\Sigma}_i| - \frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \mathbf{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i), \tag{6}$$

where the mean vector $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta})$ and the variance-covariance matrix $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ have the form given by (4) and (5). The maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ is obtained by maximizing (6) with respect to the parameters in $\boldsymbol{\theta}$. In general, the likelihood function is a complicated non-linear function of the parameters and no explicit solutions exist. Instead, an iterative numerical procedure is necessary to obtain the MLEs. The MLEs are consistent and asymptotically unbiased, although they may be biased in small samples.

For some special cases of the general model, methods have been derived to obtain unbiased estimates of the variance components in $\boldsymbol{\theta}$. For the linear mixed model, assuming no covariate effects on the latent variables, the general model reduces to $\mathbf{Y}_i = N(\boldsymbol{\beta}\mathbf{x}_i, \boldsymbol{\Sigma}_i(\boldsymbol{\alpha}))$, where $\boldsymbol{\alpha}$ correspond to the variance components, and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$. Unbiased estimates of the variance components in this model are obtained using restricted maximum likelihood (REML) (e.g. Verbeke and Molenberghs, 2000). This procedure takes into account the uncertainty due to the estimation of fixed effects, $\boldsymbol{\beta}$, in the estimation of the variance components $\boldsymbol{\alpha}$.

In the general model with mean and variance-covariance structure given in (4) and (5), some of the parameters in $\boldsymbol{\theta}$ appear both in the mean and variance-covariance expression, and restrictions of the likelihood to obtain unbiased estimates of variance components has not been discussed much in the literature. However, the difference between estimates obtained from ML and REML should not be large if number of units is large and number of fixed effects limited.

3.4 Model identification

The problem of identification has two different components: parameter identification, referring to the ability of the observed data to render unique parameter estimates for a given model, and model equivalence, referring to the fact that several different parameterizations may yield the same response distribution, even when the observed data fits the model perfectly (Skrondal, 1996). For the latter problem we have to rely on substantive theory, in choosing between models. The following discussion of identification concerns parameter identification.

There are two types of identification: *global* and *local* identification (e.g. Bollen, 1989). A parameter vector $\boldsymbol{\theta}$ is globally identified if there are no vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ such that $f(\mathbf{Y}|\mathbf{x};\boldsymbol{\theta}_1) = f(\mathbf{Y}|\mathbf{x};\boldsymbol{\theta}_2)$ unless $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. Local identification is a weaker concept of uniqueness: a parameter vector $\boldsymbol{\theta}$ is locally identified at a point $\boldsymbol{\theta}_1$, if, in the neighborhood of $\boldsymbol{\theta}_1$, there is no vector $\boldsymbol{\theta}_2$ for which $f(\mathbf{Y}|\mathbf{x};\boldsymbol{\theta}_1) = f(\mathbf{Y}|\mathbf{x};\boldsymbol{\theta}_2)$ unless $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. Global identification trivially implies local identification, and local identification is necessary but not sufficient for global identification.

One way to establish identification is analytically. For example, assuming the outcome to follow a multivariate normal distribution, i.e. $\mathbf{Y} \sim N(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, each element of $\boldsymbol{\theta}$ must be solved for in terms of one or more elements of the observed mean vector

$\boldsymbol{\mu}$ and the observed variance-covariance matrix $\boldsymbol{\Sigma}$. However, with even moderately complex models, solving the set of nonlinear equations for the unknown parameters can prove virtually impossible. For some special cases analytic checks of the model has been suggested, but there are no conditions that are both *necessary* and *sufficient* for the general model (Bollen, 1989).

Given this situation, researchers often turn to empirical tests of identification. The empirical procedures test local identification. Although establishing global identification is preferable, checks on local identification is still useful. It provides a means to detect some models that are not globally nor locally identified, since failure to achieve local identification tells us that global identification also fails.

One approach for empirical identification is to check the information matrix, defined as minus the expected value of second-order partial derivatives of the log-likelihood. The parameter $\boldsymbol{\theta}$ is locally identified at some point $\boldsymbol{\theta}_1$, if and only if the inverse of the information matrix exists, i.e. if the information matrix is non-singular (Rothenberg, 1971). However, there are sources of uncertainty to this test, since we need to rely on numerical means to evaluate the singularity of the information matrix. Further, the local identification is evaluated at the estimated value $\hat{\boldsymbol{\theta}}$, rather than the true population parameter value $\boldsymbol{\theta}$. Other empirical tests exist, such as starting the model estimation from different positions (starting values) in the parameter space to see if it converges to the same parameter estimates each time.

3.5 Model selection

For twin data, interest is primarily in the variance components, and inference about variance components is typically based on model comparisons. There are several options for choosing fit criterion (e.g. Tanaka, 1993; Skrondal, 1996). One option for comparing nested models is to perform conventional likelihood-ratio testing. For two models $M1$ and $M2$, where $M2$ is nested within $M1$, with log-likelihoods $l(\hat{\boldsymbol{\theta}}_{M1}|\mathbf{y};\mathbf{x})$ and $l(\hat{\boldsymbol{\theta}}_{M2}|\mathbf{y};\mathbf{x})$ respectively, the test statistic is

$$D = 2\left(l(\hat{\boldsymbol{\theta}}_{M1}|\mathbf{y};\mathbf{x}) - l(\hat{\boldsymbol{\theta}}_{M2}|\mathbf{y};\mathbf{x})\right).$$

Under the restricted model $M2$, this test statistic is asymptotically χ^2-distributed with *df* degrees of freedom, where *df* is the difference in number of parameters in the two models compared (e.g. Pawitan, 2001).

One of the regularity conditions under which the chi-squared approximation is valid is that the parameter values in the restricted model $M2$ are not on the boundary of the parameter space. When testing the significance of variance components, we typically fit a restricted model where the variance component is set to zero, and since variance components can not be negative the null value is on the boundary. In these situations the likelihood ratio test statistic does not follow the simple χ^2-distribution above (e.g.

Verbeke and Molenberghs, 2000).

Another model selection criterion is the Akaike information criterion (AIC), defined as $AIC = -2l(\hat{\boldsymbol{\theta}}|\mathbf{y};\mathbf{x}) + 2r$, where $\hat{\boldsymbol{\theta}}$ is the MLE and r the dimension of $\boldsymbol{\theta}$ (Akaike, 1987). Using the AIC, the log-likelihood of a model is penalized by the number of parameters in the model, which makes for a fairer comparison between models. Models with smaller AIC are preferred. AIC also allows for comparison of non-nested models. Alternative measures of goodness-of-fit have been suggested, e.g. so called goodness of fit indices (GFI) and measures of approximate fit such as the "root mean square error of approximation" (RMSEA) (e.g. Bollen, 1989).

Approximate standard errors for the parameter estimates are obtained from the inverse of the information matrix, and hence performing a Wald test is an option for making inferences provided the parameter estimates follow a normal distribution. The normal approximation fails completely fails, however, if the parameter is on the boundary of the parameter space.

4 Latent growth models for longitudinal twin data

The first step in latent growth curve modelling of longitudinal twin data is to formulate growth models that correspond to the shape of individual trajectories, which also correspond to aspects of growth that are of interest and make biologically sense. In this section a few growth models are discussed, and details of the linear model presented.

4.1 Shape of individual trajectories

In the analysis of longitudinal data one major task is to formulate the longitudinal shape of the individual trajectories. For twin data, the goal is to make inference about the genetic importance for the longitudinal process, and we want the model to correspond to a structure that we believe exists in reality and which is meaningful in this context. If little is known about the longitudinal process, the first step is to explore models for the shape by comparing different latent growth models. At this stage, the structure of within pair covariances is left completely unspecified, allowing the correlation within pairs to be different for MZ and DZ twins.

The simplest model is the linear model, which assumes the rate of change to be constant over time. This model includes individual-specific random intercepts and random slopes. For many traits the linear model is too simplistic. One possibility is to extend it by including a random quadratic term. However, it is not clear how to biologically interpret an estimated genetic or environmental effect on individual-specific quadratic terms.

A plausible alternative model for the example considered here, a study of cognitive decline in old age, is a hidden change-point model where trajectories are allowed to

be segmented and have different shapes before and after an individual-specific change-point (Slate and Turnbull, 2000). This model would allow for testing of the hypothesis that genes regulate the age for *terminal drop* (terminal decline) in cognition. According to the hypothesis of terminal drop, most people maintain quite stable into old age, and a more marked decline is an indication of impending death (e.g. Berg, 1996). The hidden change-point model would also allow for estimation of the variability in the number of years before death that terminal drop occurs. Ignoring the biometrical structure for twins, this model belongs to the domain of non-linear mixed effects model (Lindstrom and Bates, 1990). We now proceed to discuss the linear growth curve model in detail and return to the change-point model in section 7, where plans for future research are discussed.

4.2 The linear ACE model: a "Cholesky model"

When the shape of individual trajectories is established the full covariance structure can be specified. One model that has been widely used is the linear growth curve model (e.g. Reynolds et al., 2002a,b, In press; McArdle, 1986; McArdle and Hamagami, 2003) assuming an additive genetic effect as well as shared and unshared environmental factors to affect the random individual-specific level and slope. We refer to this model as the linear ACE model. The model equations for twin j in pair i at time t is

$$Y_{ijt} = \eta_{Iij} + \text{age}_{ijt}\eta_{Sij} + \epsilon_{ijt} \tag{7}$$

$$\eta_{Iij} = \boldsymbol{\gamma}_I \mathbf{x}_{Iij} + \lambda_{A1}\eta_{AIij} + \lambda_{C1}\eta_{CIij} + \lambda_{E1}\eta_{EIij} \tag{8}$$

$$\begin{aligned}\eta_{Sij} = {} & \boldsymbol{\gamma}_S \mathbf{x}_{Sij} + \lambda_{A2}\eta_{AIij} + \lambda_{C2}\eta_{CIij} + \lambda_{E2}\eta_{EIij} + \\ & + \lambda_{A3}\eta_{ASij} + \lambda_{C3}\eta_{CSij} + \lambda_{E3}\eta_{ESij},\end{aligned} \tag{9}$$

where η_{Iij} and η_{Sij} are the individual-specific random intercept and random slope respectively. The vectors $\mathbf{x}_{Iij}$ and $\mathbf{x}_{Sij}$ in (8) and (9) denote vectors of observed covariates that affect intercept and slope, with $\boldsymbol{\gamma}_I$ and $\boldsymbol{\gamma}_S$ the corresponding parameter vectors. This model is specified so that observed covariates are assumed to act on the underlying (unobserved) latent process and not directly on the observed outcome Y_{ijt}. Direct covariate effects on the outcome could be incorporated.

In model (7) the time scores to define the time scale for the slope has been expressed in terms of age, incorporating staggered entry into the study, with individuals entering at different ages. It also allows for unbalanced data, where the time distance between repeated measures are different for different individuals.

The latent variables η_{AIij}, η_{CIij} and η_{EIij} correspond to additive genetic and environmental (shared and non-shared) factors assumed to act on both the phenotypic level and on the rate of change. The latent variables η_{ASij}, η_{CSij} and η_{ESij} denote factors which only load on the rate of change. This parametrization is sometimes called a "Cholesky model", referring to the specific way of introducing correlation between

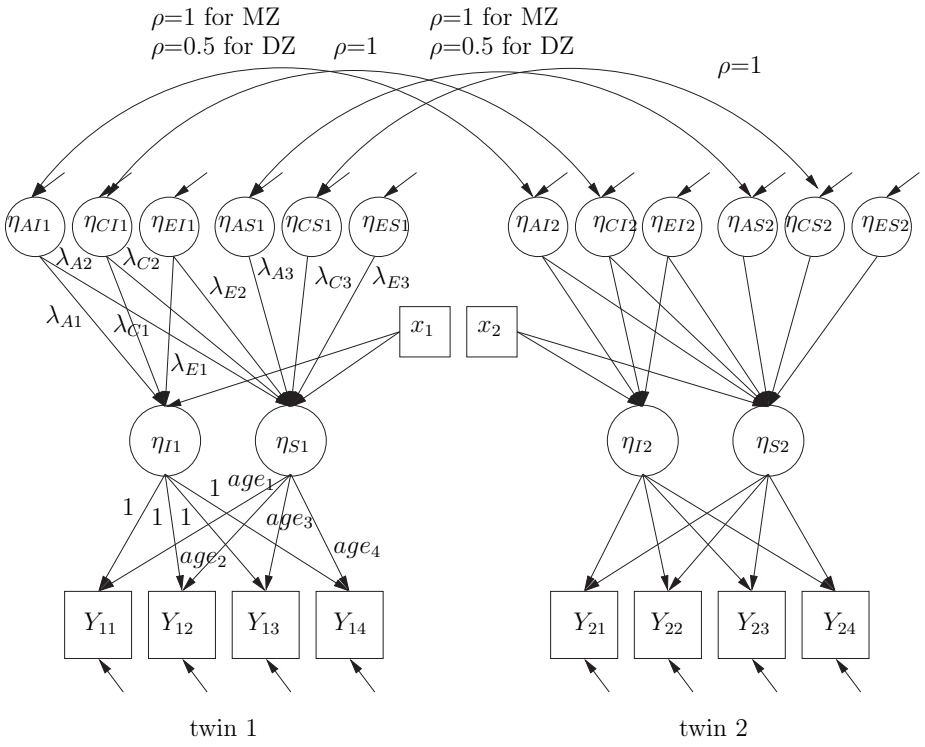

Figure 2: Path diagram for linear latent growth model for twins, assuming an additive genetic effect, as well as shared and non-shared environmental effects on level and slope.

variables by having several sets of latent variables (Loehlin, 1996). In this model the variances for the latent genetic factors, η_{AI} and η_{AS}, as well as the latent environmental factors, η_{CI}, η_{CS}, η_{EI} and η_{ES}, are all set equal to one. Instead, the factor loadings, λ_{A1}, λ_{C1}, λ_{E1}, λ_{A2}, λ_{C2}, λ_{E2}, λ_{A3}, λ_{C3} and λ_{E3}, are estimated to assess whether genetic or environmental factors are the most important for the trait level and the rate of change.

The linear ACE model specified by the model equations in (7), (8) and (9), correspond to the path diagram in Figure 2. In the picture the information available about correlations between genetic factors for twins, which is equal to 1 for MZ and equal to 0.5 for DZ twins, have been included.

In this model the correlation between level and slope enters only through the latent variables η_{AIij}, η_{CIij} and η_{EIij}, which load on both the level and the slope. This means that at least one of the factor loadings λ_{A2}, λ_{C2} and λ_{E2} will be different from zero if there is a correlation between level and slope. From this we could draw the conclusion that some of the factors that affect level and slope are correlated. However, it should not be over-interpreted as if it is the *same* genes or the *same* environmental factors.

Often, conclusions about the genetic importance for aspects of the growth are based on comparisons between model (7), and models where one of the genetic factors, η_{AI} or η_{AS}, are excluded. To test whether there is a genetic effect on the rate of change the

genetic factor only acting on the slope, η_{AS} is excluded. Also, the factor loading from the genetic factor η_{AI} to the slope, i.e. λ_{A2}, is set to zero.

As an alternative to model comparisons one could measure the importance of genes in terms of functions of the variance components. One measure that has been used is the *trajectory of heritability*, which expresses the fraction of phenotypic variance attributable to genetic sources at each age, e.g. (e.g. McArdle and Hamagami, 2003). We propose an other measure, the *heritability of change*. In the linear model this corresponds to the importance of genetic sources for the rate of change, which is assumed to be constant over time. We define the heritability of change as the fraction of variability in the random slopes, $\mathrm{Var}(\eta_S)$, attributable to genetic factors

$$h_S^2 = \frac{\lambda_{A2}^2 + \lambda_{A3}^2}{\lambda_{A2}^2 + \lambda_{A3}^2 + \lambda_{C2}^2 + \lambda_{C3}^2 + \lambda_{E2}^2 + \lambda_{E3}^2}. \tag{10}$$

We return to the concept of heritability of change in section 7 and discuss means of assessing distributional properties of h_S^2.

5 Missing values in longitudinal twin studies

To clarify the assumptions underlying the model estimation procedure for incomplete longitudinal twin data, using full information maximum likelihood, general missing data theory is first given as a background. The difference between dropout due to death and dropout for other reasons than death is pointed out. We discuss dropout mechanisms in the twin setting, and elaborate on situations in which dropout can be ignored. In this section methods for exploring the dropout process are briefly mentioned, and we refer to the manuscript in the Appendix for a more detailed description and an application to a sample from the Swedish Adoption/Twin Study of Aging (SATSA).

5.1 General missing data theory

A general treatment of statistical analysis of data with missing values is given by Little and Rubin (2002) where a useful hierarchy of missing value mechanisms are introduced. Let $\mathbf{Y}$ denote the data matrix. Further, denote the observed part of $\mathbf{Y}$ by $\mathbf{Y}_{obs}$, and the missing part by $\mathbf{Y}_{mis}$, so that $\mathbf{Y} = (\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$. An informal description of values being *missing at random* (MAR) is that the probability that a value is missing may depend on $\mathbf{Y}_{obs}$ but not on $\mathbf{Y}_{mis}$. Despite the name, MAR does not suggest that the missing values are a simple random sample of all values. The latter condition is known as *missing completely at random* (MCAR). MAR is less restrictive than MCAR because it requires only that the missing values behave like a random sample of all values within subclasses defined by observed data.

Let $\mathbf{R}$ be a matrix of indicator variables whose elements are zero or one depending

on whether the corresponding elements of $\mathbf{Y}$ are missing or observed. The probability model for $\mathbf{R}$, denoted by $P(\mathbf{R}|\mathbf{Y};\boldsymbol{\xi})$, depends on $\mathbf{Y}$ as well as some unknown parameters $\boldsymbol{\xi}$. More formally, MAR can be described in terms of the probability model for $\mathbf{R}$:

$$P(\mathbf{R}|\mathbf{Y};\boldsymbol{\xi}) = P(\mathbf{R}|\mathbf{Y}_{obs};\boldsymbol{\xi}). \tag{11}$$

If some of the data are missing the "observed data" truly consist not only of $\mathbf{Y}_{obs}$ but also of $\mathbf{R}$ and the probability distribution of the observed data is given by

$$\begin{aligned} P(\mathbf{R},\mathbf{Y}_{obs}|\boldsymbol{\theta},\boldsymbol{\xi}) &= \int P(\mathbf{R},\mathbf{Y}|\boldsymbol{\theta},\boldsymbol{\xi})d\mathbf{Y}_{mis} \\ &= \int P(\mathbf{R}|\mathbf{Y},\boldsymbol{\xi})P(\mathbf{Y}|\boldsymbol{\theta})d\mathbf{Y}_{mis}. \end{aligned}$$

When values are MAR, i.e. when condition (11) holds, the probability distribution for the observed data is

$$\begin{aligned} P(\mathbf{R},\mathbf{Y}_{obs}|\boldsymbol{\theta},\boldsymbol{\xi}) &= P(\mathbf{R}|\mathbf{Y}_{obs};\boldsymbol{\xi}) \int P(\mathbf{Y}|\boldsymbol{\theta})d\mathbf{Y}_{mis} \\ &= P(\mathbf{R}|\mathbf{Y}_{obs};\boldsymbol{\xi})P(\mathbf{Y}_{obs}|\boldsymbol{\theta}). \end{aligned}$$

If the measurement parameters, $\boldsymbol{\theta}$, and the parameters for the missing value model, $\boldsymbol{\xi}$, are *distinct*, which means that the joint parameter space of $(\boldsymbol{\theta},\boldsymbol{\xi})$ is the product of the individual parameter spaces for $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, estimation of $\boldsymbol{\theta}$ can be based on what Little and Rubin (2002) refers to as the likelihood ignoring the missing data mechanism, $L(\boldsymbol{\theta}|\mathbf{Y}_{obs}) \propto P(\mathbf{Y}_{obs}|\boldsymbol{\theta})$. Note that $\boldsymbol{\theta}$ refers to the parameters of the model for the complete data $\mathbf{Y} = (\mathbf{Y}_{obs},\mathbf{Y}_{mis})$, not the parameters for the distribution of $\mathbf{Y}_{obs}$ alone.

If data is MAR and the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ are distinct, the missing value mechanism is said to be *ignorable* (Little and Rubin, 2002). If this is not the case, the missing value mechanism is said to be *non-ignorable* or *informative*.

5.2 Full information maximum likelihood

Assuming that the complete-data model, $P(\mathbf{Y}_{obs}|\boldsymbol{\theta})$, is correct and that the missing value mechanism is ignorable, all relevant statistical information about the parameters $\boldsymbol{\theta}$ is contained in the observed-data likelihood $L(\boldsymbol{\theta}|\mathbf{Y}_{obs})$. This tends to be a complicated function of $\boldsymbol{\theta}$ and special computational tools, such as the EM algorithm (Dempster et al., 1977), are needed. We index the unique missing data patterns that appear in the sample by s, $s = 1, 2, ..., S$, and let $I(s)$ denote the subsets of the twin pairs that exhibit pattern s. Assuming that the observed data vector $\mathbf{y}_i$ for each pair i $(i = 1, 2, ..., n)$ is a sample from a multivariate normal distribution the observed-data log-likelihood is

$$l(\boldsymbol{\theta}|\mathbf{Y}_{obs}) = \sum_{s=1}^{S} \sum_{i \in I(s)} -\frac{1}{2} \ln |\boldsymbol{\Sigma}_i| - \frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i),$$

where $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta})$ is the mean vector and $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ the variance-covariance matrix for twin pair i. These vectors and matrices will be of different dimensions if some values are missing for some twin pairs. The procedure for estimating the parameters $\boldsymbol{\theta}$ based on maximization of this log-likelihood is referred to as full information maximum likelihood (FIML) or raw maximum likelihood. The EM algorithm needed to do this is described in Little and Rubin (2002). The procedure gives efficient and consistent estimators as long as the missing value mechanism is ignorable.

5.3 Ignorable missingness?

To establish whether the missing value process is ignorable we need to know whether values are MAR and if the parameters for the longitudinal process, $\boldsymbol{\theta}$, and the parameters for the missing value process, $\boldsymbol{\xi}$, are distinct. However, it is not possible to test whether these assumptions hold, since non-ignorable missingness by definition means that the missing value process can not be explained (or explored) by the observed part of the data. Hence, the options available are to either model the missing value mechanism jointly with the longitudinal process, do a sensitivity analysis on how much we can expect missingness to affect the results based on the observed data, or argue why the missing value mechanism should be ignorable based on substantive knowledge about the longitudinal as well as the missing value mechanism.

In the discussion of ignorability in the setting of longitudinal twin data, we need to formulate the processes we believe are acting, and the nature of model parameters. Here, both parameters corresponding to genetic effects, as well as parameters for environmental effects, are incorporated in the model, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_{genes}, \boldsymbol{\theta}_{env})$.

In the SATSA there are both intermittent missing values and dropouts. We believe that intermittent missing values are indeed ignorable, but suspect that this might not be the case for dropout. The main reasons for dropout is death or that individuals develop dementia, in which case values will be censored. It has been shown that both longevity (Yashin and Iachine, 1995) and dementia (Gatz, Pedersen, Berg, Johansson, Johansson, Mortimer, Posner, Viitanen, Winblad, and Ahlbom, 1997) has genetic as well as environmental components, and the parameters for the dropout process could be expressed as $\boldsymbol{\xi} = (\boldsymbol{\xi}_{genes}, \boldsymbol{\xi}_{env})$. Ignorability in this setting means that values are MAR and that genes and environmental factors affecting the longitudinal and the dropout process are acting independently. This is probably not the case when studying cognitive decline in old age, with dropouts due to development of dementia or due to death.

5.4 Dropout due to death

The standard approaches for handling missing data assume missing values to hide true values, and estimation of parameters are performed by integrating over, or imputing, missing values. However, truncation of follow up data due to death is conceptually different from dropout where the individual stays alive. For most outcomes it is not reasonable to think of missing values after death, since they are counterfactual and poorly defined. The value is "missing" not because a non-null value exists and is unobserved, but because a non-null value does not exist. The topic of truncation due to death has been subject to increasing interest in the last few years and has been discussed in the framework of estimation of causal effects of treatments (Zhang and Rubin, in press).

If death times are available one approach for incorporating death, if believed to be related to the longitudinal process, is to model longitudinal data and survival jointly (e.g. Hogan and Laird, 1997b,a). Although this procedure incorporates informative dropout it still does not account for the fundamental difference of truncation due to death and missing values for other reasons. What truncation by death means for the analysis of longitudinal twin data is not clear and how to handle it remains an open question.

5.5 Modelling the dropout

It is useful to model dropout as a function of observed data, to find evidence against the hypothesis of values being MCAR. As mentioned, it is not possible to find evidence against MAR from the observed data. However, regardless of whether the dropout process is MAR or non-ignorable, it is useful from a substantive point of view to model the dropout process and to try to understand how it is related to the data observed prior to dropout.

In the manuscript in the Appendix we model dropout from the SATSA as a discrete process in time, distinguishing between dropout due to death and dropout for other reasons than death. Methodological issues include (i) how to handle item non-response, leading to missing covariates in the dropout model, (ii) how to reduce the dimension of covariates in the dropout model, and (iii) how to incorporate within twin pair dependence in dropout. We refer to the manuscript in the Appendix for a description of the methods and the results for the SATSA.

6 Application to SATSA

The motivation for our research on methods for analyzing longitudinal twin data is the Swedish Adoption/Twin Study of Aging (SATSA), which is a study of the aging process in late life. In this section we describe the sample and the phenotype considered,

cognitive ability. Different latent growth curve models are discussed and results from fitting linear growth curve models are given.

6.1 The data

The Swedish Adoption/Twin Study of Aging (SATSA) is a study of factors influencing normal aging. The SATSA includes both questionnaire assessments and in-person testings of cognitive and functional capabilities, personality and health (Pedersen et al., 1991). The base population of the SATSA comprise all twin pairs in the Swedish Twin Registry who indicated that they had been reared apart, and a control sample of twins reared together, matched to those reared apart on gender, age and county of birth (3838 individuals). We refer to Lichtenstein et al. (2002) or Pedersen et al. (2002) for a general description of the Swedish Twin Registry. The first in-person testing (IPT1) in SATSA took place in 1986-1988 for a sub-sample of pairs and follow-up data were obtained after three (IPT2), six (IPT3) and thirteen years (IPT4). Testing took place in a location convenient to the twins, such as district nurses' offices, health-care schools, and long-term care clinics. Testing was completed during a single 4-hour visit.

The SATSA sample used in this study is restricted to twins with data on at least one cognitive measure at one testing occasion, who reached the age of 50 years during their participation in the study and for whom zygosity is known. Observations obtained after onset of dementia were excluded. The sample that we use includes 840 twins, with individuals from both complete pairs (396 pairs) and incomplete twin pairs (48 individuals). It has the following background characteristics: 59% are female, 63% are dizygotic and the average age at the first in-person testing is 61.7 years (range= $38.2 - 88.0$).

In terms of response data in our sample of 840 twins, 36% have cognitive measures from all four testing occasions, 26% from three testing occasions, 15% from two testing occasions, while 23% only have cognitive measures from a single testing occasion. Table 1 shows participation patterns, revealing that there are several reasons for incomplete observations. One concerns the study design, where individuals are selected as in Pedersen et al. (1991) and restricted to be 50 years or older at entry, with an exception of a sub-sample of twins younger than 50 years, who were recruited for another study and also were administered the same cognitive battery. Some individuals have intermittent missing values, but these are much fewer than missing values due to dropout from the study.

6.2 Measures of cognition

There are hundreds of tests of different aspects of cognition. These tests include measures of broad factors (specific cognitive abilities) such as crystallized intelligence, fluid intelligence, memory and perceptual processing speed. Crystallized intelligence refers to those cognitive processes that are imbedded in a context of cultural meaning and

IPT1	IPT2	IPT3	IPT4	Individuals
⋆	⋆	⋆	⋆	299
⋆	⋆	⋆	-	140
⋆	⋆	-	⋆	10
⋆	-	⋆	⋆	8
-	⋆	⋆	⋆	60
⋆	⋆	-	-	46
-	⋆	⋆	-	18
-	-	⋆	⋆	19
⋆	-	⋆	-	18
⋆	-	-	⋆	24
-	⋆	-	⋆	6
⋆	-	-	-	94
-	⋆	-	-	7
-	-	⋆	-	2
-	-	-	⋆	89

Table 1: Number of individuals and their patterns of participation in the SATSA. ⋆ correspond to an observation, and - is used as a marker for missing values.

reflects the store of knowledge or information that has accumulated over time. Fluid intelligence is defined as the "on-the-spot reasoning ability, a skill not basically dependant on our experience" (Belsky, 1990).

The SATSA cognitive test battery includes 11 cognitive measures drawn from various sources and chosen to assess four areas of cognitive ability (Nesselroade, Pedersen, McClearn, Plomin, and Bergeman, 1988; Pedersen, Plomin, Nessleroade, and McClearn, 1992). Crystallized abilities are tapped by tests of Information, Synonyms, and Analogies. Figure Logic, Block Design, and Card Rotation assess fluid abilities. Memory tests include Digit Span, Thurstone's Picture Memory, and Names & Faces. Finally, Symbol Digit and Figure Identification measure perceptual speed. An overview of cognitive tests and domains are found in Table 2.

The SATSA test scores are often expressed as percentage of maximum score to enable comparison of test results. The distribution from the first testing occasion is shown in Figure 3. The figure shows that for most items, but nor for all, the distributions are fairly symmetrical. The distribution for the test Symbol Digit, which we analyze here, is fairly normal and we take this as reassurance that we can apply the latent growth curve models based on the assumption that the repeated outcomes follow a multivariate normal distribution.

If data from all the 11 tests are to be analyzed as a multivariate outcome, the structure of the test subclasses, based on what aspect of cognition they measure, preferably should be built into the model. Here, we only use data from the test Symbol Digit in the longitudinal twin models. This test measures perceptual speed and is based on the

ability to verbally report digits that correspond to symbols.

Test	Domain
Information	Crystallized ability
Synonyms	Crystallized ability
Analogies	Crystallized ability
Figure Logic	Fluid ability
Block Design	Fluid ability
Card Rotations	Fluid ability
Digit Span	Memory
Thurstone's Picture Memory	Memory
Names & Faces	Memory
Symbol Digit	Perceptual Speed
Figure Identification	Perceptual Speed

Table 2: Cognitive tests and domains of cognition in the SATSA.

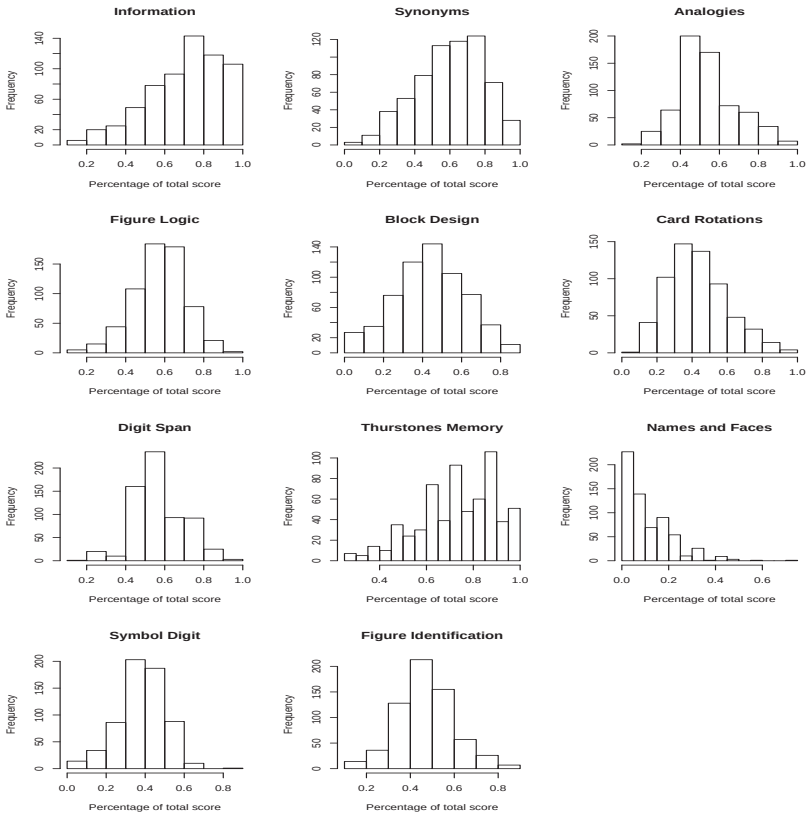

Figure 3: Distribution of test scores at IPT1 expressed as percentage of maximum score.

6.3 Shape of individual trajectories

In Figure 4 scores from the perceptual speed test Symbol Digit are plotted, including the trajectories for three individuals in the bottom, middle and top of the distribution. From the plot it is clear that individual test scores tend to decrease with increasing age. The variability between individuals is quite large for the perceptual speed level, even if trajectories were age adjusted. Individual variability in the rate of change is more difficult to assess, but it appears to be smaller in absolute terms than the variability in the perceptual speed level.

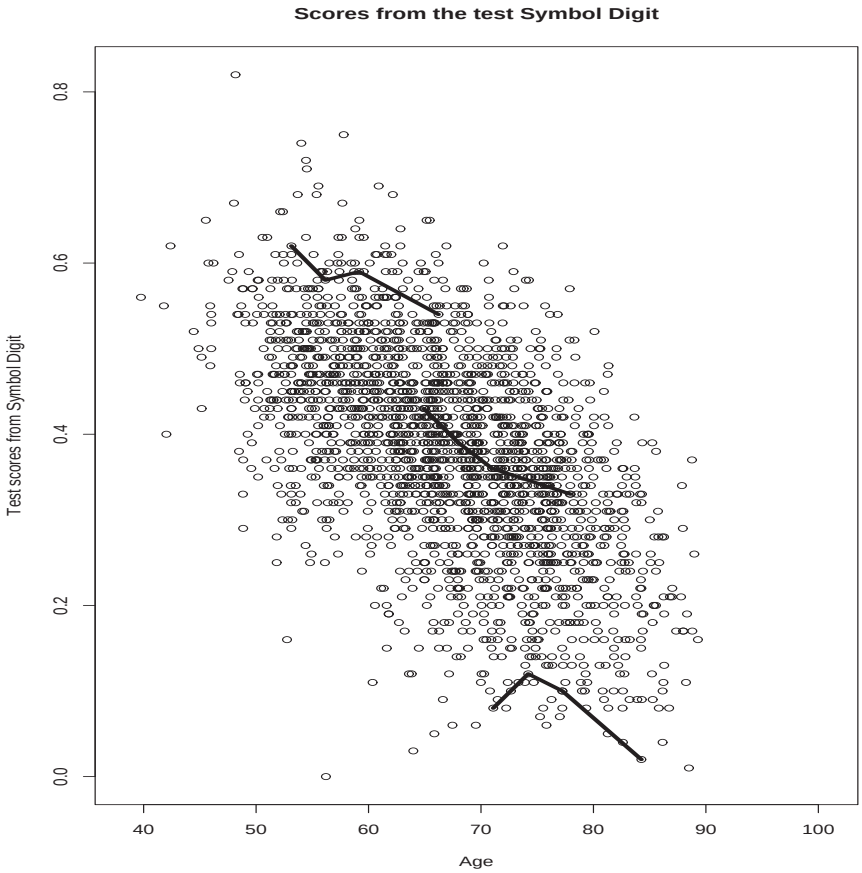

Figure 4: Test scores available for a test measuring perceptual speed, including trajectories for three individuals in the bottom, middle and top of the distribution.

In developmental aging research one argues that in the process of aging there are patterns that are related more to distance to death than to age per se. This relation is explored for the Symbol Digit test score in Figure 5 for a random sub-sample of the 130 individuals in our sample that die before the fourth testing occasion. The plot shows that the individual variability is large, and that more repeated measurements would be needed in order to infer any clear patterns.

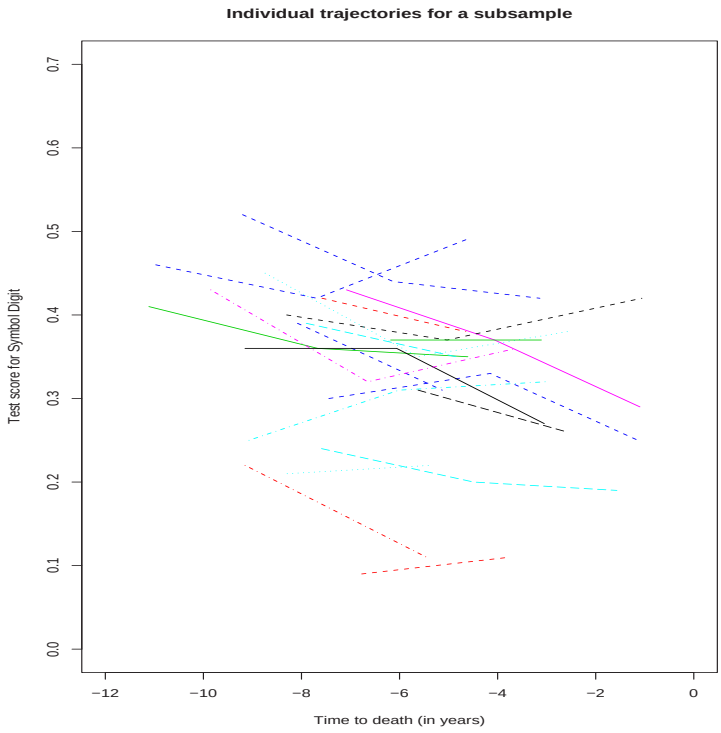

Figure 5: Trajectories for test scores for Symbol Digit as a function of time to death, for as sub-sample of the participants in SATSA that died before the fourth testing occasion.

6.4 Linear latent growth curve models

The linear ACE model, specified by model equations (7), (8) and (9), was fitted to the raw scores from the test Symbol Digit. We later refer to this model as model $M1$. The model fitting was perform using the software Mplus (Muthén and Muthén, 1998-2001), and results are shown in Table 3.

The mean slope is negative and significant, reflecting that perceptual speed tend to decrease with increasing age. Genetic factors seem to be important for the perceptual speed level, but less so for the rate of change. For the level, also non-shared environment seems to explain some of the variability.

To investigate the importance of genes for the level and slope for perceptual speed two more restricted models were fitted to the same data, one excluding the genetic effect on slope (model $M2$) and one excluding the genetic effect on both level and slope (model $M3$). The models all have the same measurement model given by (7) in section 4. The difference lies in the structural equations. Following the notation introduced in section 4, the structural equations are

		Model $M1$		Model $M2$		Model $M3$	
Effect	Par.	Est.	SE	Est.	SE	Est.	SE
Mean level	γ_I	37.4	0.41	37.4	0.41	37.4	0.41
Mean slope	γ_S	-0.75	0.03	-0.75	0.03	-0.75	0.03
η_{AI} on η_I	λ_{A1}	8.21	0.40	8.16	0.37	-	-
η_{CI} on η_I	λ_{C1}	0.62	1.94	0.99	1.28	6.52	0.39
η_{EI} on η_I	λ_{E1}	3.13	0.48	3.16	0.47	5.89	0.32
η_{AI} on η_S	λ_{A2}	0.01	0.05	-	-	-	-
η_{CI} on η_S	λ_{C2}	0.14	0.08	0.14	0.08	0.03	0.03
η_{EI} on η_S	λ_{E2}	0.10	0.05	0.10	0.05	0.05	0.03
η_{AS} on η_S	λ_{A3}	0.00	0.17	-	-	-	-
η_{CS} on η_S	λ_{C3}	0.00	0.65	0.00	0.35	0.16	0.07
η_{ES} on η_S	λ_{E3}	0.14	0.08	0.13	0.08	0.11	0.11
Residual error	σ^2	26.2	1.20	26.2	1.20	26.4	1.21
Parameters	r	12		10		9	
Log likelihood	$l(\hat{\boldsymbol{\theta}}\|\mathbf{y})$	-7314.0		-7314.0		-7334.4	

Table 3: Results from fitting linear latent growth curve models to data on Symbol Digit from the SATSA.

$$
\begin{aligned}
M1: \qquad & \eta_{Iij} = \gamma_I + \lambda_{A1}\eta_{AIij} + \lambda_{C1}\eta_{CIij} + \lambda_{E1}\eta_{EIij} \\
& \eta_{Sij} = \gamma_S + \lambda_{A2}\eta_{AIij} + \lambda_{C2}\eta_{CIij} + \lambda_{E2}\eta_{EIij} + \\
& \qquad + \lambda_{A3}\eta_{ASij} + \lambda_{C3}\eta_{CSij} + \lambda_{E3}\eta_{ESij} \\
M2: \qquad & \eta_{Iij} = \gamma_I + \lambda_{A1}\eta_{AIij} + \lambda_{C1}\eta_{CIij} + \lambda_{E1}\eta_{EIij} \\
& \eta_{Sij} = \gamma_S + \lambda_{C2}\eta_{CIij} + \lambda_{E2}\eta_{EIij} + \lambda_{C3}\eta_{CSij} + \lambda_{E3}\eta_{ESij} \\
M3: \qquad & \eta_{Iij} = \gamma_I + \lambda_{C1}\eta_{CIij} + \lambda_{E1}\eta_{EIij} \\
& \eta_{Sij} = \gamma_S + \lambda_{C2}\eta_{CIij} + \lambda_{E2}\eta_{EIij} + \lambda_{C3}\eta_{CSij} + \lambda_{E3}\eta_{ESij}.
\end{aligned}
$$

As mentioned earlier, the models are parameterized by setting the variances of the genetic factors, η_{AI} and η_{AS}, and the environmental factors, η_{CI}, η_{EI}, η_{CS} and η_{ES}, equal to one, and estimating the corresponding factor loadings. The parameter estimates are shown in Table 3. The hypothesis of no genetic effect on the slope can be tested based on the likelihood ratio, comparing model $M1$ and $M2$. The test statistic is $2(l(\hat{\boldsymbol{\theta}}_{M1}|\mathbf{Y}) - l(\hat{\boldsymbol{\theta}}_{M2}|\mathbf{Y})) = 0.0$, and we accept model $M2$. The conclusion is that there is no significant genetic effect on the slope, which is not surprising in view of the difficulty to detect slope variability from the observed data. Genetic effect estimates would require the ability to distinguish the slope similarity for MZ and DZ twins to be different.

By fitting model $M3$ and comparing it to model $M2$ the hypothesis that also the

genetic effect on level can be excluded is tested. The test statistic is now $2(l(\hat{\boldsymbol{\theta}}_{M2}|\mathbf{Y}) - l(\hat{\boldsymbol{\theta}}_{M3}|\mathbf{Y})) = 40.8$. The difference in number of free parameters in the models is 1, and hence, the test statistic should be compared with a χ^2-distribution with 1 degree of freedom. This is highly significant showing that there is indeed a genetic effect on level of perceptual speed.

In equation (10) in section 4 we proposed a new measure of heritability of change, h_S^2, assuming a constant rate of change, to be used as a measure of the importance of genes for change in trait values. The empirical estimate of the heritability of change for the cognitive test Symbol Digit is here zero:

$$h_S^2 = \frac{0.01^2 + 0.00^2}{0.01^2 + 0.00^2 + 0.14^2 + 0.00^2 + 0.10^2 + 0.14^2} \approx 0.$$

This low value also reflects that the genetic effect on rate of change is small or non-detectable in the SATSA data.

7 Future research and discussion

As data from longitudinal twin studies become more widely available the need for new statistical tools increase. More effort needs to be put on the specification of latent growth curve models corresponding to different patterns of growth, and the clarification of what models that are identifiable. Clearly, the latter depend on the number of repeated measurements available. We intend to focus on the linear, the quadratic, and the change-point model. The first step will be to formulate the change-point model for longitudinal twin data, and define a measure of heritability of change based on this model. The statistical properties of the measures of heritability of change need to be explored and the usefulness of these measures, as an alternative to a nested testing procedure, assessed. Distributional properties of the likelihood-ratio test statistic for testing variance components in this setting also must be investigated further.

Another topic for future research is to investigate how the suggested measures are affected by the dropout process, and whether or not the assumption of ignorability simplifies the argument for how dependence in dropout affects heritability of change.

The issues mentioned above will be addressed in a simulation study. Identification of, and precision in, model parameters will be studied for the linear growth model and the hidden change-point model, as a function of number of repeated measurements. Properties of measures of heritability in change based on these models will be assessed based on Monte Carlo simulations, without any missing data.

Based on simulations we also aim to appraise the effect of dropout on estimates of heritability of change. First MAR, with a dependence in dropout within twin pairs, will be assumed to confirm that the full information maximum likelihood produce consistent estimates of the parameters in the latent growth curve models. Data will

also be simulated based on different scenarios of non-ignorable dropout, and the bias in measures of heritability assessed. We hope this to give an indication of the severity of ignoring the dropout process.

One major goal for statistical research on twin models is to provide a method for analyzing longitudinal twin data that account for informative dropout. Here, a joint model for the longitudinal data and the dropout process may be useful. Another option, that may be even more useful, is to incorporate a sensitivity parameter in latent growth curve models yielding bounds for measures of heritability.

To incorporate multivariate outcomes at each repeated measurement would add yet another layer of complexity to the latent growth curve models. Such an extension would be of scientific interest, since there is still much to learn about the dynamics of behavioral traits and how they are related.

References

Akaike, H. (1987), "Factor analysis and AIC," *Psychometrika*, 52.

Belsky, J. (1990), *The psychology of aging: theory, research, and interventions*, Pacific Grove: Brooks/Cole Publishing Company.

Berg, S. (1996), *Handbook of the psychology of aging*, Academic Press, chap. 18: Aging, behavior, and terminal decline, 4th ed., pp. 323–337.

Bollen, K. (1989), *Structural equations with latent variables*, John Wiley & Sons.

Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum likelihood for incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1–38.

Diggle, P., Liang, K.-Y., and Zeger, S. (1994), *Analysis of longitudinal data*, Oxford University Press.

Gatz, M., Pedersen, N., Berg, S., Johansson, B., Johansson, K., Mortimer, J., Posner, S., Viitanen, M., Winblad, B., and Ahlbom, A. (1997), "Heritability for alzheimers-disease - the study of dementia in swedish twins," *Journals of Gerontology series A - Biological sciences and medical sciences*, 52, 117–125.

Hindsberger, C. (2001), "Robust analysis of quantitative twin data: Genetic and environmental influences on obesity," Ph.D. thesis, Department of Biostatistics, University of Copenhagen and Danish Epidemiology Science Center, Institute of Preventive Medicine, Copenhagen University Hospital.

Hogan, J. W. and Laird, N. M. (1997a), "Mixture models for the joint distribution of repeated measures and event times," *Statistics in Medicine*, 16, 239–257.

— (1997b), "Model-based approaches to analysing incomplete longitudinal and failure time data," *Statistics in Medicine*, 16, 259–272.

Lange, K. (1978), "Central limit theorems for pedigrees," *Journal of mathematical biology*, 6, 59–66.

Lichtenstein, P., deFaire, U., Floderus, B., Svartengren, M., Svedberg, P., and Pedersen, N. (2002), "The Swedish Twin Registry: A unique resourse for clinical, epidemiological and genetic studies," *Journal of Internal Medicine*, 252, 184–205.

Lindstrom, M. and Bates, D. (1990), "Nonlinear mixed effects models for repeated measures data," *Biometrics*, 46, 673–687.

Little, R. J. A. and Rubin, D. B. (2002), *Statistical analysis with missing data*, John Wiley and Sons, 2nd ed.

Loehlin, J. (1996), "The Cholesky approach: a cautionary note," *Behavior Genetics*, 26, 65–69.

McArdle, J. J. (1986), "Latent variable growth within behavior genetic models," *Behavior Genetics*, 16, 163–200.

McArdle, J. J. and Hamagami, F. (2003), "Structural equation models for evaluating dynamic concepts within longitudinal twin analyses," *Behavior Genetics Special Issue on Aging*, 33, 137–159.

Muthén, B. (2002), "Beyond SEM: General latent variable modeling," *Behaviormetrika*, 29, 81–117.

Muthén, B. and Muthén, L. (1998-2001), *Mplus User's Guide*, Muthén & Muthén, Los Angeles.

Nesselroade, J., Pedersen, N., McClearn, G., Plomin, R., and Bergeman, C. (1988), "Factorial and criterion validities of telephone-assessed cognitive ability measures: Age and gender comparisons in adult twins," *Research on Aging*, 10, 220–234.

Pawitan, Y. (2001), *In all likelihood: statistical modelling and inference using likelihood*, New York: Oxford University Press.

Pedersen, N. L., Lichtenstein, P., and Svedberg, P. (2002), "The Swedish Twin Registry in the Third Millenium," *Twin Research*, 5, 427–432.

Pedersen, N. L., McClearn, G. E., Plomin, R., Nessleroade, J. R., Berg, S., and de Faire, U. (1991), "The Swedish Adoption/Twin Study of Aging: An update," *Acta Geneticae Medicae et Gemellologiae*, 40, 7–20.

Pedersen, N. L., Plomin, R., Nessleroade, J. R., and McClearn, G. E. (1992), "Quantitative genetic analysis of cognitive abilities during the second half of the lifespan," *Psychological Science*, 3, 346–353.

Pedersen, N. L., Ripatti, S., Berg, S., Reynolds, C., Hofer, S., Finkel, D., Gatz, M., and Palmgren, J. (2003), "The influence of mortality on twin models of change: addressing missingness through multiple imputation," *Behavior Genetics Special Issue on Aging*, 3, 161–169.

Plomin, R., DeFries, J., McClearn, G., and McGuffin, P. (2001), *Behavioral genetics*, New York: Worth Publishers, 4th ed.

Reynolds, C. A., Finkel, D., Gatz, M., and Pedersen, N. L. (2002a), "Sources of influence on rate of cognitive change over time in Swedish twins: An application of latent growth models," *Experimental Aging Research*, 28, 407–433.

Reynolds, C. A., Finkel, D., McArdle, J., Gatz, M., and Pedersen, N. L. (In press), "Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood," *Developmental Psychology*.

Reynolds, C. A., Gatz, M., and Pedersen, N. L. (2002b), "Individual variation for cognitive decline: Quantative methods for describing patterns of change," *Psychology & Aging*, 17, 271–287.

Rothenberg, T. (1971), "Identification in parametric models," *Econometrica*, 39, 577–591.

Skrondal, A. (1996), "Latent trait, multilevel and repeated measurement modelling with incomplete data of mixed measurement levels," Ph.D. thesis, Section of Medical Statistics, University of Oslo.

Skrondal, A. and Rabe-Hesketh, S. (2004), *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*, Chapman and Hall.

Slate, E. and Turnbull, B. (2000), "Statistical models for longitudinal biomarkers of disease onset," *Statistics in Medicine*, 19, 617–637.

Tanaka, J. (1993), *Multifaceted conceptions of fit in structural equation models*, Newbury Park: Sage Publications, chap. 2, Testing structural equation models, K. Bollen and J.S. Long (eds.), pp. 10–39.

Verbeke, G. and Molenberghs, G. (2000), *Linear mixed models for longitudinal data*, New York: Springer-Verlag.

Yashin, A. and Iachine, I. (1995), "Genetic analysis of durations: correlated frailty model applied to survival of Danish twins." *Genetic Epidemiology*, 12, 529–538.

Zhang, J. and Rubin, D. (In press), "Estimation of causal effects via principal stratification when some outcomes are truncated by "death"," *Journal of Education and Behavioral Statistics*.

Appendix: manuscript

Models for dropout mechanisms in a longitudinal twin study

Annica Dominicus[1,2], Juni Palmgren[1,2], Nancy L. Pedersen[2,3]

[1]Department of Mathematical Statistics, Stockholm University, Stockholm, Sweden
[2]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[3]Department of Psychology, University of Southern California, Riverside, USA

Keywords: Longitudinal twin study, dropout, cognition, bivariate logistic regression, generalized estimating equations, multiple imputation.

Abstract

Longitudinal studies typically suffer from dropouts due to death and dropouts for other reasons than death, especially when studying processes in late life. We apply two methods for modelling dropout, generalized estimating equations (GEE) and bivariate logistic regression (BLR), to data from the Swedish Adoption/Twin Study of Aging (SATSA) (Pedersen, McClearn, Plomin, Nessleroade, Berg, and de Faire, 1991). The aim is to assess (i) to what extent the level and rate of change in earlier measures of cognition are associated with dropout and (ii) the pattern of within-twin-pair dependency of dropout. The phenotype considered is repeated measures of cognition. The results show that dropout is associated with low levels of earlier measures of all domains of cognition considered here: crystallized ability, fluid ability, memory and perceptual speed. The hypothesis that rate of change in cognition is a predictor for dropout could not be confirmed, probably due to low power. The BLR analysis shows that the within-twin-pair dependency of dropout is significant, and that the association is larger for MZ compared to DZ twins, although the MZ/DZ difference is not significant. We discuss possible implications of these findings for the assessment of heritability of rate of change for cognition in an elderly population.

1 Introduction

Longitudinal data are increasingly available in behavioral science. Longitudinal twin studies offer a possibility to assess the relative importance of genes and environmental factors for the dynamics of behavioral traits. When studying processes in late life, there is often a substantive loss to follow up due to death or for other reasons than death. Selective participation may endanger a study's validity, and more consideration needs to be put on assessing what it means for inferences on heritability of change based on incomplete longitudinal twin data.

Analysis of longitudinal twin data often involves fitting latent growth curve models (LGCM), based on maximum likelihood methods. If the data is incomplete, full information maximum likelihood (FIML), is used (e.g. McArdle and Hamagami, 2003). Following the notation of Little and Rubin (2002), this procedure gives consistent estimates of model parameters if values are missing completely at random (MCAR) or missing at random (MAR). If not, estimates will be inconsistent. A discussion of dropout from longitudinal twin studies is given in (Pedersen, Ripatti, Berg, Reynolds, Hofer, Finkel, Gatz, and Palmgren, 2003).

The aim of this paper is to investigate dropout from a study of cognition in old age, using data from the Swedish Adoption/Twin Study of Aging (SATSA), by modelling dropout as a function of earlier measures of cognition. We make a distinction between dropout due to death and dropout for other reasons. Based on these analysis, the hypothesis of data being MCAR can be tested from observed data. By modelling the within-twin-pair association of dropout we hope to clarify whether genetic factors are expected to affect the dropout mechanism and whether, as a consequence of this, estimates of the heritability of change are affected by the dropout. Cognition is measured by a battery of tests designed to represent four domains of cognition: crystallized ability, fluid ability, memory and perceptual speed. Different domains have different etiologies, and possibly they have different connections to dropout for various reasons (e.g. Reynolds, Finkel, Gatz, and Pedersen, 2002).

Methodological issues addressed in this paper include how to handle item nonresponse, resulting in incomplete test scores. These missing values are assumed to be MAR and are handled with multiple imputation techniques, imputing values for missing responses and summarizing the conclusions based on a set of imputed data sets. Another issue concerns the summarizing of test results that are measuring the same domain of cognition. This is done using the first principal component for each of the four domains of cognition. A key methodological issue concerns how to address the dependence between twins. We approach this issue in two ways: using generalized estimating equations (GEE) (Liang and Zeger, 1986) and bivariate logistic regression (BLR) (Dale, 1986; Palmgren, 1989), respectively. The latter allows maximum likelihood estimation of the within-twin-pair odds ratio of dropout and inferences on contrasts between MZ and DZ twin pairs.

In section 2 the data from the SATSA is described. Models for dropout and related methodological issues are presented in section 3. The results from mod-

elling the dropout from the SATSA is given in section 4, and implications of the findings are discussed in section 5.

2 The data

The Swedish Adoption/Twin Study of Aging (SATSA) is a study of the aging process in late life. It includes both questionnaire assessments and in-person testings of cognitive and functional capabilities, personality and health, and have been described in detail elsewhere (Pedersen et al., 1991). The participants were identified via the Swedish Twin Registry (Lichtenstein, deFaire, Floderus, Svartengren, Svedberg, and Pedersen, 2002; Pedersen, Lichtenstein, and Svedberg, 2002). The first in-person testing (IPT1) took place in 1986-1988 and follow-up data were obtained after three (IPT2), six (IPT3) and thirteen (IPT4) years. Testing took place in a location convenient to the twins, such as district nurses' offices, healthcare schools, and long-term care clinics. Testing was completed during a single 4-hour visit.

The SATSA sample used in this study is restricted to twins with data on at least one of the nine cognitive measures considered here, at one of the three first testing occasions of SATSA. A further restriction is that they reached the age of 50 years during their participation in the study and that the zygosity is known. Observations obtained after onset of dementia were excluded. The sample that we use includes 721 twins, with individuals from both complete pairs (341 pairs) and incomplete twin pairs (39 individuals). It has the following background characteristics: 59% are female and 64% are dizygotic and the average age at the first in-person testing (IPT1) is 63.8 years (range$= 41.8 - 88.0$).

The SATSA cognitive test battery includes 11 cognitive measures drawn from various sources and chosen to assess four domains of cognition (Nesselroade, Pedersen, McClearn, Plomin, and Bergeman, 1988; Pedersen, Plomin, Nessleroade, and McClearn, 1992). This study is restricted to nine tests, that are believed to best capture crystallized ability, fluid ability, memory and perceptual speed. Crystallized abilities are tapped by tests of Information and Synonyms. Figure Logic and Block Design assess fluid abilities. Memory tests include Digit Span, Thurstone's Picture Memory, and Names & Faces. Finally, Symbol Digit and Figure Identification measure perceptual speed. An overview of the cognitive tests and domains are found in Table 1. To enable comparisons of test results, the test scores are expressed as percentage of maximum score.

Table 2 show the number of participants in each of the first three testings, for whom data is available for at least one of the nine cognitive tests. Of the 589 individuals observed at IPT1, 94 (16%) drop out before IPT2. At IPT2 additional participants enter the study and a total of 570 individuals are observed, of which 53 (9%) drop out before IPT3. Yet other participants enter at IPT3, resulting in totally 564 participants at IPT3. 178 (32%) of those drop out before IPT4. Through matching the data set to the Swedish Death Registry, death dates are available for the 130 participants in the sample, who died before IPT4. The

Test	Cognitive ability
Information	Crystallized ability
Synonyms	Crystallized ability
Figure Logic	Fluid ability
Block Design	Fluid ability
Digit Span	Memory
Thurstone's Picture Memory	Memory
Names & Faces	Memory
Symbol Digit	Perceptual Speed
Figure Identification	Perceptual Speed

Table 1: Domains of cognition and cognitive tests used in this study, available from the SATSA.

number of years to death from the last testing occasion ranged from 0.4 to 11.5 years, with a mean of 3.7 years.

When modelling the probability of dropout between any two adjacent testing occasions, we distinguish between dropout due to death and dropout for other reasons than death (the latter include dropout due to cognitive decline). Dropout due to death was defined as dropout where the person died within three years after the last testing occasion, which corresponds to the planned time for follow up for all testings except for the last one.

From the first three testing occasions 1723 observations on cognitive performance, tapped by the nine tests mentioned above, are available. Of these observations 87% include test results for all nine items, 7% have information missing on one item and 6% have information missing on two or more items.

IPT	Participants	Dropouts
1	589	94 (20 deaths, 74 dropouts)
2	570	53 (15 deaths, 38 dropouts)
3	564	178 (29 deaths, 149 dropouts)

Table 2: Number of participants and dropouts, due to death and for other reasons, in the sample from the first three testing occasions in the SATSA.

3 Methods

3.1 Multiple imputation for item non-response

Item non-response, resulting in incomplete test results, was assumed to be MAR, and was handled with multiple imputation techniques (Rubin, 1987; Schafer, 1997). Multiple imputation is a Bayesian Monte Carlo data augmentation approach and shares the same underlying philosophy as the non-Bayesian EM procedure: solving an incomplete-data problem by repeatedly solving the complete-data version. The missing values are replaced by simulated values from the conditional predictive distribution given the observed data, resulting in a number

of complete data sets. Each of the completed data sets is analyzed by standard complete-data methods. The variability in the results over the imputed data sets provides a measure of the uncertainty due to missing data, and combined with measures of sample variation, inferential statements can be made about the parameters of interest based on the observed incomplete data. The total variance is divided into two parts: the *within-imputation variance*, $\bar{U}$, and the *between-imputation variance*, B,

$$\bar{U} = \frac{1}{m}\sum_{t=1}^{m} U^{(t)}, \qquad B = \frac{1}{m-1}\sum_{t=1}^{m}(\hat{Q}^{(t)} - \bar{Q})^2,$$

where $\hat{Q}^{(t)}$ is the parameter estimate based on the tth imputed data set ($t = 1, ..., m$), $\bar{Q}$ is the mean of the parameter estimates, and $U^{(t)}$ is the estimated variance for the parameter estimate $\hat{Q}^{(t)}$. The total variance is

$$T = \bar{U} + (1 + m^{-1})B.$$

A useful diagnostic tool for assessing how the missing data contribute to inferential uncertainty about the parameter of interest is the *relative increase in variance due to non-response*

$$r = \frac{(1 + m^{-1})B}{\bar{U}}.$$

We use this measure to assess how much the item non-response contribute to the uncertainty in parameter estimates in the dropout model.

3.2 Principal components

To decrease the number of explanatory variables when using cognitive test scores as predictors for dropout, the best single composite variable is produced for each domain of cognition, using the method of principal components (Jolliffe, 1986). The idea is to generate a single linear composite function of the original variables which maximally discriminates the individuals in the data set. The principal components are linear functions of the original variables, $X_1, ..., X_n$, of the form $W_j = e_{1j}X_1 + e_{2j}X_2 + \cdots + e_{nj}X_n$, with the constraint $e_{1j}^2 + e_{2j}^2 + \cdots + e_{nj}^2 = 1$, for $j = 1, ..., n$. The first principal component is the linear function of this form, where the variance of W_1 has the maximum variance over all possible linear functions of the original variables, subject to the constraint on the loadings. We apply the method of principal components to the standardized test scores and use the first principal component score for each cognitive domain as a single measure of that ability.

3.3 Modelling dropout for twin pairs

We model the probability for an individual to drop out between any two adjacent testing occasions, and assume time lags to be conditionally independent. When using level of cognition at last testing occasion as a predictor for dropout, the dropout indicators for the first, second and third time lag are modelled jointly. When using rate of change in cognition as a predictor for dropout we need data from at least two preceding testing occasions, and hence, only dropout for the second and third time lag could be used. The within-twin-pair dependence in dropout is accounted for in two different ways: using a semi-parametric approach, generalized estimating equations, and a parametric approach, bivariate logistic regression.

Generalized estimating equation
Generalized estimating equation (GEE), proposed by Liang and Zeger (1986), is based on the idea of specifying the mean and variance structure for an outcome without explicitly specifying the probability distribution. For the binary outcomes from the two members of twin pair i, denoted by Y_{i1} and Y_{i2}, a natural model for the means, $E[Y_{i1}] = p_{i1}$ and $E[Y_{i2}] = p_{i2}$, are

$$\begin{aligned} \text{logit}(p_{i1}) &= \boldsymbol{\beta}\mathbf{x}_{i1} \\ \text{logit}(p_{i2}) &= \boldsymbol{\beta}\mathbf{x}_{i2}, \end{aligned}$$

where $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$ are covariate vectors. The variance expressions for the binary outcomes are $\text{Var}(Y_{i1}) = p_{i1}(1-p_{i1})$ and $\text{Var}(Y_{i2}) = p_{i2}(1-p_{i2})$. Using vector notation, the outcome from twin pair i is $\mathbf{Y}_i = (Y_{i1}, Y_{i2})$ with mean $\mathbf{p}_i = (p_{i1}, p_{i2})$. The variance-covariance matrix for $\mathbf{Y}_i$, denoted by $\mathbf{V}_i$, is specified by assuming a 'working' correlation structure, denoted by $\mathbf{R}_i(\alpha)$. We use an exchangeable correlation structure $\mathbf{R}(\alpha) = \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}$, assuming the correlation structure to be the same for all twin pairs. The variance-covariance matrix is

$$\mathbf{V}_i = \phi \mathbf{A}_i^{1/2} \mathbf{R}(\alpha) \mathbf{A}_i^{1/2},$$

where $\mathbf{A}_i$ is a diagonal matrix with the variances of Y_{i1} and Y_{i2} on the diagonal, and ϕ is a dispersion parameter. Given the mean and covariance specifications, the GEE estimate $\hat{\boldsymbol{\beta}}$ is the solution of the system of estimating equations

$$\sum_i \frac{\partial \mathbf{p}_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{p}_i) = 0.$$

The estimating equations are solved for $\boldsymbol{\beta}$ and α using a two-stage procedure iterating between the estimation of α given $\boldsymbol{\beta}$ and the estimation of $\boldsymbol{\beta}$ given α. The GEE yields consistent estimates of $\boldsymbol{\beta}$, even if the assumed covariance

structure is not correct (Liang and Zeger, 1986). We use the implementation of this algorithm in the procedure xtgee in STATA (StataCorp, 2003).

Bivariate logistic regression
An alternative approach is to specify a full parametric model for the bivariate binary outcome $\mathbf{Y}_i = (Y_{i1}, Y_{i2})$. Bivariate logistic regression is based on the assumption that $\mathbf{Y}_i$ follow a multinomial distribution with the four probabilities $p_{ilm} = P(Y_{i1} = l, Y_{i2} = m)$, where $l, m = 0, 1$. The marginal probabilities are given by $p_{i1} = p_{i11} + p_{i10}$ and $p_{i2} = p_{i11} + p_{i01}$. Further, $\psi_i = p_{i11}p_{i00}/p_{i10}p_{i01}$ denotes the odds ratio. The probability p_{i11} can be expressed in terms of p_{i1}, p_{i2} and ψ_i

$$p_{i11} = \begin{cases} \frac{1}{2}(\psi_i - 1)^{-1}(a_i - \sqrt{a_i^2 + b_i}) & \text{if } \psi_i \neq 1 \\ p_{i1}p_{i2} & \text{if } \psi_i = 1, \end{cases}$$

where $a_i = 1 + (p_{i1} + p_{i2})(\psi_i - 1)$ and $b_i = -4\psi_i(\psi_i - 1)p_{i1}p_{i2}$ (Dale, 1986; Palmgren, 1989). The other multinomial probabilities follow from the marginal probabilities p_{i1} and p_{i2}. The bivariate logistic regression model is specified by expressing $\text{logit}(p_{i1})$, $\text{logit}(p_{i2})$ and $\log(\psi_i)$ as linear predictors $\boldsymbol{\beta}\mathbf{x}_{i1}$, $\boldsymbol{\beta}\mathbf{x}_{i2}$ and $\boldsymbol{\gamma}\mathbf{x}_{i3}$, respectively

$$\begin{aligned} \text{logit}(p_{i1}) &= \boldsymbol{\beta}\mathbf{x}_{i1} \\ \text{logit}(p_{i2}) &= \boldsymbol{\beta}\mathbf{x}_{i2} \\ \log \psi_i &= \boldsymbol{\gamma}\mathbf{x}_{i3}. \end{aligned}$$

Estimates of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are obtained by maximum likelihood estimation. The contribution from twin pair i to the log-likelihood is

$$l(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{cases} y_{i1}y_{i2}\ln(p_{i11}) + y_{i1}(1 - y_{i2})\ln(p_{i10}) + & \\ (1 - y_{i1})y_{i2}\ln(p_{i01}) + (1 - y_{i1})(1 - y_{i2})\ln(p_{i00}) & \text{if } y_{i1} \text{ and } y_{i2} \text{ observed} \\ y_{i1}\ln(p_{i1}) + (1 - y_{i1})\ln(1 - p_{i1}) & \text{if only } y_{i1} \text{ observed} \\ y_{i2}\ln(p_{i2}) + (1 - y_{i2})\ln(1 - p_{i2}) & \text{if only } y_{i2} \text{ observed.} \end{cases}$$

The maximum likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ were calculated using the ml facility in STATA (Gould and Sribney, 1999). Needed for this calculation is a program that specifies the log-likelihood, refer to the Appendix for the program code.

4 Results

Results from tests designed to measure the same domain of cognition were summarized into one measure using the first principal component score for each domain.

The loadings and the proportion of variance explained by the first principal component for each domain are given in Table 3. The proportion of variance explained by the first principal component range from 0.8 to 0.85 for crystallized ability, fluid ability and perceptual speed. The corresponding proportion for memory is somewhat lower, 0.6, which is connected to the fact that there are three tests for memory, but only two tests for the other domains of cognition. These high numbers indicate that the use of principal components to combine test scores designed to measure the same ability is adequate.

Test	Crystallized	Fluid	Memory	Perceptual Speed
Information	0.69	-	-	-
Synonyms	0.72	-	-	-
Figure Logic	-	0.52	-	-
Block Design	-	0.85	-	-
Digit Span	-	-	0.39	-
Thurstone's Picture	-	-	0.83	-
Names & Faces	-	-	0.41	-
Symbol Digit	-	-	-	0.66
Figure Identification	-	-	-	0.75
Variance explained	0.84	0.80	0.60	0.85

Table 3: Principal component loadings and proportion of variance explained by first principal component for each domain of cognition.

The difference between individuals that drop out and those who remain in the study was explored graphically. In Figure 1 the distribution of the first principal components for crystallized ability, fluid ability, memory and perceptual speed at last testing occasion, stratified on dropout, are plotted. The figure indicates that dropout is associated with low levels of all four domains of cognition. The patterns are not very clear though, due to the small number of individuals in some strata when studying different time lags separately.

To further investigate the association between dropout and earlier measures of cognition, the probability of dropout between any two adjacent testing occasions was modelled as a function of age (centered at 65 years), length of time lag, sex and earlier measures of cognition. The cognitive measures used are the first principal component scores at the last testing for each domain, and the relative change in these scores from the next last to the last testing. The model for twin j in pair i is

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 age_{ij} + \beta_2 time_{ij} + \beta_3 sex_{ij} + \beta_4 cognition_{ij},$$

Using multiple imputation techniques implemented in the package norm in R (Novo and Shafer, 2002) to account for item non-response, five data sets were imputed and a multiple imputation analysis performed. This first investigation of what domains of cognition that predict dropout, was based on the method of GEE, assuming the within pair correlation of dropout to be the same for all twin pairs.

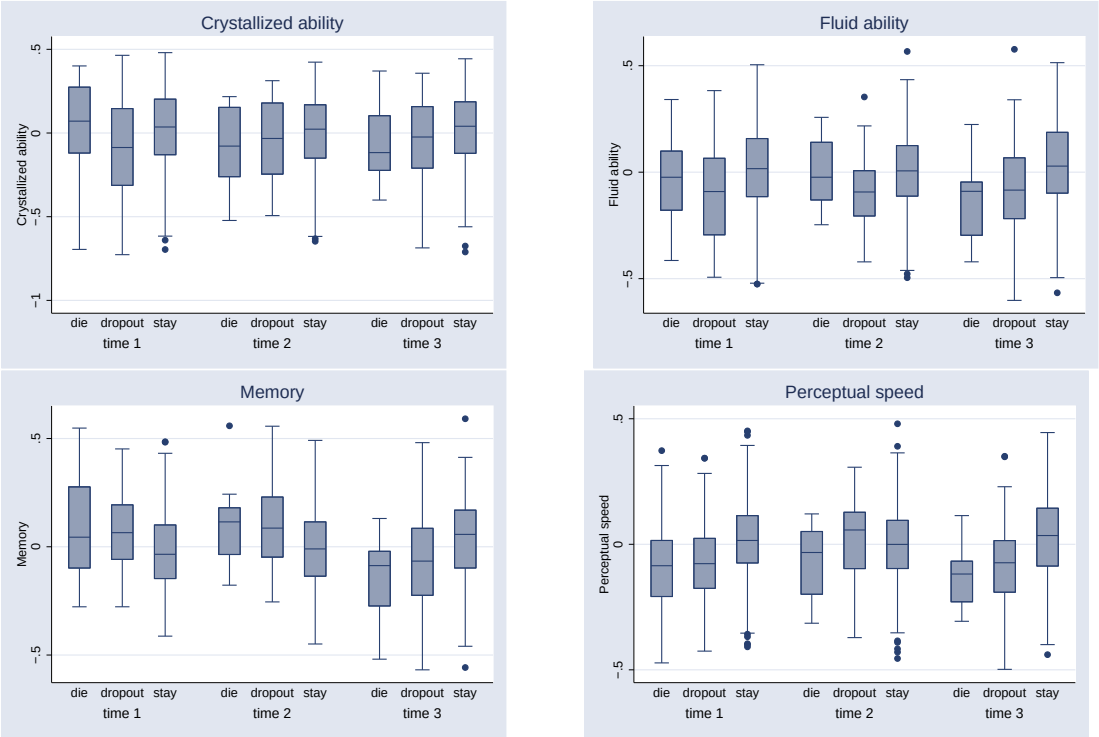

Figure 1: First principal component scores for each domain of cognition at first three testing occasions, stratified on dropout.

Covariate	Par.	Crystallized	Fluid	Memory	Perceptual speed
Intercept	β_0	-2.48 (0.55)	-2.51 (0.55)	-2.74 (0.59)	-2.56 (0.56)
Age	β_1	0.18 (0.02)	0.18 (0.02)	0.17 (0.02)	0.17 (0.02)
Time lag	β_2	0.12 (0.07)	0.12 (0.07)	0.12 (0.07)	0.12 (0.07)
Sex	β_3	-1.54 (0.29)	-1.50 (0.29)	-1.40 (0.31)	-1.48 (0.30)
Cognition	β_4	-0.85 (0.60)	-0.63 (0.82)	-2.02 (0.89)	-1.39 (1.17)
r		0.06	0.04	0.17	0.08

Table 4: GEE estimates (and standard errors) for dropout due to death. The covariate cognition refers to the level of the first principal component score. r is the relative increase in variance of $\hat{\beta}_4$ due to non-response, obtained from the multiple imputation analysis.

Covariate	Par.	Crystallized	Fluid	Memory	Perceptual speed
Intercept	β_0	-2.97 (0.32)	-3.04 (0.32)	-3.29 (0.33)	-3.17 (0.33)
Age	β_1	0.06 (0.01)	0.05 (0.01)	0.05 (0.01)	0.04 (0.01)
Time lag	β_2	0.27 (0.04)	0.27 (0.04)	0.28 (0.04)	0.27 (0.04)
Sex	β_3	-0.08 (0.16)	-0.05 (0.16)	0.09 (0.16)	0.04 (0.16)
Cognition	β_4	-0.95 (0.32)	-1.74 (0.42)	-1.76 (0.41)	-2.10 (0.59)
r		0.02	0.07	0.06	0.10

Table 5: GEE estimates (and standard errors) for dropout for other reasons than death. The covariate cognition refers to the level of the first principal component score.

The results in Table 4 show that the probability of dropout due to death increase with age and the length of time lag to the next testing occasion. Males have a significantly larger probability of dropout compared to females. A high probability of dropout due to death is significantly associated with low levels of memory at the last testing occasion. The trend is the same for other domains of cognition, although non-significant. The parameter estimates in Table 5, for the models for dropout for other reasons than death, point in the same direction as those for dropout due to death, except for sex. The results show that dropout, when excluding individuals that die within three years after the last testing occasion, is associated with low values for all the cognitive domains: crystallized ability, fluid ability, memory and perceptual speed.

Included in the tables is the measure of relative increase in variance of the parameter estimate $\hat{\beta}_4$ due to non-response, r, that was defined in section 3.1. This measure ranges from 4% to 20% for the different domains of cognition, indicating that the relative increase in variance due to item non-response was modest.

To explore if the different domains of cognition together explain more about the dropout than what they do separately, we fit a model for dropout for other reasons that death, using memory and one additional measure of cognition as predictors. The results, given in Table 6, show that both fluid ability and perceptual speed are significant in these models. Additional modelling revealed that there is no gain in extending these dropout models to include more than two measures of cognition as covariates.

Covariate	Par.	Crystallized	Fluid	Perceptual speed
Int.	β_0	-3.23 (0.33)	-3.22 (0.33)	-3.32 (0.33)
Age	β_1	0.05 (0.01)	0.04 (0.01)	0.04 (0.01)
Time lag	β_2	0.28 (.04)	0.28 (0.04)	0.28 (0.04)
Sex	β_3	0.05 (0.16)	0.03 (0.16)	0.10 (0.16)
Memory	β_4	-1.49 (0.46)	-1.14 (0.48)	-1.22 (0.49)
Other	β_5	-0.42 (0.37)	-1.20 (0.50)	-1.40 (0.69)

Table 6: GEE Estimates (and standard errors) for dropout due to other reason than death, including level of memory and one additional cognitive ability as predictors.

It has been suggested that change in cognition has a large predictive value for dropout. We investigate this by modelling dropout between IPT2 and IPT3 and dropout between IPT3 and IPT4 jointly, using relative change between the two previous testings as predictors. The results, given in Table 7 and Table 8, for dropout due to death and dropout for other reasons respectively, reveal that the parameter estimates for change in cognition are non-significant. Hence the hypothesis could not be confirmed. The relative increase in variance due to non-response, r, also turned out to be extremely large for some of these analysis.

Based on BLR, a model for the within-twin-pair odds ratio for dropout was specified, in addition to the model for the probability of dropout used in the GEE analysis. The log odds ratio was modelled as a linear function of zygosity, allowing the dropout dependency to be different for MZ and DZ twins. Results are

Covariate	Par.	Crystallized	Fluid	Memory	Perceptual speed
Intercept	β_0	-3.43 (0.80)	-3.42 (0.80)	-3.40 (0.80)	-3.41 (0.79)
Age	β_1	0.21 (0.03)	0.21 (0.03)	0.21 (0.03)	0.21 (0.03)
Time lag	β_2	0.21 (0.10)	0.21 (0.10)	0.20 (0.10)	0.20 (0.10)
Sex	β_3	-1.52 (0.40)	-1.52 (0.40)	-1.55 (0.40)	-1.53 (0.39)
Change	β_4	-0.004 (0.009)	0.007 (0.008)	0.01 (0.03)	-0.002 (0.003)
r		0.40	0.04	4.57	0.37

Table 7: GEE estimates (and standard errors) for dropout due to death. The covariate change refers to the relative change between the results from cognitive tests from the two previous testings.

Covariate	Par.	Crystallized	Fluid	Memory	Perceptual speed
Intercept	β_0	-3.60 (0.44)	-3.62 (0.44)	-3.58 (0.44)	-3.60 (0.44)
Age	β_1	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)
Time lag	β_2	0.39 (0.06)	0.39 (0.06)	0.39 (0.06)	0.39 (0.06)
Sex	β_3	-0.30 (0.21)	-0.29 (0.21)	-0.30 (0.21)	-0.29 (0.21)
Change	β_4	-0.01 (0.01)	0.003 (0.009)	$\approx$0.00 (0.01)	0.001 (0.003)
r		1.79	0.99	1.54	0.34

Table 8: GEE estimates (and standard errors) for dropout for other reasons than death. The covariate change refers to the relative change between the results from cognitive tests from the two previous testings.

shown in Table 9. For dropout due to other reasons than death the log odds ratio of dropout is significant. The odds ratio of dropout is larger for MZ compared to DZ twins, even though this difference is not significant. In the analysis of dropout due to death very few twin pairs enter into the analysis resulting in extremely large standard errors for the log odds ratio parameters. In fact, for dropout due to death, there are 568 observations where both twins stay in the study to the next testing, 34 pairs where one twin drop out, but only 2 pairs where both twins drop out due to death. The corresponding numbers for dropout for other reasons that death are: 568 observations where both twins stay, 107 pairs where one twin dropout, and 40 pairs where both twins drop out.

Covariate	Par.	Dropout due to death	Dropout for other reasons
Intercept	β_0	-2.70 (0.52)	-3.31 (0.35)
Age	β_1	0.17 (0.02)	0.05 (0.01)
Time lag	β_2	0.13 (0.07)	0.28 (0.04)
Sex	β_3	-1.47 (0.31)	0.10 (0.16)
Memory	β_4	-2.05 (0.84)	-1.66 (0.52)
Intercept (OR)	γ_0	-30.9 (1567)	2.18 (1.03)
Zygosity (OR)	γ_1	15 (783)	-0.37 (0.60)

Table 9: Parameter estimates (standard errors) based on BLR including first principal component score for memory at last measurement as predictor for dropout.

5 Discussion

We have shown that dropout due to death is associated with low scores on earlier measures of memory. The trend was the same for other domains of cognition, although not significant. The dropout for other reasons than death is related to low levels of earlier measures of all the domains of cognition considered: crystallized ability, fluid ability, memory and perceptual speed. This is evidence against the hypothesis that dropout for other reasons than death is MCAR. It reveals that it is not only truncation due to death that may introduce a bias in latent growth curve modelling, based on the observed longitudinal data, where the dropout process is ignored.

The models used for modelling dropout all assume that data is missing at random (MAR), i.e. that dropout only depend on observed quantities. The assumption of MAR is not testable. Extensions that incorporate informative dropout in longitudinal studies have been suggested, such as the models proposed by Diggle and Kenward (1994), which combine a multivariate linear model for the underlying longitudinal response with a logistic regression model for the dropout process. The latter incorporates dependence of probability of dropout on unobserved, missing, observations. However, these models have been criticized (see the discussion of the paper by Diggle and Kenward), since estimating the "unestimable" can be accomplished only by making distributional assumptions or assumptions about associations.

The major fear when modelling longitudinal twin data with dropout is that within-twin-pair dependency of dropout is different for MZ and DZ twins, and that this may affect conclusions about the genetic importance for the longitudinal process when basing the analysis on the observed part of the longitudinal data and ignoring the dropout. In the BLR modelling we show that there is indeed a strong within-twin-pair dependency of dropout, even when excluding those twins who die within three years after the last testing and adjusting for age, sex, time lag and earlier measures of cognition. However, the difference between MZ and DZ twins is not significant, and there is no strong indication in the observed dropout process that it should bias conclusions about heritability when ignoring the dropout. However, the findings that the difference for MZ and DZ twins is non-significant may be due to low power. Clearly, more research on the role of dropout in longitudinal twin modelling is needed.

References

Dale, J. R. (1986), "Global Cross-Ratio Models for Bivariate, Discrete, Ordered Responses," *Biometrics*, 42, 909–917.

Diggle, P. and Kenward, M. (1994), "Informative drop-out in longitudinal data analysis," *Applied Statistics*, 43, 49–93.

Gould, W. and Sribney, W. (1999), *Maximum likelihood estimation with Stata*, Stata Press.

Jolliffe, I. (1986), *Principal component analysis*, New York: Springer-Verlag.

Liang, K. and Zeger, S. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22.

Lichtenstein, P., deFaire, U., Floderus, B., Svartengren, M., Svedberg, P., and Pedersen, N. (2002), "The Swedish Twin Registry: A unique resourse for clinical, epidemiological and genetic studies," *Journal of Internal Medicine*, 252, 184–205.

Little, R. J. A. and Rubin, D. B. (2002), *Statistical analysis with missing data*, John Wiley and Sons, 2nd ed.

McArdle, J. J. and Hamagami, F. (2003), "Structural equation models for evaluating dynamic concepts within longitudinal twin analyses," *Behavior Genetics Special Issue on Aging*, 33, 137–159.

Nesselroade, J., Pedersen, N., McClearn, G., Plomin, R., and Bergeman, C. (1988), "Factorial and criterion validities of telephone-assessed cognitive ability measures: Age and gender comparisons in adult twins," *Research on Aging*, 10, 220–234.

Novo, A. and Shafer, J. (2002), *The norm package: Analysis of multivariate normal datasets with missing values*.

Palmgren, J. (1989), "Regression models for bivariate binary resonses," Tech. Rep. 101, Deptartment of Biostatistics School of Public Health and Community Medicine in Seattle, Washington.

Pedersen, N. L., Lichtenstein, P., and Svedberg, P. (2002), "The Swedish Twin Registry in the Third Millenium," *Twin Research*, 5, 427–432.

Pedersen, N. L., McClearn, G. E., Plomin, R., Nessleroade, J. R., Berg, S., and de Faire, U. (1991), "The Swedish Adoption/Twin Study of Aging: An update," *Acta Geneticae Medicae et Gemellologiae*, 40, 7–20.

Pedersen, N. L., Plomin, R., Nessleroade, J. R., and McClearn, G. E. (1992), "Quantitative genetic analysis of cognitive abilities during the second half of the lifespan," *Psychological Science*, 3, 346–353.

Pedersen, N. L., Ripatti, S., Berg, S., Reynolds, C., Hofer, S., Finkel, D., Gatz, M., and Palmgren, J. (2003), "The influence of mortality on twin models of change: addressing missingness through multiple imputation," *Behavior Genetics Special Issue on Aging*, 3, 161–169.

Reynolds, C. A., Finkel, D., Gatz, M., and Pedersen, N. L. (2002), "Sources of influence on rate of cognitive change over time in Swedish twins: An application of latent growth models," *Experimental Aging Research*, 28, 407–433.

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

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, Chapman and Hall.

StataCorp (2003), *Stata Statistical Software: Release 8.0*, Stata Corporation, College Station, TX.

Appendix: STATA program for bivariate logistic regression

```
/*************************************
 BIVARIATE LOGISTIC REGRESSION IN STATA
*************************************

The program blr specified below define the log-likelihood based on
the multinomial distribution for a bivariate binary outcome
y=(y1,y2), specified in terms of the marginal probabilities and
the odds ratio.

The model is specified in the command: ml model lf blr () () () ()
() (). Equation 1, i.e. the first parenthesis, defines the linear
predictor for the log odds ratio, leave empty to only include an
intercept. The following equations correspond to the parameters in
the linear predictor for the logits of marginal probabilities. In
this program there is one intercept (equation 2) and four
covariate parameters (equation 3-6), corresponding to covariates
named x11-x14 for y1 and x21-x24 for y2.
*************************************************************/

program define blr
    args lnf t1 t2 t3 t4 t5 t6
    tempvar psi ex1 p1 ex2 p2 a b p11
    quietly gen double `psi'=exp(`t1')
    quietly gen double `ex1'=1
    quietly gen double `p1'=1
    quietly gen double `ex2'=1
    quietly gen double `p2'=1
    quietly gen double `a'=1
    quietly gen double `b'=1
    quietly gen double `p11'=1
    quietly replace `ex1'=exp(`t2'+`t3'*x11+`t4'*x12+`t5'*x13+`t6'*x14)
    quietly replace `p1'=`ex1'/(1+`ex1')
    quietly replace `ex2'=exp(`t2'+`t3'*x21+`t4'*x22+`t5'*x23+`t6'*x24)
    quietly replace `p2'=`ex2'/(1+`ex2')
```

```
    quietly replace `a'=1+(`p1'+`p2')*(`psi'-1)
    quietly replace `b'=-4*`psi'*(`psi'-1)*`p1'*`p2'
    if `psi'==1 quietly replace `p11'=`p1'*`p2'
    if `psi'~=1 quietly replace `p11'=(`a'-sqrt(`a'^2+`b'))/(2*(`psi'-1))
    quietly replace `lnf'=ln(1-`p1'-`p2'+`p11') if y1==0 & y2==0
    quietly replace `lnf'=ln(`p1'-`p11') if y1==1 & y2==0
    quietly replace `lnf'=ln(`p2'-`p11') if y1==0 & y2==1
    quietly replace `lnf'=ln(`p11') if y1==1 & y2==1
    quietly replace `lnf'=ln(1-`p1') if y1==0 & y2==.
    quietly replace `lnf'=ln(`p1') if y1==1 & y2==.
    quietly replace `lnf'=ln(1-`p2') if y1==. & y2==0
    quietly replace `lnf'=ln(`p2') if y1==. & y2==1
    quietly replace `lnf'=0 if y1==. & y2==.
end

ml model lf blr () () () () () ()
ml maximize
```